

Секция по компютърна лингвистика
Институт за български език „Професор Любомир Андрейчин“
Българска академия на науките



**ПОЛУАВТОМАТИЧНО РАЗРАБОТВАНЕ НА
МНОГОЕЗИКОВИ ТЕРМИНОЛОГИЧНИ РЕСУРСИ**

**Дисертация за присъждане на образователна и научна степен
„доктор“**

Област на висше образование: 2. Хуманитарни науки;

Професионално направление: 2.1. Филология;

Специалност: Общо и сравнително езиковедие

Автор:

Научен ръководител:

Виктория Петрова-Любенова

Проф. д-р Светла Коева

София

2022

Декларация за оригиналност и достоверност

от Виктория Красимирова Петрова-Любенова

Във връзка с провеждането на процедура за получаване на образователната и научната степен „доктор“ към Института за български език „Професор Любомир Андрейчин“ при Българската академия на науките, декларирам, че:

Резултатите и приносите в дисертацията на тема „Полуавтоматично разработване на многоезикови терминологични ресурси“, които представям във връзка с процедурата, са оригинални.

Всички чужди публикации, статии и книги, използвани за дисертацията, са надлежно цитирани и описани в библиографията, приложена към този труд. Научната разработка в цялост, както и нейните отделни части, не са представяни в подкрепа на друга процедура за придобиване на образователна и/или научна степен в университет или друга образователна институция в България и чужбина.

Представената от мен информация във вид на документи и публикации, лично съставени справки и други съответства на обективната истина.

10. 10. 2022 г.

гр. София

ДЕКЛАРАТОР:

/Виктория Петрова-Любенова/

Благодарности

Благодаря на колегите от Института за български език за това, че винаги бяха готови да ми дадат съвет, информация или разяснение по интересуващите ме въпроси, особено що се касае до правилниците на Института и на БАН като цяло.

Специални благодарности на гл. ас. д-р Валентина Стефанова.

Благодаря на съпруга ми за подкрепата и търпението през последните две години.

Благодаря на научния ми ръководител – проф. Коева за съветите, насоките, мотивацията и изобщо за времето. Без нейната помощ тази дисертация нямаше да види бял свят.

Посвещавам този труд на родителите ми.

Съдържание

УВОД	8
Актуалност на изследването	8
Мотивация за изследването	9
Предмет на изследването	10
Цели и задачи	10
Методи, използвани при работата по дисертацията	13
Композиция на изследването	14
ПЪРВА ГЛАВА	
Системи за компютърноподпомогнат превод и техните компоненти	16
I. Системи за компютърноподпомогнат превод и техните основни компоненти	17
I.1. Преводна памет	20
I.2. Машинен превод	24
I.3. Терминологични бази	34
I.4. Приложение на многоезикови терминологични бази в системите за компютърноподпомогнат превод	35
I.5. Нови функции на системите за компютърноподпомогнат превод	36
I.6. Използване на системите за компютърноподпомогнат превод в зависимост от лингвистичните характеристики на различните типове текстове	37
II. Отношението на преводачите към технологиите за компютърноподпомогнат превод	40
II.1. Изисквания на системите за компютърноподпомогнат превод към уменията на преводачите	40
II.2. Профилът на преводача в съвременната епоха	42
II.3. Нови професии, създадени заради технологиите за компютърноподпомогнат превод	44
II.4. Изисквания на преводачите към функционалностите на системите за компютърноподпомогнат превод	48
II.5. Проучване сред преводачите в България	51
ВТОРА ГЛАВА	
Терминологични бази в системите за компютърноподпомогнат превод	64
I. Термини, терминологични речници и терминологични бази	67
I.1. Термините в контекста на превода	67

I.2. Терминологични речници	70
I.3. Терминологични бази	72
II. Сравнение на терминологичните бази с терминологичните речници	81
III. Връзка на терминологичните бази с преводната памет	84
IV. Създаване и управление на терминологичните бази в системите за компютърноподпомогнат превод	85
IV.1. Притурки към основната програма	87
IV.2. Системи за управление на терминология	88
IV.3. Използване на контролирани езици	91
V. Умения за работа с терминология и терминологични бази	93
ТРЕТА ГЛАВА	
Полуавтоматично създаване на многоезикови терминологични бази	97
I. Създаване на терминологични ресурси в областта на компютърната терминология	102
I.1. Особенности на компютърната терминология	102
I.2. Речници с компютърна терминология	104
I.3. Автоматично извличане на термини: методи и добри практики	109
I.4. Инструменти за автоматично извличане на термини	113
I.5. Сравнение на инструментите за автоматично извличане на термини	114
II. Методика за полуавтоматично създаване на терминологична база в областта на компютърната терминология	124
II.1. Двуетиков корпус за извличане на термини	125
II.2. Извличане на термините от корпуса	126
II.3. Ръчна проверка и техническо оформление на терминологичната база	133
III. Описание на характеристиките на създадената терминологична база от данни в областта на компютърната терминология	135
ЧЕТВЪРТА ГЛАВА	
Измерване на качеството в системите за компютърноподпомогнат превод и техните компоненти	139
I. Измерване на точността на превода със системите за компютърноподпомогнат превод	142
I.1. Измерване на точността на превода без терминологични бази	143
I.2. Измерване на точността на превода с терминологични бази	144
II. Сравнение на ефективността на програмите за оценка на превода	148

III. Класификация на грешките при превод	153
IV. Сравнение на резултатите от автоматичните средства за измерване на качеството на превода	156
V. Международни стандарти за качеството на превода	162
Приноси на дисертацията	176
Приложение 1	177
Приложение 2	192
Приложение 3	196
Списък на някои от използвани (по-рядко употребявани) термини:	219
БИБЛИОГРАФИЯ	220

Фигури

Фигура 1: Интерфейс на Традос 2021

Фигура 2: Интерфейс на мемоКю 2021

Фигура 3: Интерфейс на Мемсорс

Фигура 4: Технологии за превод, използвани през 2021 и 2022 според Проучване на европейската езикова индустрия

Фигура 5: Професионалната заетост на участващите в анкетата

Фигура 6: Честота на използване на инструментите за компютърноподпомогнат превод

Фигура 7: Оценяване на инструментите за компютърноподпомогнат превод

Фигура 8: Отговори на преводачите относно употребата на преводна памет

Фигура 9: Отговори на преводачите относно употребата на терминологичните бази

Фигура 10: Произход на терминологичните бази

Фигура 11: Значение на терминологичните бази за качеството на превода според преводачите

Фигура 12: Как преводачите оценяват работата с терминологичните бази

Фигура 13: Познаване на международните стандарти за качество на превода

Фигура 14: Честотата на употреба на машинния превод от страна на българските преводачи

Фигура 15: Интерфейсът на Тием таун

Фигура 16: Резултати от извличането на термини с Фаиффилтърс

Фигура 17: Извличане на термини с програмата ИАКЕ.

Фигура 18: Част от извлечените термини от корпуса с програмата Термостат

Фигура 19: Търсене на еквивалентни термини на български

Фигура 20: Визуализиране на разпознат термин и неговия еквивалент за български в програмата Мейткат

Фигура 21: Резултати от проверката с Традос

Фигура 22: Резултати от проверката на новосъздадената база с МемоКю

Фигура 23: Методите за проверка на качеството на превода, използвани от българските преводачи

Таблицы

Таблица 1: Процентни съвпадения в преводна памет

Таблица 2: Сравнение на техническите характеристики на програмите за извличане на термини, направено през 2016 година от А. Зарецкая и др.

Таблица 3: Сравнение на техническите характеристики на програмите за извличане на термини, направено през 2022 година

Таблица 4: Съпоставка на техническите характеристики на самостоятелни програми за проверка на качеството на превода

Таблица 5: Класификация на грешки при превод и тяхното влияние за правилното пренасяне на съдържанието

Таблица 6: Допълнителна схема за оценяване на програмите за проверка на качеството на превода

УВОД

Актуалност на изследването

В периода от 60-те до 80-те години на миналия век надеждите, че машинният превод ще се развие достатъчно, че да замести човешкия превод, се оказват неоправдани¹. Затова усилията се пренасочват към разработването на програми, които да помогнат на преводачите. В днешно време по-голямата част от текстовете (най-вече специализираните) се превеждат с помощта на машинен превод (с последваща обработка от човек) и системите за компютърнопомогнат превод. Нарастващото използване на компютърнобазирани инструменти и ресурси за превод революционизира и променя начина на работа на преводачите и теорията на превода. Новите технологии са вече толкова неразривно свързани с преводаческата професията, че към днешна дата преводът изглежда немислим, без да се използва някакъв компютърен инструмент (било то и най-обикновен текстов редактор). Много учени са съгласни, че що се отнася до превода през 21 век, не може да се пренебрегне не само въздействието на компютърните и информационните технологии, но и на процеси като глобализацията, интернационализацията, локализацията на продуктите, което води до нови и специфични предизвикателства за преводачите. В това отношение трябва да се спомене нуждата от осигуряване на качество и покриването на изискванията за производителност, което предизвиква значителна преоценка на „модерните“ преводи (Имре 2020: 16).

Сериозното ангажиране с технологиите в процеса на превода бавно навлиза в изследванията по превод и набира скорост едва през второто десетилетие на нашия век. Технологиите, които са свързани с писането и превода, се появяват и развиват с течение на времето и са пряко обвързани с културните и социално-икономическите нужди на обществото, което ги създава и използва. Въпреки това, когато учените в тази област обсъждат технологиите за превод, фокусът е сравнително ограничен до дигиталните технологии и най-вече до преводната памет и машинния превод.

„Преводът може да е необходим по различни причини. [...] Много често преводачът иска читателят да приеме даден превод като текст сам по себе си, без да разбира, че това е превод. За да постигне това, преводачът трябва да коригира някои характеристики на изходния текст в съответствие с целевия език“ (М. Бейкър 1992: 112).

¹ Най-вече след доклада „Automatic Language Processing Advisory Committee“ или ALPAC от 1966 г.

„Дискусията за същината, важността и техниката на преводаческата работа получи в последните десетилетия тласък, сравним с малко други сфери на филологическата материя, и в това няма нищо случайно. (...) Всеки ползвател на преводен текст би могъл да открие в нея мотиви, които са част от собствения му езиков опит или са формирали личната му култура. Не по-малко значимо е, че с нарастващата важност на превеждането като дейност се увеличава и всеобщата зависимост от изкуството и професионализма на преводача. А всяка зависимост крие и рискове, като след тях рискът от съзнателна или несъзнателна манипулация е съществен проблем“ (Д. Карапеткова 2016: 9).

Преводът заема ключова роля в създаването на многоезикови новинарски статии за международната преса, телевизионните канали, интернет, социалните медии, блогове, Уикипедия и други (Хаус 2015: 101). Ролята на преводача се променя и може да се предвиди, че в недалечно бъдеще работата на преводача ще бъде почти изцяло съсредоточена върху последващото редактиране на машинен превод или върху предоставянето на висококачествени преводи на „чувствителни“ текстове като рекламни материали, литературни произведения и правна документация (Гамбие 2019: 350).

Мотивация за изследването

Необходимостта да се намерят евтини и ефективни решения налага промени в употребата на технологиите по отношение на инструментите за превод. „Универсалната революция в превода замества писалката и хартията на преводачите в свят на работни станции и инструменти за компютърноподпомогнат превод, които стават все по-сложни и превръщат преводача в нещо много по-близко до пилот на самолет, който е свързан с външния свят чрез невероятно разнообразие от инструменти, предназначени да му помогнат при намирането на информация, общуването с другите и извършването на всички задачи, очаквани от него“ (Гаудек 2007: XIII).

Една от най-важните цели на използването на софтуер за превод е да се преведе изходен текст, чийто език не е познат на потребителя, така че да се обясни съдържанието му на език, познат на потребителя. Установено е, че по-голямата част от търговските системи за машинен превод са за превод между индоевропейски езици, или между „основни“ езици с голям брой носители. Софтуерът за превод, който включва и други езици, е относително малък на брой (Син-уай 2015: 42). Най-голяма е необходимостта от специализирани и технически преводи, тъй като тяхното количество расте с голяма бързина. Такива преводи не могат да бъдат осъществени без познаването на правилната терминология за конкретната област. Ситуацията се усложнява допълнително от липсата на езикови ресурси за малки езици като българския. Затова настоящото изследване цели

да анализира каква е ситуацията в момента, и да предложи начин за преодоляване (поне доколкото е възможно) на липсата на нужните терминологични ресурси.

Обект на изследването

Настоящият дисертационен труд има следния обект на изследване: проучване, анализ и подбор на достъпни и ефективни методи за създаване на терминологични ресурси, предназначени за системите за компютърноподпомогнат превод. Обектът на изследване включва както съществуващите методи за създаване на терминологични бази, така и проучване на възможностите за интегриране на други технологии, които да спомогнат за улесняването и ускоряването (където е възможно) на процеса по създаване на нови терминологични ресурси.

Предмет на изследването

Приложението на терминологичните бази се разглежда от гледна точка на професионалните преводачи в контекста на системите за компютърноподпомогнат превод. Предмет на изследване са терминологичните бази и тяхната употреба в системите за компютърноподпомогнат превод, както и начините, по които съвременните технологии позволяват и улесняват създаването и употребата им. Изследването е фокусирано върху функционалностите на терминологичните бази в системите за компютърноподпомогнат превод и външните програми за извличане на терминология и нейната обработка (извличане, съпоставяне и конвертиране). Изследвани са нагласите на преводачите и средствата, които те използват за създаването на терминологични бази, анализирани са най-популярните инструменти в тази област (както платени, така и безплатни).

Езиците, които са обект на внимание са български и английски, а посоката на превод, която се разглежда в дисертацията, е от английски на български. В текста са описани ситуации, основаващи се на реална работна среда, за периода 2018 – 2022 година.

Създаването на терминологични ресурси се илюстрира с паралелни корпуси, терминологични речници и терминологична база в областта на компютърната терминология. Постоянното развитие на технологиите налага и съответното обновяване на компютърната терминология.

Цели и задачи

Основната цел на изследването е да предложи методика за създаване на терминологични ресурси за дадена специализирана област. Поради това са анализирани възможностите на програмите за компютърноподпомогнат превод за

създаването на терминологични бази, програмите за извличане на термини, както и наличните към момента терминологични ресурси за български език.

За постигането на **основната цел** се подхожда комплексно посредством следните **по-конкретни цели**:

Първа глава:

- Да се опишат в детайли съвременните системи за компютърноподпомогнат превод, като се акцентира върху техните основни компоненти (преводна памет, машинен превод, терминологични бази, включително многоезикови).
- Да се опишат новите познания и умения, които се изискват от преводачите във връзка с използването на технологиите за компютърноподпомогнат превод.
- Да се очертае съвременното състояние на използване на системите за компютърноподпомогнат превод, техните компоненти и нови функционалности.
- Да се проведе проучване сред преводачите в България относно използването на системите за компютърноподпомогнат превод и да се обобщят и анализират резултатите от него.

Втора глава

- Да се представят видовете терминологични речници и тяхната структурата в съпоставка със структурата на терминологичните бази. Целта е да се очертаят приликите и разликите, както и преимуществата при използването на терминологични бази в процеса на работа със системи за компютърноподпомогнат превод.
- Да се опише структурата и особеностите на терминологичните бази. Представяне за световноизвестни терминологични бази, които могат да се използват във или извън системите за компютърноподпомогнат превод.
- Да се представи създаването и управлението на терминологичните бази в системите за компютърноподпомогнат превод. Ще бъдат разгледани основните начини за създаване и управление на терминологични бази в системи за компютърноподпомогнат превод. В допълнение ще се разгледат притурки, които могат да се използват заедно със системите за компютърноподпомогнат превод, както и самостоятелно функциониращи системи за работа с терминология, които могат да бъдат използвани паралелно.

- Да се покаже използване на терминологичните бази в зависимост от различните типове текст за превод и приложението на контролирани езици. Тук основно внимание ще се обърне на изискванията за последователност и непротиворечивост при употребата на термини и възможността да се наложат специфични ограничения. Ще бъдат разгледани накратко контролираните езици и тяхното приложение в специфични области.

Трета глава

- Да се представят накратко особеностите на компютърната терминология, като стремежът е към обобщение от гледна точка на включването на компютърните термини в терминологични ресурси.
- Да се опишат начините за създаване на терминологични ресурси, като се започне от традиционните отпечатани терминологични речници, премине се през онлайн терминологичните речници и се достигне до автоматичното извличане на термини и представянето на съществуващите инструменти за тази цел. Описанието на инструментите ще бъде съпътствано от сравнение на техните функционалности в рамките на определен период от време.
- Да се създаде методика за полуавтоматично създаване на терминологична база в областта на компютърната терминология и да се опишат конкретните етапи от работата по създаването на двуезикова терминологична база (за английски и български език).
- Да се представят характеристиките на създадената двуезикова терминологична база.
- Да се предложи унифициран начин на работа при създаването на терминологични бази, основаващ се от най-добрите практики и вземащ под внимание възможностите и начина на работа на преводачите.

Четвърта глава

- Да се представи измерването на точността на превода със системите за компютърноподпомогнат превод, съответно без или с използването на терминологични бази.
- Да се представи измерването на точността на превода с функционалности, вградени във или външни за инструментите за компютърноподпомогнат превод.

- Да се предложи класификация на грешките при превод със системи за компютърноподпомогнат превод с оглед на тяхната значимост на съдържанието на текста.
- Да се направи сравнение на ефективността на програмите за оценка на превода по предварително зададена методика.

Методи, използвани при работата по дисертацията

При разработването на дисертацията са използвани различни съвременни методи (статистически, за компютърна обработка, корпусен), както и някои традиционни (описателен, аналитичен, експериментален). Основни подходи за постигане на изследователските цели са анализът и описанието. Към описателния метод на работа могат да се причислят прегледът на съществуващата научна литература в тази област, прегледът на известни терминология бази и прегледът на техническите инструменти, използвани в ежедневната работа на преводачите и под. Анализ се прилага при подбора на подходящи езикови ресурси и средства за тяхната обработка.

Използват се и негативни резултати, за да се вземе конкретно решение или да се направят необходимите изводи. Например, някои от предварително избраните програми за автоматично извличане на термини се оказаха неприложими за изпълнението на поставените задачи.

Формулирани са хипотези, например за нагласите на преводачите в България към новите средства за работа при превод или за количеството термини, които могат да бъдат извлечени в зависимост от изходния паралелен корпус. Хипотезите се проверяват експериментално по различен начин: посредством анкета, проведена онлайн; посредством работа с програми за автоматична обработка на текстове: за автоматично извличане на термини, за автоматично подравняване на паралелни документи по изречения, за автоматично създаване на двуезикови терминологични бази. При работата с всички програми за компютърноподпомогнат превод се използват съвременни средства за компютърна обработка на текстове; създаване на преводна памет, на терминологични бази, използване на машинен превод.

Използвани са и статистически методи за обобщение на събраната информация: от анкетата сред преводачите и при автоматичното извличане на термини. Корпусния подход се прилага при аотирането на преводни текстове с грешки с цел да се проверят функционалностите на програми за проверка на качеството на превода.

Работата по дисертацията изискваше професионална работа с поредица от системи за компютърноподпомогнат превод и техните компоненти (преводна памет и терминологична база), автономни системи за автоматично извличане на информация, системи за автоматично извличане на термини и подравняване на паралелни текстове по изречения и не на последно място – автономни системи за проверка на качеството на преода.

Композиция на изследването

Научната разработка се състои от четири самостоятелни глави.

Първа глава – *Системи за компютърноподпомогнат превод и техните компоненти*, представя системите за компютърноподпомогнат превод и техните основни компоненти (преводна памет, машинен превод и терминологични бази). Разгледани са нови функции на системите за компютърноподпомогнат превод, тяхното използване в зависимост от лингвистичните характеристики на различните типове текстове и изискванията на системите за компютърноподпомогнат превод към уменията на преводачите. Очертан е профилът на преводача в съвременната епоха и са показани нови професии, създадени във връзка с технологиите за компютърноподпомогнат превод. Направено е проучване сред преводачите в България, за да се обобщят и анализират степента на използване на новите технологии сред преводачите и техните нагласи по отношение на машинния превод и системите за компютърноподпомогнат превод.

Втора глава – *Терминологични бази в системите за компютърноподпомогнат превод* представя термините, терминологичните речници и терминологичните бази в контекста на превода. Направено е сравнение на терминологичните бази с терминологичните речници и на преводната памет и терминологичните бази. Анализирани са методите за създаване и управление на терминологичните бази в системите за компютърноподпомогнат превод, притурките към основната програма, системите за управление на терминология и уменията за работа с терминология и терминологични бази.

Трета глава – *Полуавтоматично създаване на многоезикови терминологични бази* представя създаването на терминологични ресурси в областта на компютърната терминология и особеностите ѝ. Анализирани са средствата за автоматично извличане на термини и е направено сравнение между тях. Разработена е методика за полуавтоматично създаване на терминологична база в областта на компютърната терминология чрез двуезиков паралелен корпус. След извличането на термините е

направена ръчна проверка и техническо оформление на терминологичната база. Описани са характеристиките на създадената терминологична база в областта на компютърната терминология.

Четвърта глава – *Измерване на качеството в системите за компютърноподпомогнат превод и техните компоненти* представя измерването на точността на превода със системите за компютърноподпомогнат превод без терминологични бази, с терминологични бази и с програми, външни на инструментите за компютърноподпомогнат превод. Предложена е класификация на грешките и е направено сравнение на резултатите от автоматичните средства за измерване на качеството на превода. Разгледани са международни стандарти, метрики и модели, свързани с качеството на превода.

В дисертацията има три приложения, показващи в пълнота резултатите от направената анкета сред преводачите в България, автоматично извлечените термини на английски в областта на компютърната терминология и английско българската терминологична база в областта на компютърната терминология.

ПЪРВА ГЛАВА

Системи за компютърноподпомогнат превод и техните компоненти

Настоящата глава представя системите за компютърноподпомогнат превод и техните компоненти. Има за основна цел да направи преглед на техническите им характеристики, както и на измененията, които налагат в работните процеси при превод, в образованието и в обучението на преводачите.

По-конкретно, Първа глава е насочена към изпълнението на следните цели:

- Да се опишат в детайли съвременните системи за компютърноподпомогнат превод, като се акцентира върху техните основни компоненти (преводна памет, машинен превод, терминологични бази, включително многоезикови).
- Да се характеризира използването на системите за компютърноподпомогнат превод в зависимост от лингвистичните характеристики на различните типове текст за превод.
- Да се опишат новите познания и умения, които се изискват от преводачите във връзка с използването на технологиите за компютърноподпомогнат превод.
- Да се очертае съвременното състояние на използване на системите за компютърноподпомогнат превод, техните компоненти и нови функционалности.
- Да се покажат изискванията на съвременните преводачи към съвременните средства, които подпомагат превода: системите за компютърноподпомогнат превод, машинния превод и терминологичните бази.
- Да се проведе проучване сред преводачите в България във връзка с използването на системите за компютърноподпомогнат превод и да се обобщят и анализират резултатите от него.

Както се вижда, съдържанието на главата е ориентирано около два основни фокуса: системите за компютърноподпомогнат превод и техните компоненти, от една страна, и качествата и уменията на съвременните преводачи, които използват системи за компютърноподпомогнат превод, от друга страна.

I. Системи за компютърноподпомогнат превод и техните основни компоненти

След неуспеха на първите прототипи на машинен превод, между края на 60-те и началото на 70-те години, се предлага нов подход: система, която не трябва да превежда автоматично, а да улеснява работата на преводача. Става въпрос за системите за компютърноподпомогнат превод (Computer Assisted / Aided Translation Tools) – софтуер, чрез който преводачът да може да превежда максимално бързо и лесно възложените му преводи. В тези системи преводачите имат възможност да редактират, създават, съхраняват и управляват съдържанието, докато в същото време имат на разположение терминологични бази, предишни преводи, машинен превод, проверка на правописа и други полезни функции (например визуализация на оригиналния документ, функция за използване и комбиниране на макроси, опция за диктуване и др.). Основното предимство на този вид програми е възможността им да разделят големи документи на сегменти (отделни изречения или абзаци, в зависимост от настройките), които се съхраняват в база от данни и които могат да бъдат използвани отново. Първата такава система е *Система за поддръжка на превода* (Translation Support System или TSS), разработена от американската компания Алпнет в средата на 80-те години (Гарсия 2015: 70). Поради високата си цена, тя е достъпна само за големи компании, извършващи огромни количества преводи. През втората половина на 80-те, холандската компания ИНК (или INK) разработва ТекстТулс (TextTools). През 1984 г. в Щутгарт се разработва Традос², която и до днес се приема за най-използваната система за компютърноподпомогнат превод. През 90-те години разширяването на пазара на инструменти за компютърноподпомогнат превод прави софтуера достъпен за малкия бизнес и преводачите на свободна практика. Достъпността на интернет пък дава възможност на преводачите да обменят информация без оглед на географските ограничения в световен мащаб. Променят се също и сроковете за изпълнение на възложените преводи и методите на работа. Всичко това води до нуждата от въвеждане на общоприети стандарти за качество като ISO (описани в последната глава на този труд). Така технологиите продължават да се развиват и през 2015 г. са изброени шестдесет и седем налични системи на пазара (Гарсия 2015: 22). През 2020 г. бройката се покачва на 164³, като вече не се вземат предвид системите за компютърноподпомогнат превод в

² <https://www.trados.com/>

³ Според годишния доклад на Нимзи за 2020 г. <https://www.nimdzi.com/nimdzi-language-technology-atlas-2020/?hilite=trados>

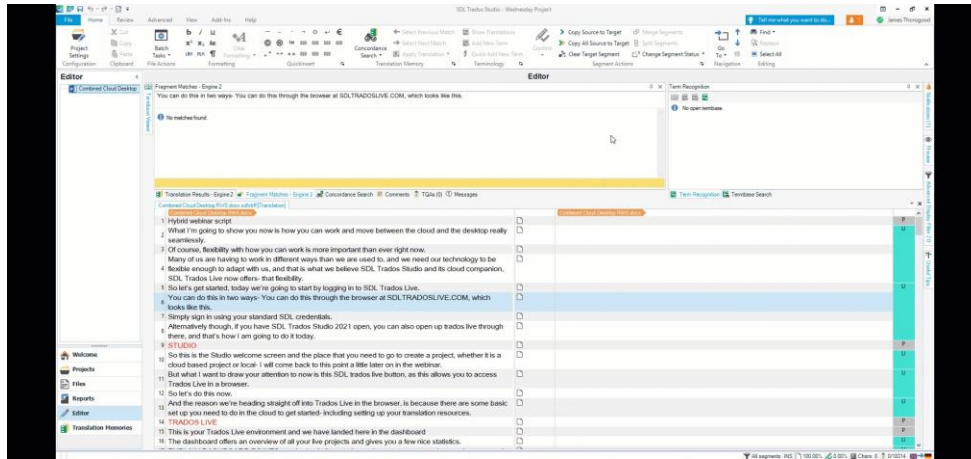
традиционния им вид, а „системите за управление на превод“ (Translation Management Systems). Последните обхващат нови услуги като десктоп версия, сървърна версия, локализация за програмисти, js прокси локализация и интеграция с външни програми и инструменти (например с програми за управление на проекти за превод и техния бюджет).

Създателите на системите за компютърнопомогнат превод дават следните дефиниции за този вид програми: „Инструментите за компютърнопомогнат превод са софтуерни приложения, които помагат за превода на текстове от един език на друг. По-конкретно, инструментите за компютърнопомогнат превод се използват за възлагане, редактиране, управление и съхранение на преводи“ според Мемсорс⁴; „Инструментите за компютърнопомогнат превод разделят големи многоезикови документи на сегменти (изречения и абзаци), които се съхраняват в база от данни, наречена преводна памет: вече преведени материали могат да бъдат използвани повторно по всяко време“ според мемоКю⁵; „Това е софтуер, използван от преводачи и лингвисти. Има различно предназначение, но преди всичко подпомага процеса на превода. Позволява редакция, създаване, управление на превода“ според Традос⁶. Изображенията по-долу илюстрират интерфейса на няколко такива програми. Най-голямо пространство на екрана заема текстовият редактор. Той е разделен на две части: от лявата страна е изходният текст, разделен на отделни сегменти, а от дясната – целевият текст. Към всеки целеви сегмент са посочени процентите на съвпадение с наличната информация в преводната памет. Както се вижда от Фигура 2, сегментът върху който се работи, е ограден в рамка. При кликване върху него, програмата визуализира всички налични ресурси, свързани със сегмента. Програмата от Фигура 3 дава възможност и за визуализиране на оригиналния документ, което може да е от полза за преводача при избора на най-подходящия превод на даден сегмент. Над езиковия редактор (Фигура 1) и от дясната му страна (във Фигури 2 и 3) се показват наличните за конкретния сегмент ресурси. Всички програми имат лента с допълнителни настройки.

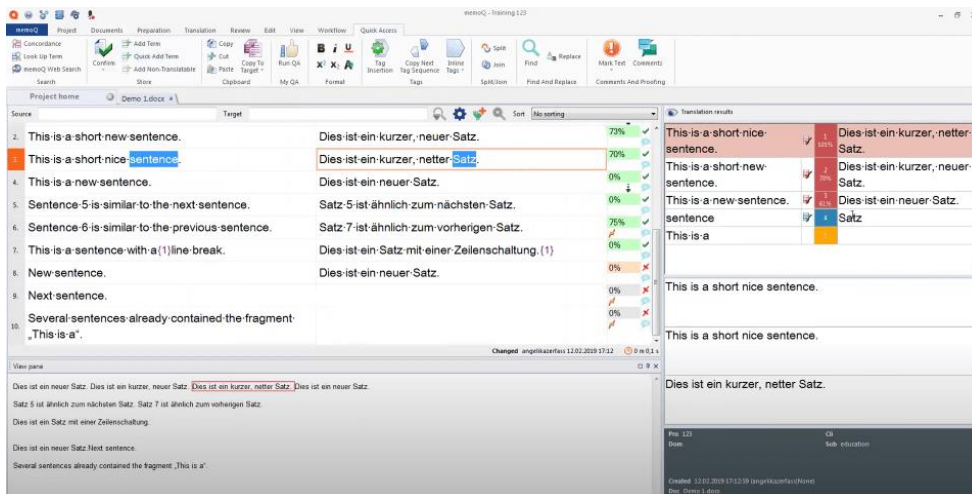
⁴ <https://www.memsource.com/what-are-cat-tools/>

⁵ <https://www.memoq.com/tools/what-is-a-cat-tool>

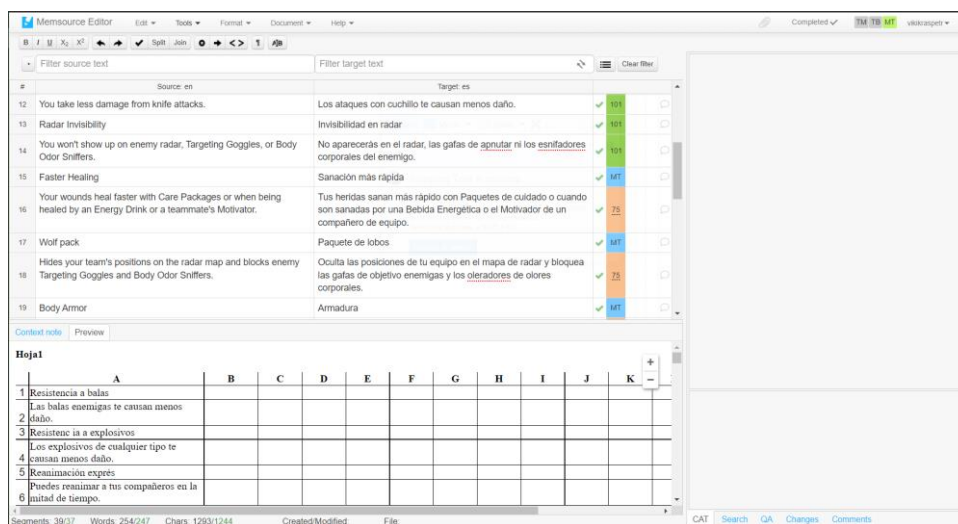
⁶ <https://www.trados.com/solutions/cat-tools/translation-101-what-is-a-cat-tool.html>



Фигура 1: Интерфейс на Традос 2021



Фигура 2: Интерфейс на мемоКю 2021



Фигура 3: Интерфейс на Мемсорс

Софтуерът за управление на превод е предназначен да организира и управлява целия процес на превод и обикновено обхваща и функции, които не са свързани конкретно с превода. Пример за такива функции са: управлението на връзката с клиента (client management), управлението при създаване на споразумения, инструкции, спецификации и процедури (management of agreements, briefs, specifications and procedures), създаването на оферти (generation of estimates), отчитането на работното натоварване и планирането на работния поток (workload and work flow scheduling), както и анализът след завършването на проектите (post-mortem project analysis).

Въпреки различията във функционалностите, доколкото ни е известно, всички системи за компютърноподпомогнат превод имат две общи характеристики: наличието на преводна памет и терминологична база. Свързването с машинен превод също е възможно, но то води до промени в работния процес и налага друг вид технически познания на преводачите. Поради тази причина връзката с машинния превод се разглежда отделно.

I.1. Преводна памет

Преводната памет (Translation Memory или ТМ) е база от данни, която съхранява изрази от вече преведени от човек документи. При въвеждането в програмата за компютърноподпомогнат превод текстът се сегментира (по подразбиране на всеки пунктуационен знак за край на изречение – точка, въпросителна или удивителна, две точки и точка и запетая, но са възможни и персонализирани настройки) и подравнява (aligned) в съответстващи си двойки изходни и целеви езикови единици (заглавия, словосъчетания, изречения), които са готови за повторна употреба, след което преводачът превежда сегмент по сегмент. Всички сегменти от изходния език и целевия език се съхраняват в паметта. Когато даден израз бъде засечен отново в същия или нов документ, преводът му се предлага от паметта. Въпреки че качеството на машинния превод значително се подобрява през последните няколко години (най-вече с напредъка в областта на невронния машинен превод), все още се предпочита използването на преводна памет. По този начин не само е възможно да се превеждат „перфектно“ вече преведени изречения, но също така се предлага „почти перфектно“ качество на превода, когато се извличат от паметта подобни изречения (Булте и Тецкан 2019).

Основната единица в паметта, сегментът, обикновено е разграничена с пунктуационен знак и следователно обикновено е изречение, но може да бъде и заглавие, мото, формула и др. Сегмент на изходния език (source segment), свързан със своя превод (target segment), плюс съответните метаданни (например час, дата и име на преводача

или редактора, извършил корекцията, име на клиента, тематична област и т.н.) образуват заедно т. нар. „единица за превод“ (Translation Unit или TU). Паметта съдържа също и алгоритъм за разпознаване на съвпадения (matches): когато идентичен или подобен сегмент се срещне в нов текст. Ако в паметта се открие изходен сегмент в базата от данни, който точно съвпада със сегмента от новия текст, се предлага съответният превод като точно съвпадение или 100% съвпадение. При такава ситуация това, което преводачът трябва да направи, е да провери дали сегментът може да бъде използван повторно такъв, какъвто е, или са необходими корекции заради разлики в контекста. Ако в паметта се открие изходен сегмент в базата от данни, който е подобен на преждания сегмент, той се предлага като частично съвпадение (Fuzzy Match) заедно със степента му на подобие, посочена като процент. Подобие то се изчислява по Разстоянието на Левенщайн⁷, т.е. минималния брой вмъквания, изтривания или замествания на символи; след това преводачът преценява дали предложеният превод може да бъде адаптиран, или по-малко усилие ще представлява да се направи изцяло нов превод.

Според настройките по подразбиране обикновено се предлагат и визуализират само сегменти със съвпадение от над 70% (за Традос е 75%), тъй като се приема, че съвпаденията с по-нисък процент са в по-голяма степен пречка, отколкото полезни за преводача, ако в паметта не може да бъде намерен сегмент, надвишаващ предварително зададения праг на съвпадение⁸. Случаят се нарича праг на частично съвпадение (Fuzzy Match Threshold). Когато от паметта няма предложения, е налице „липса на съвпадение“ (No Match Found) и преводачът трябва да преведе този сегмент по традиционния начин (Гарсия 2015: 71-72). По-новите версии на системите за компютърнопомогнат превод имат функция за контекстни съвпадения (Context Matches). Контекстното съвпадение е известно още като „101% съвпадение“ и се наблюдава, когато има точно съвпадение със сегмент в преводната памет, както и съвпадение на неговия контекст: т.е. предходният и следходният или следходните сегменти също са точно съвпадение, т.е. 100%. В Таблица 1 са показани няколко примера с различните проценти на съвпадение. Необходимо уточнение е, че примерите не са реални⁹ и имат само илюстративен

⁷ Разстоянието на Левенщайн между два низа е броят на изтривания, вмъквания или замествания на единичен символ, необходими за трансформиране на един низ в друг. То е известно още и като разстояние за редактиране.

⁸ Трябва да се отбележи, че са възможни разлики в изчисляването на частичните съвпадения. Влияние в по-малка или голяма степен могат да окажат: словоредът, пунктуацията (понякога влиянието не е важно, а друг път може напълно да промени значението на сегмента), форматирането и таговете (елементи, които съдържат кодирана информация за формата и структурата на дадена дума, фраза или сегмент).

⁹ Част от примерите са взети от статия на преводаческата платформа Тиетаун: <https://www.tm-town.com/>

характер. Причината за това е конфиденциалността на информацията в истинската преводна памет. Обикновено тя е собственост на преводаческа агенция или компания.

Съвпадение	Изходен език	Целеви език
1% – 74%	She asked for the store. she asked for the store!	Тя попита къде е магазинът. Тя попита за магазина. Тя поиска да знае къде е магазинът. Тя искаше да знае къде се намира магазинът.
75% – 99%	Bob went to the store. Bob went to the store	<u>Боб</u> отиде до магазина. <u>Борис</u> отиде до магазина.
100%	John went to the store. John went to the store.	Джон отиде до магазина. Джон отиде до магазина.
101%	It was a rainy day. John went to the store. He needed to buy some eggs.	<u>Беше дъждовен ден.</u> Джон отиде до магазина. <u>Трябваше да купи яйца.</u>

Таблица 1: Процентни съвпадения в преводна памет

Две допълнителни функции в помощ на преводача са вътрешното частично съвпадение и възстановяването на частично съвпадение. Вътрешното частично съвпадение (Internal Fuzzy Match) е частично съвпадение в изходния текст. Това се случва, когато две изречения в документа съвпадат. В този случай съвпадението не произтича от паметта, а от текущия документ, който се превежда. Възстановяването на частично съвпадение (Fuzzy Match Repair) включва използването на машинен превод, който се опитва да „поправи“ частта от сегмента в паметта, която не съвпада с изходния сегмент. Това е сравнително нова функция, която все още не е налична във всички програми за компютърнопомогнат превод. Освен като бази от данни преводната памет може да бъде характеризирана също и като паралелни корпуси, които се създават от системите за управление на преводната памет (Translation Memory Management

System или TMMS). Растящата популярност на програмите за компютърнопомогнат превод налага възприемането на общ стандарт за преводна памет. Такъв е .tmx (съкратено от Translation Memory eXchange). Той позволява паметта да бъде съвместима, дори когато е създадена чрез различни приложения и се използва от разнообразни системи за управление на преводна памет. Повторното използване е и една от причините преводната памет да се продава¹⁰, а отделни институции да позволяват ползването ѝ безвъзмездно¹¹.

Преводната памет обикновено е частна собственост, тоест принадлежи на отделните преводачи или преводачески компании. Това обаче не означава, че не може да бъде извлечена от публично достъпни паралелни корпуси. Например някои много големи многоезикови паралелни корпуси като Юролекс¹² и Юропарл¹³ или Аки Комюнотер¹⁴ (цялото законодателство на ЕС, състоящо се от над един милиард думи) са достъпни, както като подравнени паралелни корпуси, така и в стандартен формат на преводна памет (Занетин 2015: 445).

Преводната памет се нуждае от три различни вида софтуер, за да функционира правилно:

- система за подравняване на сегменти (aligner), посредством която се създадат файловете за преводна памет. Може да бъде както самостоятелна програма, така и вградена функция в системите за компютърнопомогнат превод. В нея се съпоставят изходните и целевите сегменти от вече преведения текст, за да се създаде преводната памет;
- система за управление на формата, която е вградена опция в програмите за компютърнопомогнат превод. Тя следи форматирането на изходния текст да съвпада с форматирането на преводния текст;
- система за управление на преводната памет, която дава възможност за преместване на съдържание между различна преводна памет, изтриване,

¹⁰ Двете най-големи платформи за обмяна на преводна памет са: <https://www.tm-town.com/> и <https://www.ttmem.com/>

¹¹ Още от 2007 г. Генерална дирекция „Писмени преводи“ към Европейската комисия предоставя своите паметни за свободно ползване: <https://ec.europa.eu/jrc/en/language-technologies/dgt-translation-memory>

¹² ЕуроЛекс (EuroLex) – онлайн портал за правото на ЕС, който дава достъп до правните документи на всички официални езици на ЕС и се актуализира ежедневно. <https://eur-lex.europa.eu/homepage.html?locale=bg>

¹³ Еуропар (EuroPar) – портал с актуална информация относно работата на Парламента на ЕС: <https://www.europarl.europa.eu/portal/bg>

¹⁴ Аки Комюнотер (Acquis Communautaire) – френски термин, отнасящ се до съвкупността от закони на Европейската общност. <https://www.eurofound.europa.eu/observatories/eurwork/industrial-relations-dictionary/acquis-communautaire>

експортирате на части от съдържанието, както и историята на всеки сегмент, намиращ се в паметта (Гаудек 2007: 271 – 272).

Създадени са огромни по обем преводни паметни, които се поддържат и разширяват постоянно. Още през 2007 г. Д. Гаудек прогнозира, че „преводната памет ще намери своя път на пазара. Очертава се нов бизнес сегмент и профил на работа [...], като се започне от преводна памет с обща тематика и се стигне до специализирани и ясно дефинирани области (превод на баланси, договори, инструкции за работа и др.)“ (Гаудек 2007: 273). Според него от преводачите зависи да използват сегментите, които са потвърдени като еквивалентни по значение, т.е. качеството на потвърдения сегмент не трябва да подлежи на съмнение, когато се използва отново. Последното е изключително важно. При стандартен превод, след като преводачът е приключил с даден сегмент, сегментът става част от паметта. Това на свой ред отваря възможността за грешки – ако сегменти от преводната памет са потвърдени невнимателно. Даден преводач може да приеме съвпадение от 100%, без да отдели необходимото време за проверка. Поради тази причина не е необичайно преводачи и преводачески агенции да предприемат ревизия на преводната памет от гледна точка на начина на използване на езика при подбора на думите и изразите. Според Д. Гаудек в този контекст „начин на използване на езика“ трябва да се разбира като: „(а) дискурса и стереотипите на изказване; (б) съвкупността от фрази и формулировки, специфични за конкретен автор, тип документ, професионална област и т.н.“ (Гаудек 2007: 274). Причината преводната памет да е изключително ценна, е, че спестява време и средства на притежателите си и спомага за увеличаване на ефективността и производителността на преводачите. Преводачът прекарва по-малко време в търсене на правилния превод, защото трябва само да сравни повторенията, идентифицирани от паметта. На свой ред, по-малкото време за търсене на правилния превод позволява извършването на по-голямо количество преводи (Мичъл-Шуитевърдер 2020: 5).

I.2. Машинен превод

В продължение на много години към машинния превод има утопично очакване, че ще замени човека, и дори се възприема като заплаха за професията на преводачите. След отрицателния доклад на Консултативния комитет за автоматична обработка на езика¹⁵ по отношение на развитието на машинния превод през 1966 г., опитите за

¹⁵ Automatic Language Processing Advisory Committee – комисия от седем учени, създадена през 1964 г. от правителството на Съединените щати. <https://blog.pangeanic.com/alpac-report>

подобряване на машинния превод намаляват, но не спират. През последните години, чрез безплатните сайтове, машинният превод става достъпен за всички и в същото време се превръща в съществена част от преводаческата индустрия. Що се отнася до обикновените потребители, през 2016 г. 143 милиарда думи са превеждат всеки ден от Гугъл транслейт (Ууд¹⁶). Според статия на Д. Адауси от август 2021 г., „Гугъл транслейт се използва за превод на над сто милиарда думи на ден“¹⁷. Лесният достъп и начин на използване на тези инструменти допринасят преводът да се превърне в широко разпространена дейност. Това на свой ред е свързано с изучаването на езици, с туризма и пътуванията, както и с възможността за контакти с други култури (Пим и Торес-Симон 2021: 479).

Константин Дранч¹⁸ оценява размера на пазара на машинния превод на 440 милиона щатски долара за 2021 г., което е с 16% ръст, въз основа на изследвания, направени през последните две години. Числата за отделните компании са приблизителни, защото само 50% от компаниите са съгласни да споделят своите продажби. Константин Дранч уточнява, че делът на машинния превод „е поне два пъти по-голям от системите за преводни памети и расте по-бързо“.

Най-общо казано, машинният превод е автоматизиран процес, в който се използва специален софтуер, за да превежда текст от един естествен език на друг естествен език, без да има човешка намеса. Става въпрос за компютърногенериран превод, базиран на набор от специфични алгоритми. Обикновено този вид програми са бързи и лесни за използване, макар и не винаги да предлагат най-доброто качество за превод. Необходимо уточнение е техническата разликата между отделните системи. Двата най-популярни и разпространени модела са статистическият и невронният машинен превод. При статистическия машинен превод (Statistical Machine Translation или SMT) декодерът, който „превежда“, е по същество алгоритъм за търсене. За всяка дума и група думи в изречението за превод алгоритъмът проверява наличните ресурси, съдържащи подравнени думи и групи от думи, и извлича най-добрия еквивалент. Този метод може да се определи и като парадигматичен: всяка изходна дума създава слот, който може да бъде попълнен от дадена дума от наличните преводни думи и фрази

¹⁶ Twenty years of building for everyone. Company Announcements: <https://www.blog.google/inside-google/company-announcements/twenty-years-building-everyone/>

¹⁷ Информацията е публикувана в Scientific Editing, преводаческа агенция, специализирана в превода и редакцията на научна литература <https://www.scientific-editing.info/blog/everything-you-need-to-know-about-google-translate/#>

¹⁸ Журналист, преводач и съосновател на Custom.MT, специализирана в създаването на персонализирани услуги за невронен машинен превод. <https://custom.mt/>

(Боукър и Сиро 2019: 37). Невронният машинен превод (Neural Machine Translation или NMT) работи по различен начин. Декодерът не търси конкретен елемент от наличните ресурси. Използва първо невронните мрежи, за да се научи и след това да идентифицира най-добрите последователности за превод на цели изречения. Декодерът се обучава от големи по обем паралелни данни. Невронният машинен превод се опитва да изгради последователност от думи при превода линейно. Всяка последователност от предишни думи определя следващата дума.

Основните тенденции при компютърнопомогнатия превод в момента също поставят все по-голям акцент върху машинния превод. Разликата днес е в това, че за машинния превод се говори през призмата на изкуствения интелект и дълбоките невронни мрежи. Два примера за употребата на машинния превод в комбинация с машинното самообучение, са стратегиите на Ибей (Ebay) и Еърбиендби (Airbnb). През 2014 година Ибей въвежда свой собствен машинен превод (eBay Machine Translation или eMT), който има за цел да превежда продуктите заглавия на сайта. Изследователите, направили проучване, установяват, че въвеждането на eMT (което носи със себе си значително подобрение на качеството в превода) увеличава износа на eBay с 10,9%. Износът нараства най-вече за артикули с по-дълги заглавия, по-евтини продукти и при купувачи с по-малко опит. Учените заключват, че „въвеждането на eMT е еквивалентно на увеличение на износа от намаляване на разстоянията между страните с 26,1%. Това сравнение показва, че възпрепятстващият търговията ефект на езиковите бариери е от първостепенно значение“ (Бриньолфсон и др. 2019: 11). На 9 ноември 2021 г. Еърбиендби¹⁹ обявява, че компанията е внедрила машинен превод, който позволява на потребителите автоматично да четат преводи на рецензии и описания на повече от 60 езика, без да се налага да натискат бутон за превод. Става въпрос за проект с отворен код, в който са предоставени предварително обучени модели. Те непрекъснато се подобряват въз основа на корекции от лингвистите, работили върху съдържанието на компанията през последните години. Според Ербиенби, „Машинният превод подобрява качеството на повече от 99% от обявите на Ербиенби“.

В продължение на много години технологиите за машинен превод и тези за компютърнопомогнат превод са противопоставяни една на друга. В последните години обаче машинният превод (най-вече след навлизането на невронния машинен превод) започва да заема все по-централна роля в работните процеси при

¹⁹ <https://news.airbnb.com/2021-winter-release/>

компютърноподпомогнатия превод. Например в най-новата версия на Традос използването на машинен превод става неразделна част от процеса на всеки нов проект: след като документите за превод, паметта и терминологичната база бъдат заредени, програмата създава машинен превод специално за този проект²⁰.

Първите опити за интеграция между преводната памет и машинния превод се отнасят до възможността те да се комбинират по два начина. Единият е потребителят да получава за всеки сегмент предложения от машинния превод, преводната памет и терминологичната база и те да бъдат визуализирани едновременно. Този подход е изключително полезен, когато няма точно или частично съвпадение между сегментите в преводната памет. В такива ситуации машинният превод остава единствена възможност за помощ при превода. Според проучване на Канавос и Картсаклис още от 2010 г. предложенията, дадени от машинния превод, увеличават производителността за сегменти, които имат съвпадения с изходния език под 75–80%. Същото важи и за преводачите по време на индивидуална работа (Канавос и Картсаклис 2010: 11). Вторият начин за комбиниране е използването на двете технологии заедно, за да се подобрят резултатите в целевия език и по този начин да се увеличи производителността и да се намалят усилията след редактиране. Sybase IT²¹ отчита увеличаване на производителността след такова комбиниране, при условие обаче, че машинният превод е обучен със специфични данни (Байер 2012).

В същото време, въпреки че използването на технологиите за машинен превод е неизбежно, практически погледнато резултатът от всеки машинен превод може да изисква някаква форма на човешка намеса. Най-важната причина машинният превод да не може да се използва самостоятелно²², си остава качеството на изходния текст. Това важи както за свободnodостъпния, така и за специализирания машинен превод. Един начин за подобряване на качеството на машинния превод е предварителната обработка на текстовете, използвани за обучение: те се филтрират, за да се премахнат неизвестни термини, неясноти и многозначност от различен характер; текстовете се преобразуват в „опростен“ език с кратки изречения. Чрез такава обработка преди машинния превод всички неизвестни или недостатъчно ясни елементи биват идентифицирани и преведени от човек или биват обработени, така че компютърът да може да въведе необходимите

²⁰ Свързването става по различен начин за по-новите и по-старите версии на програмата. <https://www.trados.com/products/machine-translation/>

²¹ Днешното име на компанията е Sap – <https://www.sap.com/index.html>

²² Не се взема под внимание употребата на машинния превод за лични цели.

еквиваленти в системата. Друг начин за подобряване на качеството на машинния превод е чрез „обучение“ с помощта на преводна памет. От гледна точка на качеството този метод се приема за по-добър, тъй като обикновено потвърдените сегменти в паметта са одобрени от професионални преводачи. През 2013 г. група учени предлагат системата TMTPrime²³, която има за цел да улесни интегрирането на машинния превод и преводната памет (Дара и др. 2013: 10). Опити в тази посока са правени и в комбинация със статистически машинен превод и след това с невронен (Еригуки и др. 2019).

Яп фон Дер Мер определя инструментите за машинен превод като „най-доброто решение за улесняване на интегрирането на машинен превод в работния процес на професионалните преводачи“. Основните ограничения затова са цената (повечето са собственост на компания и са достъпни само като част от разпространението на продукта) и ограничения във възможностите за промяна (Дер Мер 2020: 293).

Някои инструменти за компютърноподпомогнат превод използват машинен превод, основаващ се на предишни примери (example-based machine translation – EBMT), за да осигурят така нареченото сегментно сглобяване (segment assembly). Става въпрос за търсене на отделни фрагменти от изходния сегмент в преводната памет и терминологичната база, извличане на техните преводи и комбинирането им, за да се получи превод за целия целеви сегмент. Частите от текста, за които не е намерено съвпадение в нито един от източниците, се вмъкват в предложението за превод на оригиналния език. При този случай обаче не се използва усъвършенствана технология на машинен превод, а само се прави търсене, подобно на нормалното търсене в преводната памет. Използват се и по-сложни техники за конструиране на целеви сегменти от по-къси фрагменти, извлечени от преводна памет (например при Мемсорс). Най-новите версии на инструментите за компютърноподпомогнат превод използват усъвършенствани технологии за машинен превод. Предизвикателство остава как машинният превод може да бъде интегриран в работния процес на преводачите по най-удобния за потребителите начин. Развитието на технологиите дава възможност много от системите за машинен превод да бъдат снабдени с плъгин, чрез който да се свържат със съответния инструмент за компютърноподпомогнат превод.

Нарастващото интегриране на машинния превод в работния процес на проектите за превод води до така наречения „разширен превод“ (augmented translation). Разширеният превод комбинира уменията на преводачите и машинния превод. А.

²³ <http://www.meta-net.eu/events/german-lt-day-2013/posters/cngl.pdf>

Аренас определя тази дейност като „преразглеждане на предварително преведен текст, генериран чрез машинен превод на изходния текст, и коригиране на възможните грешки с цел да бъдат постигнати предварително заложените критерии за качество“²⁴ (Аренас 2020: 333). Може да се направи разграничение на два вида редактиране на машинния превод: частично (light) и пълно (full) редактиране. В първия случай текстът трябва да бъде разбираем, като се допускат някои граматически и правописни грешки. Във втория случай текстът трябва да бъде не само разбираем, но също стилът, граматиката, правописът и терминологията трябва да бъдат сравними с тези на текст, преведан от човек. Тези процедури отнемат значително време и са необходими големи обеми от данни, за да бъде ефективна работата на машинния превод. Необходими са и специализирани технически познания, които не само че не принадлежат към компетенциите на преводачите, но и най-често не представляват интерес за тях. Всичко това повдига въпроса за достъпността на тези технологии. Големите преводачески агенции и доставчици на езикови услуги могат да обучават специфични за дадена област модели за машинен превод, но преводачите на свободна практика нямат такава възможност и често отказват да използват машинния превод като помощно средство. Освен това малките преводачески агенции рядко разполагат с ресурси за създаването на качествен машинен превод за всички езикови двойки, от които се нуждаят. Допълнително ограничение налагат някои клиенти, които забраняват на преводачите да използват машинен превод поради риск от проблеми с поверителността.

Независимо от недостатъците и ограниченията, употребата на технологиите за машинен превод не само няма да намалее, но и ще продължи да се увеличава все повече. Подобряването на качеството на машинния превод и неговото разпространение се доказват и от факта, че в началото на новото хилядолетие нито един инструмент за компютърноподпомогнат превод не е притежавал функционалност за вграждане на машинен превод, докато през последните години не само се предоставят различни възможности за работа с машинен превод, но и се дава възможност за включване на машинния превод като част от даден проект за превод. В дългосрочен план доставчиците на програми за превод подчертават като важна характеристика на своите продукти „помощта“ за преводачите и включват в разработването им машинния превод чрез плъгин, който може да бъде активиран или деактивиран по време на превода.

²⁴ Важно е да се подчертае, че критериите за качество се определят заедно с възложителя на проекта преди започването на работата.

Видовете системи за машинен превод варират между готови за употреба, адаптивни системи (в които се използва обратната връзка с потребителя, за да не се повтаря дадена грешка) и специализирани, създадени по поръчка системи. Възможностите за свързване с инструменти за компютърноподпомогнат превод също варират. Това се дължи и на новите модели за компютърноподпомогнат превод (които, от една страна, комбинират в себе си много повече функционалности, а от друга страна, са изцяло в облак или са насочени към локализация). Най-новите версии²⁵ на мемоКю и Традос позволяват интеграция с машинен превод чрез така наречения приложно-програмен интерфейс (Application Programming Interface или API). Модерн МТ²⁶, Амазон Транслейт²⁷ и Дийпел²⁸ предлагат подобна интеграция.

Някои от компаниите, които разработват компютърноподпомогнат превод, създават свои системи за машинен превод. Такъв е случаят с Традос: от момента, в който е бил интегриран невронният машинен превод на Традос (в обратната връзка се включват 40-50% от потребителите)²⁹, той се използва около два пъти повече от Гугъл или Дийпел при компютърноподпомогнатия превод със системата.

Има няколко компонента, които трябва да се вземат предвид, когато става въпрос за сравнение на достъпните програми за машинен превод. Дийпел работи с 26³⁰ езика (и всяка комбинация на тези езици посредством преходни езици (pivot language, т.е. език, който служи като мост между два други езика); Гугъл работи с 109 езика³¹ (и също използва преходни езици), а Традос работи с 52 езика³² (не използва преходни езици, но за някои от езиците, включително за испански, португалски, немски, арабски и италиански, се предлагат до седем директни комбинации към други езици (освен английски), плюс китайски и японски).

Въпреки многото и различни технологии няма машинен превод, който да е изцяло автоматизиран и да не се нуждае от някакъв вид човешка намеса. Аналогично – трудно е да си представим, че човешкият превод в днешно време може да бъде извършен без помощта на компютри (Уанг и Пинг 2020: 3). Както стана ясно от описаното по-горе,

²⁵ Тази информацията е валидна към 1 октомври 2021 г. https://marcinbas.github.io/api_mt/

²⁶ Модерн МТ <https://www.modernmt.com/>

²⁷ Амазон Транслейт (AWS) <https://aws.amazon.com/translate/>

²⁸ Дийпел <https://www.deepl.com/translator>

²⁹ Главен продуктов мениджър за продуктите на Традос Транслейшън Продъктивити като част от портфолиото от технически продукти на Ердабълюес Груп (RWS group).

³⁰ <https://support.deepl.com/hc/en-us/articles/360019925219-Languages-included-in-DeepL-Pro> (посетен на 20.07.2022)

³¹ <https://cloud.google.com/translate/docs/languages> (посетен на 20.07.2022).

³² <https://docs.rws.com/980998/338184/trados-studio-2022/supported-languages> (посетен на 20.07.2022).

визуализирането на резултатите от интегрирания машинен превод в програмите за компютърноподпомогнат превод е идентично с това на преводната памет. Този вид визуализиране на предложенията би могъл да доведе до объркване (езиковите резултати от паметта и машинния превод биха могли да съвпадат) и да затрудни търсенето на подходящи за преводача езикови ресурси. Би трябвало да се показват правилните предложения от различните източници, които са свързани с програмата за компютърноподпомогнат превод (машинен превод, преводна памет), но няма гаранция, че предложенията на машинния превод ще съдържат търсените от преводача думи и изрази. Например, може да има няколко синонима в две или три различни предложения, но не и тази дума, която е необходима³³.

Друг много важен аспект е човешкият фактор, защото дискусиите относно инструментите за компютърноподпомогнат и машинен превод често могат да бъдат емоционално натоварени, особено от страна на преводачите, които от една страна, не искат да използват новите технологии, защото им се струват прекалено сложни, а от друга, се страхуват от тях и ги виждат като заплаха за работното си място. Последният вид страх е отдавна известен: някои преводачи мислят, че компютърните помощни средства ще елиминират професията на преводача и ще заместят хората (Шилдс 1999: 4). Резултатите от изследвания, проведени през първото десетилетие на новото хилядолетие в тази сфера, сочат и допълнителни проблеми като ниските нива на удовлетвореност от работата, непригодността на инструментите за компютърноподпомогнат превод за нуждите на упражняване на свободна професия и високата цена на лицензите за използване на тези инструменти. Забелязва се консервативно отношение към технологичните инвестиции (Зафра 2006)³⁴. С времето всъщност не малко преводачи започват да разглеждат интернет технологиите като „необходимо зло“: повишаването на производителността става наложително. От една страна, защото възложителите започват да го изискват, а от друга, на фона на нарастващата конкуренция между самите преводачи. Поне към момента, производителността може да се увеличи само чрез наличието на по-добри технологии, например повече компютърна памет, по-голяма компютърна мощност, по-ефективни софтуерни инструменти (например разработване на потребителски интерфейс, мобилна

³³ Тул Бокс Джърнал - <http://www.internationalwriters.com/toolkit/current.html#LETTER.BLOCK48>

³⁴ Вероятно поради тази причина започват да се появяват и по-необичайни изследвания, като изследването на „любовните“ писма от преводачите към техните технологични другари, когато инструментите са полезни в работата, и писма за раздяла, когато не са (Руоконен и Коскинен 2017).

версия на инструментите за компютърноподпомогнат превод или по-широк и бърз достъп до документация във всички необходими формати). Същото констатира и Ч. Син-уи: „Независимо дали ни харесва, или не, технологичната революция в превода вече е тук. Никой не знае как е започнал преводът, нито какво е бъдещето му. Това, което знаем, е, че преводът е с нас от незапомнени времена; [...] това е една от най-старите професии в света. Каквото и да е, преводът вече не е това, което е бил и ще продължи да се променя според обстоятелствата“ (Син-уи 2017: 258). Според проучване от същата година преводачите се „примирият“ с присъствието на технологиите. Допълнителните проблеми, описани по-горе, в комбинация със страха от загуба на творчеството и изкривяването на представите за възможностите на технологиите, и най-вече на машинния превод, е това, което безпокои преводачите в най-голяма степен. За да стигне до това заключение, Л. Виеира използва автоматично извличане на ключови думи (конкретно съществителни имена) от дискусии във форуми за преводачи. По този метод очаквано се извличат като най-много думи, свързани с инструментите за компютърноподпомогнат превод, машинен превод и продуктивността (Виеира 2017).

Сравнително добрите резултати на невронния машинен превод отново пораждат нереалистични очаквания към машинния превод и твърденията, че в кратки срокове машинният превод ще замени преводачите. Според проучване на Б. Уанг и И. Пинг от 2020 г. асоциациите на преводачи и специализираните издания като АйТиАй (ITI)³⁵ и Мултилингвал (MultiLingual)³⁶ имат доста консервативна, макар и положителна позиция по отношение на приложенията на машинния превод, докато неспециализираните медии го определят като потенциална заплаха за преводачите. В проучването са разгледани само публикации на китайски медии, но те са съпоставими с публикациите в други медии по света. В неспециализираните издания се твърди, че е въпрос на време машинният превод да замени човека в ерата на изкуствения интелект, което естествено може да предизвика негативно отношение към новите технологии сред професионалните преводачи. Според специализираните издания обаче единствената възможна алтернатива е сътрудничеството между човека и машината. Например, използването на корпуси дава възможност за по-сложни методи за превод, като разпознаване на фрази, превеждане на идиоми и изолиране на противоречия (Уанг и Пинг 2020: 11). Въпреки това спорът дали машинният превод и инструментите за

³⁵ Институтът за транслейшън енд интерпретинг (Institute for Translation and Interpreting) <https://www.itl.org.uk/>

³⁶ <https://multilingual.com/>

компютърноподпомогнат превод могат да превърнат преводачите в обикновени редактори, което да ги направи по-малко важни, а в близко бъдеще и излишни, е друг аспект, свързан с технологичния напредък, който се дискутира често. „Използването на машинния превод винаги е било противоречиво, но вече се налага да бъде поставено в ползрението на много по-големия контекст на изкуствения интелект, най-вече заради невронния машинен превод“ (доклад на SDL³⁷ от 2020). Причините за отхвърлянето на новите технологии от страна на някои преводачи са различни: не е само поради страха от евентуална загуба на работа и професионален престиж, но и поради опасения за спад в качеството на текстовете и за начина на тяхното възприемане. От такива хора преводът се възприема като изкуство, което притежава собствени естетически критерии, които са свързани с творчеството и силата на въображението. Това се отнася най-вече за специфични видове превод, например на литературни текстове, където многозначността, конотацията и стилът играят важна роля (Чарчунеско и др. 2004). По време на форума Транслейтинг Юроп 2021³⁸ Крис Дурбан заявява, че „един качествен преводач трябва не само да бъде грамотен, но да бъде и добър писател“³⁹. През 2020 г. учени от Университета Помпеу Фабра разглеждат нагласите на преводачите, работещи с машинен превод. В своя анкета те питат 50 преводачи за тяхното мнение. 38% (или 19 от общо 50) смятат последващата редакция за по-изморителна дейност от превода. Според техните заключения преводачи, които редактират машинен превод за първи път, имат предразсъдъци и цялостно негативно отношение (към което се добавят и външни елементи като заплащането и бъдещето на професията на преводача). Въпреки това, след приключването на последващото редактиране, мненията им вече не са толкова негативни и повечето от тях признават, че биха извършвали тази дейност, въпреки че всички предпочитат да се занимават с превод (Видал и др. 2020: 65).

Някои учени (Кени 2012; О’Хаган 2012; Дохърти 2016; Кристенсен и др. 2017:) посочват наличието на взаимозависимост между човешкия и машинния превод, които до голяма степен се дължат на факта, че съвременният машинен превод се основава на възможността човешки преводи да се използват отново (например преводната памет от инструментите за компютърноподпомогнат превод, използвана за обучението на машинния превод) и резултатите от него често също се обработват от хора. Дори

³⁷ <https://www.sdl.com/software-and-services/translation-software/research/>

³⁸ https://ec.europa.eu/info/events/2021TEF_bg

³⁹ <https://chrisdurbanblog.com/about/>

интерфейсът на тези програми е съотносим с този на инструментите за компютърноподпомогнат превод.

И накрая, бъдещето на превода и технологиите за превод неизбежно се преплита с настоящото и бъдещето развитие на изкуствения интелект. Реалистично е да се очаква, че машинният превод ще продължи да улеснява определени типове превод и свързани с превода задачи, както вече се вижда от някои приложения за машинен превод и бизнес автоматизация (Олохам 2020: 578), например решенията на корпорации като Ербиенби или Ибей. Съвременният софтуер за машинен превод често позволява избор на различни области (например право, икономика, медицина, ботаника и т.н.). Целта е да се подобри качеството на резултата чрез ограничаване на допустимите значения на думите. Качеството може да се подобри и чрез човешка намеса. Например, някои системи са в състояние да превеждат по-точно, ако потребителят е определил еднозначно кои думи в текста представляват имена. В някои случаи дори може да се получи резултат, използваем без допълнително редактиране. Въпреки това необходимо е разбирането, че съвременните системи за машинен превод все още не са в състояние да конкурират хората

I.3. Терминологични бази

Терминологичните бази са централизирани бази от данни, които съдържат специфични за дадена тематична област или даден проект за превод термини. Обикновено тези термини са предварително одобрени (в много случаи самите клиенти ги съставят). Основното предимство на терминологичната база е, че прави процеса на превод по-бърз (особено ако той също е автоматизиран), защото спестява на преводачите нуждата да търсят необходимия термин. Както при преводната памет, програмата за компютърноподпомогнат превод сканира всеки нов сегмент, за да провери дали има съвпадение с терминологичната база. Този процес на работа гарантира също и високо ниво на последователност, защото един и същ термин се превежда по идентичен начин.

Поради значително количество от общи характеристики, терминологичната база се нарича още „речник“, „глосар“ или „лексикон“ (макар да се различават в доста отношения), защото като стандартния речник при използването на терминологична база се проверяват значенията, преводите и изписването на думите. Основната разлика между терминологичната база и речника е, че речникът не може да бъде променян. Друга разлика е форматът: терминологичната база обикновено е във формат, който може да бъде включен автоматично в софтуера за превод (.tbx), а речникът обикновено се използва във формати като Ексел или Уърд. Предимството на използването на тези

формати е, че те могат лесно да бъдат преобразувани във формати за терминологични бази. Детайлно описание на терминологичната база и на нейните характеристики е направено в следващата глава на този труд.

I.4 Приложение на многоезикови терминологични бази в системите за компютърноподпомогнат превод

На пръв поглед терминологичните бази са като речници – съдържат в себе си списък от термини. За разлика от тях обаче, терминологичните бази могат да бъдат конфигурирани. Потребителят (или собственикът) на базата има пълната свобода във всеки един момент да добавя, изтрива или модифицира термини. Друга характеристика на по-новите поколения терминологични бази е възможността за групова работа. Различни преводачи могат да извършват промени по едно и също време. Съществува и възможност да се класифицират термините в различни категории, например забранен термин; термин, който да бъде използван само в определен контекст, и др.

Добавянето на нови термини става по доста опростен начин. Едната възможност е чрез преводната памет. Преводачът маркира желания термин и с избор от менюто го включва в паметта. Ако структурата на преводната памет го позволява (значително количество от сегментите са термини), този процес може да се автоматизира. Друга възможност е автоматичното импортиране на предварително подготвени списъци с термини (най-често е таблица с формат в Ексел), които после се конвертират във формата на базата). Съществува възможността част от термините в базата да бъдат експортирани, за да се създаде нова база. Термините могат да имат няколко значения и следователно няколко превода. Терминологичните бази улесняват избора на преводача при селекцията на най-добрия термин в конкретния случай. Терминологичните бази позволяват съхранението на метаданни (клиент, област на употреба, бележка).

Многоезиковите терминологични бази следват същата структура и функционират по същия начин като двуезиковите. Това важи също и за избора на език, използването на метаданни, импортирането на данни в табличен формат. От техническа гледна точка няма съществена разлика между двуезиковите и многоезиковите терминологични бази. Редица учени обаче сигнализират за възможни усложнения по отношение на лексикалното значение на термините. А. Имре и Б. Барабас отбелязват, че в работата с многоезикова терминологична база броят на възможните езикови комбинации автоматично се увеличава от две на най-малко шест (в случай на три изходни езика) (Имре и Барабас 2020: 228). Например, една база с езикова комбинация английски-български е възможно да има само обратната посока: български-английски,

докато база от английски, български и италиански може да се комбинира, както следва: английски-италиански-български, български-английски-италиански, български-италиански-английски, италиански-български-английски и италиански-английски-български. Според същото проучване добавянето на повече от два езика води до неефективност, защото заради твърде многото възможни езикови комбинации, основното значение на определени термини може да се загуби в процеса на преминаване през няколко езикови посоки.

I.5. Нови функции на системите за компютърноподпомогнат превод

Към средата на 2020 г. в инструментите за компютърноподпомогнат превод се разработват нови функционалности за интеграция. МемоКю например интегрира допълнителни възможности за работа: Регекс асистънт (Regex Assistant⁴⁰ – функция с готови регулярни изрази, която спестява на преводачите нуждата сами да програмират), Антидот (Antidote⁴¹ – интеграция на автоматичния редактор за английски и френски Антидот коректор), Адобе Експириънс Мениджър (Adobe Experience Manager⁴² или АЕМ – цялостно решение за управление на съдържанието за изграждане на уебсайтове, мобилни приложения и формуляри), Джира (Jira⁴³ – един от най-популярните софтуери за проследяване на работата по даден проект), БиЛейзи (BeLazy⁴⁴ – платформа, която автоматизира получаването, управлението и доставката на проекти за превод), Гридли (Gridly⁴⁵ – позволява по-бърза актуализации и локализация на съдържанието на множество езици) и WS API⁴⁶ за интеграция с всяка платформа, поддържаща стандартни уебслужби, включително приложения, разработени в .NET или Java. Нито един от тези инструменти няма пряка връзка с превода. Става въпрос за инструменти за управление на задачи, файлови формати, обработка на информация и комуникация между различни специалисти.

Съществуват и редица безплатни приложения, външни на системите за компютърноподпомогнат превод, с които преводачите могат да улеснят работата си. Пример за такива безплатни приложения са: Клокифай (Clockify⁴⁷) – приложение за

⁴⁰ <https://docs.memoq.com/ggl-tst/Places/regex-assistant.html>

⁴¹ <https://docs.memoq.com/ggl-tst/Places/antidote-corrector.html>

⁴² Този вид интеграция позволява на потребителите да започнат и да управляват превода чрез интерфейса на АЕМ: <https://www.memoq.com/integrations/content-management/aem-integration>

⁴³ <https://www.memoq.com/integrations/project-management-systems/jira-integration>

⁴⁴ Интеграцията позволява извличането на проекти от сървъра на мемоКю, както и от тези на потребителите: <https://www.memoq.com/integrations/integration-with-belazy>

⁴⁵ <https://www.memoq.com/integrations/content-management/gridly>

⁴⁶ <https://docs.memoq.com/current/api-docs/wsapi/memoqservices/index.html>

⁴⁷ <https://clockify.me/>

проследяване на свободното време и на работните часове в проекти; Протемос (Protemos⁴⁸) – бесплатна система за управление на преводи, клиенти, проекти и финанси; Гугъл таскс (Google tasks⁴⁹) – бесплатна добавка за Джимейл за създаване на списъци със задачи по проекта и за напомняне.

В последните версии на някои от програмите за компютърноподпомогнат превод (от 2022 година) са добавени нови функционалности като диктуване (за преводачи, за които диктуването е по-удобно от писането), обработка на файлови формати като .pdf и подобрени OCR функционалности за преобразуване и разчитане на изображения.

„Относно изкуствения интелект, според разработчиците на Традос, той ще е следващата стъпка след невронния машинен превод. Например чрез анализ на съдържание може да се извлече информация въз основа на класификация по тематични области и ключови думи. Компанията прогнозира, че в бъдеще изкуственият интелект ще автоматизира все повече дейности, като ресурсите за превод се преразпределят много по-ефективно. Това може да помогне автоматично да се избере правилният работен поток, да се подобри обработката на сложни типове файлове или да се определи къде редакторите трябва да обръщат специално внимание на качеството на превода. Това би могло да премахне необходимостта от поддръжка на различни машини за превод за различни случаи на употреба, като се конфигурират такива въз основа на анализ на всеки конкретен случай“⁵⁰.

I.6. Използване на системите за компютърноподпомогнат превод в зависимост от лингвистичните характеристики на различните типове текстове

Поради функционалностите на системите за компютърноподпомогнат превод не е препоръчително всички видове текст да бъдат превеждани с тяхна помощ. Както бе споменато по-горе, всеки нов текст за превод бива разделен на отделни сегменти, които после преводната памет сканира и сравнява със сегментите, съхранени в нея. Съпоставянето между нови и стари текстове се използва, за да се определи дали някоя част от текста вече е преведена и е част от паметта. Този начин на работа прави много по-предпочитани документи със стандартизирано и еднотипно съдържание (каквито са

⁴⁸ <https://protemos.com/>

⁴⁹ <https://apps.apple.com/us/app/google-tasks-get-things-done/id1353634006>

⁵⁰ eBook: Going places with Trados https://www.trados.com/download/going-places-with-trados/202621/?utm_campaign=Trados%20monthly%20newsletters&utm_source=hs_email&utm_medium=email&utm_content=212011200&_hsenc=p2ANqtz-9wSWGqdfdyXJBqBGwEnLFN5otgQoaetozTZ441ckycBGhCtCJQIWVlINZGiZH4n7gRNHF7EAF0-SJrap120N0X7sQT_A

административните, правните, техническите и др.) и е по-малко вероятно да се наблюдава между два различни литературни текста.

Съществуват редица области, които позволяват системното ограничаване на изразните средства, които се използват за да се подобри автоматичната обработка на текстове или взаимодействието между хората и машините. В контролирания език тези подобрения се постигат чрез внимателно ограничаване на граматиката и речника за конкретна тематична област, с цел намаляване или премахване на неяснотата и сложността. Всяка от различните области има специфични изисквания за проектиране на контролирания език и дори в рамките на една област често съществуват конкуриращи се формално еквивалентни езикови конструкции и не винаги е ясно коя от тях е най-подходяща при работа. Без съмнение в това направление има място за сравнителна оценка, за да се определи кои езикови конструкции са най-подходящи за използване за определена потребителска група (Швитер 2015: 501). Механизмът, по който работят инструментите за компютърноподпомогнат превод, ги прави изключително подходящи за еднотипни текстове, защото повтарящите се езикови единици се превеждат значително по-бързо. Подобряването на скоростта на превода посредством технологията за съпоставяне, описана по-горе, спестява нуждата един и същ превод да бъде набран отново. Стандартизацията на текстовете помага и за точността на превода. Чрез преводната памет вече не се налага преводачите да помнят какво са превеждали, за да превеждат по еквивалентен начин дадени фрагменти в големи по обем текстове.

Естествено преводът на художествена литература не влиза в горепосочените характеристики. Стремежът да се създадат системи за превод, които да се справят с художествени текстове, не е нов. Подобни усилия засягат предимно машинния превод, но се правят и опити литературни текстове да се превеждат с помощта на системите за компютърноподпомаган превод, въпреки че в обичайния си вариант те са пригодени за текстове, в които се използва техническа, финансова или специфична за индустрията терминология. Чрез препращане към вътрешни или външни езикови бази от данни, интегрирани в инструментите за компютърноподпомогнат превод, може дадени нюанси на езиковата употреба да бъдат интерпретирани по-лесно и това да ускори превода.

Паралелно с традиционната форма на превод („предаване на съдържанието на устен или писмен текст от един на друг език“⁵¹) се развиват и други процеси, като локализация, интернационализация и творческия превод, отличаващи се с адаптиране на

⁵¹ <https://ibl.bas.bg/rbe/lang/bg/превод/>

оригиналното значение. **Локализацията** (Localization) е адаптирането на текст към конкретни физически, технически, езикови, културни, етични, религиозни, философски, търговски, маркетингови или други специфични особености на дадена държава. Най-общо казано, локализацията може да се опише като едновременното превеждане и адаптиране на даден текст към културата и традициите. Този процес може да бъде разделен на проблематика, свързана с езикови въпроси (linguistic issues), самото съдържание на текстовете, проблематика, свързана с различните култури (content and cultural issues), и технически въпроси (technical issues) (Тофтедал 2020: 135). Повечето хора неправилно използват термина „локализация“, за да опишат единствено превода на съдържанието. Всъщност много малка част от процеса по „прехвърляне“ от един език на друг би могла да се определи като превод (Гаудек 2007: 37 – 38). За да бъде достъпен даден продукт на много различни езици, той трябва да бъде локализиран чрез превод. Локализацията представлява не само превод, но и преработена спрямо различната култура версия на описанието на даден продукт и е от съществено значение за разширяването с нови пазари, тъй като достъпът до информация за даден продукт на разбираем език увеличава търсенето му. По този начин може да се каже, че преводът има значение за световната икономика: той „приспособява“ продукти, за да отговори на нуждите на местните пазари.

С термина **творчески превод** (transcreation – съчетание от думите translation и creation) се определя творческата адаптация на текст или елемент от творба, което се използва най-често с рекламна цел. Произходът на термина датира от периода 1960 – 1980 г., но се утвърждава в преводната индустрия след началото на новото хилядолетие. През 2010 г. Р. Рей и Н. Кели публикуват своя доклад *Достигане до нови пазари чрез транскреация: когато преводът просто не е достатъчен* (Reaching New Markets through Transcreation. When translation simply isn't enough), който допринася за установяването на творческия превод (Рей и Кели 2010). Характеристики на този тип текстове са: творческата адаптация на рекламното съдържание, промяната в оригиналните термини и значения (което има за цел въздействието върху целевата аудитория) и посланието на търговската марка (Карейра 2020: 4).

Въпреки наличието на подобни термини има учени (Сдобников 2018; 72; Тимко 2021), които не са съгласни с употребата им. Според тях, технологиите, използвани при превод, не променят същността на познавателната дейност на преводача и локализацията, творческия превод или „транскултуризацията“ (от. англ. transculturation)

и не трябва да се разглеждат като отделни видове превод, а са прояви на едно и също явление – на превода (Тимко 2021: 2).

II. Отношението на преводачите към технологиите за компютърноподпомогнат превод

II.1. Изисквания на системите за компютърноподпомогнат превод към уменията на преводачите

Някои хора смятат, че професионалният превод е само въпрос на „знание на чужди езици“ и че всеки, който е учил даден език в училище или притежава езиков сертификат, може да стане преводач. Трябва да се подчертае, че професионалните преводачи са висококвалифицирани експерти, както заради съдържанието, което превеждат, така и заради различните и все по-сложни инструменти и софтуер, които трябва да умеят да използват.

Преводачите и преводаческите услуги също са засегнати от промените, наложени от технологичното развитие. Все по-значимото присъствие на интернет прави необходимо и възможно постоянното обновяване на съдържание (текстове, графики, видео и т.н.). Това води до създаването на големи количества текстове, видеоматериали, презентации и др., които се нуждаят от превод. От друга страна, навлизането на новите технологии в превода носи със себе си както положителни, така и отрицателни последици, които на свой ред водят до противоречиви мнения, свързани с използването на новите технологии. Академичните среди започват да анализират различни аспекти от работните процеси, свързани с технологиите за превод и влиянието им върху преводачите. Някои от идентифицираните недостатъци при употребата на тези технологии са затруднението при определяне на контекста в резултат на сегментирането на текста от преводния софтуер (Биау-Гил и Пим 2006: 5, Чарчунеско и др. 2004: 8, Боукър 2005: 134) или целенасоченото опростяване на даден израз само за да се постигне по-честа повторна употреба (Чарчунеско 2004), която на свой ред се посочва като едно от основните предимства на тези инструменти. В основата на професионалните изисквания към преводачите, както и в основата на много учебни програми водеща роля имат две основни заинтересовани страни в превода и устния превод: професионалните организации на преводачите и големите работодатели. Тяхната визия за професионалната практика отдавна се налага при създаването на учебни програми, поне в Европа, където се приема, че всички студенти трябва да работят с поне два езика или че каноничната посока на превода е към родния език на преводача въпреки многото

доказателства за професионалисти с един чужд език, работещи и в двете посоки на превод. В частния сектор по-големите преводачески компании провеждат кратки курсове за въвеждане и развитие на персонала, за да допълнят университетските програми. Много от тях по същество обхващат технологичните аспекти на професионалния превод и локализация, области, в които се предполага, че университетите не подготвят достатъчно своите възпитаници. В публичния сектор международните организации като Европейския съюз имат интерес да насърчават специфични тенденции в обучението. Програмата за обучение на устни преводачи на Европейската комисия, започнала през 1964 г. и продължила до 1997 г., е добър пример за това. Ролята, изпълнявана от тази програма, сега се изпълнява от магистърската програма по устен превод (European Master's in Conference Interpreting или EMCI) като Генералната дирекция за устни преводи работи заедно с университетите, за да предложи помощ при обучението (Кели и Мартин 2020: 593 - 594).

Малко преводачи разглеждат (интернет) технологиите като необходимост: повишаването на производителността трябва да се направи на фона на нарастващата конкуренция, а производителността може да се увеличи само благодарение на по-добри технологии (повече пространство в паметта, по-голяма компютърна мощност и скорост, по-ефективни софтуерни инструменти, по-широк и бърз достъп до документация и т.н.). „Ако преводът е бил засегнат от използването на компютри, същото важи и за начина, по който се обучават преводачите. Обучението задължително трябва да включва използването на технологии“ (Боукър 2015: 88). Необходимостта от интегриране на обучението за използване на инструменти за компютърнопомогнат превод в програмите за обучение на преводачите на свой ред поражда множество предизвикателства и въпроси, например кои видове инструменти са подходящи, какво е необходимо да се научи за технологиите, как трябва да се преподава използването на компютърни помощни средства и кога трябва да се въведат технологиите в учебните програми. След овладяването на текстовата обработка (въпреки че най-вероятно се използва само част от нейния потенциал) преводачите трябва да се справят с материалите за превод, които може да са в различни формати или създадени с доста сложни софтуерни инструменти, което може да изисква инсталирането на софтуер. Много от преводачите се сблъскват с тези промени, без изобщо да са преминали подходящо обучение по компютърни науки и технологии.

Според Ч. Син-уи (2015), за да се съгласи да работи с машинен превод, даден преводач трябва да има контрол (controllability) и възможност за персонализация

(customizability) на системата, която на свой ред трябва да е съвместима с различни файлови формати (compatibility) и езикови изисквания и да предлага толкова добър превод (simulativity) или дори по-добър (emulativity) от превода на човек, като му позволява да работи заедно с други специалисти (collaborativity) за създаването на качествени преводи (productivity). Авторът идентифицира посочените седем основни понятия като особено важни при компютърноподпомогнатия превод (Син-уай 2015: 32). Преводачите (като потребители) са отдалечени от основния компютърен код и по този начин и от контрола върху технологиите, с които работят, и от ресурсите, с които разполагат. Например някои от преводните технологии вече се разработват в мобилни приложения за публикуване на машинен превод (Торес-Хостенч и др. 2017). Боукър (2015: 92) посочва, че „системите за машинен превод с времето се възприемат като още един инструмент за компютърноподпомогнат превод, който взаимодейства с преводача, например под формата на предварително или последващо редактиране“. Атила Имре ги оприличава на „триъгълник на превода“, който има три върха – машинен превод, програма за компютърноподпомогнат превод и човешки превод, като всяко едно от решенията се опитва да демонстрира превъзходството си над останалите (Имре 2015: 83).

Сравнително отскоро преводачите са изправени пред избора дали да се придържат единствено към превода на текста, или да инвестират в технологичното си развитие с надеждата да достигнат до по-изгодните пазари, свързани с информационните технологии. За всички преводачи това означава да правят все по-големи инвестиции, ако искат да отговорят на изискванията на своите доставчици и клиенти, за да могат да бъдат наречени „технологично ефективни“. Всеки, който не желае или не може да участва в надпреварата за овладяване на повече технологии, ще бъде принизен към позицията на „обикновен преводач на текст“, който вече не може да се стреми към добавената стойност, която се получава с помощта на сложен софтуер или обработка на компютърен код (Имре 2021: 11).

II.2. Профилът на преводача в съвременната епоха

Преводът с помощта на компютър не е същата дейност като традиционния превод, защото компютърните програми осигуряват обработка на текста, която е много по-разнообразна от линейното четене. Интернет, със своя универсален достъп до информация и незабавна комуникация между потребителите, води до физическата и географска „свобода“ за преводачите, немислима в миналото. За това обаче са необходими технически познания, които не всички преводачи желаят или могат да

придобият. Преводачите трябва да приемат новите технологии и да се научат как да ги използват максимално пълно като средство за повишаване на производителността и подобряването на качеството (Чарчунеско и др. 2004).

Д. Гаудек посочва, че интернет предоставя на преводача важна информация за пазарите, темата на изходния материал, наличните ресурси, хардуерни и софтуерни решения, като същевременно дава възможност на преводача да поддържа връзка с клиенти и потенциални специалисти (Гаудек 2007: 264). Използването на приложения „в облака“ спестява на преводача управлението, поддръжката и конфигурирането на инструментите за превод. Онлайн моделът на използване на споделени инструменти прави преводача член на международна виртуална общност, тъй като актуализациите и новите версии са достъпни незабавно и всички потребители се възползват от тях. Такива услуги в превода са насочени към професионални или случайни преводачи, както и към агенции, институции и компании (Гамбие 2019: 351). От друга страна, този начин на работа създава зависимост от интернет връзката, което е обвързано и с риска от възможни проблеми с нарушаването на сигурността и поверителността.

Профилът на преводача придобива нови измерения след налагането на технологиите. Необходимо е да се разграничат преводачите на художествени текстове от преводачите на специализирани текстове. Изискванията към последните се променят сериозно през последните години и продължават да се променят, като налагат работа с най-новите инструменти за компютърноподпомогнат превод като Традос или други инструменти и платформи. Преводачите изпълняват широк спектър от задачи, които включват, но не се ограничават до: превод, редактиране, преглед, употребата на езика, управление на терминология и много други.

Ясно е, че информационните технологии оказват голямо влияние върху професията на преводача и започват да създават разрив между тези преводачи, които са в състояние и желаят да ги използват пълноценно, и тези, които не искат или нямат възможността за това. Едната гледна точка е, че новите технологии предоставят добавена стойност към работата на преводача. Другата гледна точка е, че посредством новите технологии преводачите стават все по-зависими от преводаческите агенции, които решават дали да инвестират в тези технологии и обикновено са собственици на ресурсите за компютърноподпомогнат превод. Допълнително разграничение се получава между тези преводачи, които смятат за достатъчно използването на основна комбинация от инструменти за компютърноподпомогнат превод (състоящи се от текстов редактор, преводна памет и система за управление на терминологията, заедно с връзка

към интернет), и тези, които използват по-сложни услуги и обработват специфични видове материали с по-сложни софтуерни системи (приложения за субтитри, приложения за локализация, системи за управление на проекти за превод, обработка на изображения и работа с персонализиран клиентски софтуер). Последните имат по-добри възможности в сравнение с тези, чиито познания се ограничават само до инструментите за компютърнопомогнат превод (Гаудек 2007: 280).

Няколко практически проблема в професионалната практика на превода (включително локализацията, последващото редактиране на машинен превод, работата с терминология, управлението на проекти или други дейности) са свързани с нивото на познаване на инструментите за превод. Важно е да се определи кои умения са от значение само за конкретна задача и каква част от тях се припокрива с други подобни, но не идентични умения. Например, ако един специалист се развие като добър преводач, колко от неговите умения биха помогнали в работата му в последващото редактиране на машинен превод или в локализиране на текстове (Шрев 2020:153).

Заради новия вид работа тип поточна линия всички процеси или дейности, за които е възможно, биват частично или напълно автоматизирани. Необходима е работа с текстообработваща програма, управление на преводна памет, управление на терминологична база, овладяването на пълен обем от интернет функционалности, към които спадат уменията за търсене и намиране на нужната информация. Преводачите трябва да се придържат към този стандарт на работа, да могат да обработват изображения, разпознаване на глас, както и да могат да обработват файлове във формати като HTML, XML, XHTML и да се справят с други технологии, които ще се наложат при превода. Логично следствие от това е разработването на средства, които ускорят, улеснят и оптимизират процеса на работа. Стандартни задачи за превод в днешно време не са само преводите на текст, а и създаването на друг вид съдържание като добавянето на субтитри (което налага технически познания за работа със съответната програма), превода на съдържание от изображения или работата по интернет страници (необходимо е познание за това как се създават сайтове и как се борави с тях).

II.3. Нови професии, създадени заради технологиите за компютърнопомогнат превод

Всичко описано дотук показва и друг аспект: новите технологии са причина за създаването на множество нови професии. Машинният превод е един от основните двигатели за промяна в технологията на превода. С него идва и последващото му редактиране, сравнително скорошно допълнение към списъка със задачи, които

преводачите трябва да изпълняват. Относително ново е, че преводачите са склонни все повече да се идентифицират с професионална група въз основа на специализацията си, до степен да създадат свои собствени организации, асоциации или форуми. Такъв е случаят с медийните преводачи (media translators), които обхващат дейности, свързани с въвеждането на субтитри (subtitlers) и дублиращи автори (dubbing authors). Преводачът терминолог е друг профил на преводач, който добива популярност (Гаудек 2007: 111). За това допринася и разпространението на софтуера за управление на терминология и по-голямото значение на терминологичната съгласуваност.

Умението да се превежда се състои от много и различни „подумения“ (sub-skills). В случай че някое от тях липсва, неизбежна е появата на проблеми в процеса на превод. Преводаческата компетентност е свързана, от една страна, с нивото на професионализъм на преводача. Например професионализмът на преводача може да бъде оценен по отношение на качеството на превода, изготвен в рамките на определен срок, способността на преводача да поддържа както качество, така и производителност, да работи съвместно с други експерти и т.н. (Али 2017: 50).

Изследването на знанията и уменията на преводачите не предизвиква интерес отскоро. Още преди две десетилетия групата за придобиване на преводачески компетенции (Process in the Acquisition of Translation Competence and Evaluation или РАСТЕ)⁵² излиза със становището, че преводаческите компетентности могат да се определят като „система от знания, способности и нагласи, необходими за извършването на превод: система от компетенции, които си взаимодействат, йерархични са и подлежат на вариации“. Няма обаче съгласие относно това от какво се състоят преводаческите компетентности, нито консенсус относно това къде трябва да се очертаят границите между подкомпетентностите. Групата за придобиване на преводачески компетенции изгражда първия си модел за компетентност в областта на превода през 1998 г. въз основа на емпирични проучвания. Прави се разлика между компетентност (основната система от знания) и ефективност (превод). Според тях двуезиковата компетентност е само един от компонентите на преводаческата компетентност и всички останали видове компетентност са взаимосвързани. Нещо повече, преводаческата компетентност се смята за експертно познание, което е предимно процедурно. По-късно в модела са включени стратегически и психофизиологични компоненти. Според тях преводаческата компетентност се различава от двуезиковата компетентност, която е основната система

⁵² <http://www3.uji.es/~aferna/EA0921/3b-Translation-competence-model.pdf>

от знания, необходими за превод. Преводаческата компетентност е експертно знание, което включва както декларативни, така и процесуални знания. Тя обхваща подкомпетентности като езикова подкомпетентност (отличното познаване както на изходния, така и на целевия език), екстралингвистична подкомпетентност, инструментална / професионална подкомпетентност, подкомпетентност за прехвърляне на съдържание (способността на преводача да разбере изходния текст и да го изрази повторно на целевия език, като вземе предвид целта на превода и характеристиките на четящите) и стратегическа подкомпетентност (Кобан 2015: 709).

Обхватът от подходи (Пим 2003), които отчитат компетентностите, свързани с извършването на преводаческа дейност, се предлага от експертната група на Европейската магистърска програма по превод (European Master's in Translation или ЕМТ) през 2009. Тези компетентности включват подкомпетенции, които не са специфични за превода, но са необходими и в други професионални области като обслужване на клиенти, и са представени в „Компетентност за предоставяне на преводачески услуги“ на модела за превод на Европейската магистърска програма по превод.

Съществуват и други фактори, които трябва да се имат предвид при анализа на преводаческата компетентност. Ако приемем, че преводът е продукт и услуга, то тогава преводачите на свободна практика се възприемат като доставчици на услуги или като (проектни) мениджъри по време на процеса на превода. Следователно има случаи, в които се налага специализация, а в други се изисква паралелно владение на всички умения, описани по-горе. Например, преводаческите компании разполагат с персонал, а преводачът на свободна практика трябва да се справя сам с всички задачи. В това число влиза работата с финанси, реклама и управление на клиенти (Имре 2020: 17). „Чист“ превод е, когато преводачът се занимава единствено с прехвърляне на съдържанието от езика източник на целевия език (и култура). Това означава, че преводачът просто извършва стандартните дейности, съпътстващи превода: търсене на информация, извличане на данни, търсене на подходящи термини, превод и редакция. Преводачите правят „чист“ превод, когато работят в екип, където други специалисти извършват останалите специализирани дейности. Съществуват и разширени услуги, при които преводачите могат да разработват уебсайтове или софтуерни пакети, да пишат скриптове, да подреждат субтитри, да подравняват нов превод със съществуваща преводна памет, да създават речници, да „клонират“ уебсайтове след превод на съдържанието им, да подготвят преводи за печат или за публикуване в интернет и много

други. В такива случаи предоставената услуга е „разширена“ или „подобнена“ и организацията на работа на преводача е различна, защото включва използване на по-широк набор от инструменти (Гаудек 2007: 104). Поради огромния брой от необходими знания, умения и качества преводачите всъщност са експерти и в области като комуникация, информационните технологии и др.

Езиковите познания са от съществено значение, но в много случаи са недостатъчни в днешно време. Техническите, юридическите, счетоводните и маркетинговите познания, съчетани с пълно разбиране на тематичната област, в рамките на която се превежда, също може да са наложителни.

Както се прогнозира от Д. Гаудек: поради взаимосвързаността никой преводач не може да се надява да оцелее без способността да общува ефективно и гладко – както професионално, така и лично – с многобройни партньори: клиенти, колеги, доставчици на информационни технологии и терминология, редактори, работодатели, доставчици на интернет услуги и много други. Доброто разбиране на маркетинга, управлението на проекти и счетоводството също са преимущество (Гаудек 2007: XIII).

Преводачите трябва да са компетентни в много области, част от които са описани по-горе. В това число влиза и доставката на преведения материал, която на свой ред също е свързана с допълнителна работа. В зависимост от ситуацията може да се наложи на преводача да положи големи грижи за връщането на файла за превод и преводната памет, както и да използва автоматичните функционалности за качество (те ще бъдат описани по-подробно в последната глава на този труд).

Въпреки всичко описано дотук има някои общи аспекти на преводаческата работа, които могат да бъдат валидни за всички. П. Нюмарк в своята книга *Учебник по превод* (A Textbook in Translation) от 1988 г. заявява: „Винаги е възможен задоволителен превод, но добрият преводач никога не е доволен от него. Обикновено може да го подобри. Няма перфектен, идеален или „правилен“ превод. Преводачът винаги се опитва да разшири знанията си и да подобри изразните си средства; той винаги преследва факти и думи. Преводачът работи на четири нива: преводът е първо наука, която включва познаването и проверката на фактите и на начина, по който ги описва – на това равнище може да се идентифицират фактически грешки; второ, това е умение, което изисква подходящ език и правилна употреба; трето, това е изкуство, което отличава добрия, творческия, интуитивния, понякога вдъхновен превод от безличното предаване на съдържание на друг език; и накрая, това е въпрос на вкус и разнообразието от качествени преводи е отражение на индивидуалните различия на преводачите“ (Нюмарк 1988: 6).

II.4. Изисквания на преводачите към функционалностите на системите за компютърноподпомогнат превод

Един от начините, използвани през годините, за да се проверят нагласите и мненията на преводачите, са допитванията и анкетите. Целта им е да се разбере как се чувстват преводачите, докато използват инструментите за компютърноподпомогнат превод, дали ги смятат за полезни, какви нови функционалности биха искали да им бъдат добавени, дали използват традиционни средства като отпечатани речници и др. Технологиите до такава степен са станали неизменна част от работата на преводачите, че дори и когато биват анкетирани за други цели, темата за технологиите обикновено също се включва. Проучването⁵³, проведено през 2017 г. съвместно от Представителството на Европейската комисия в Обединеното кралство, Представителния институт по лингвистика (Chartered Institute of Linguists) и Института по писмен и устен превод (Institute of Translation and Interpreting), се фокусира върху технологиите при повече от половината от поставените въпроси. В допълнение към въпросите за най-широко използваните инструменти при превода в проучването се търси информация колко важни са специфичните инструменти за работа, дали технологиите дават конкурентно предимство на преводачите и обратно, дали технологиите могат да заместят преводачите в бъдеще (Олохам 2020: 574 - 576).

Според проучване от 2016 г. сред 63 студенти по английска филология в Педагогическия университет в Краков относно ресурсите, които използват по време на превод, 30% от респондентите казват, че никога не използват традиционни речници. От останалите 70%, които използват речници, 46% го правят само от време на време, 16% избират отговора „обикновено“, а само 8% признават, че винаги ги използват. Що се отнася до онлайн речниците, те се използват от 100% от анкетираните, от които 75% са избрали „винаги“ и 25% „обикновено“. Причината е, че традиционните речници остаряват с оглед на бързото развитие на науката и технологиите. Високите цени на специализираните речници също са възпиращ фактор за тях (Кодура 2016). Потребителите предпочитат електронните версии на речниците. Според проучване от 2017 г. сред 50 студенти на възраст между 24 и 40 години от Магистърската програма по превод в Ислямския университет в Азад, базирано на модела Блу, няма значителна разлика между преводите, направени с помощта на онлайн речници, и тези с печатни

⁵³<https://www.ciol.org.uk/report-uk-translator-survey-2016-17>
<http://www.ciol.org.uk/sites/default/files/UKTS2016-Final-Report-Web.pdf>

издания по отношение на прецизност и плавност на превода. Разликата е значителна само по отношение на скоростта (Зарей и Фумани 2017). Проучване от същата година на Катедрата по английски език в университета Джадара в Йордания разделя 40 студенти (предварително избрани по подобни умения и знания) в две групи спрямо използваните ресурси за превод, едната група използва мобилни приложения на своите телефони, а другата – традиционни речници. Изследователите не откриват сериозни разлики в представянето на двете групи. Разлики се наблюдават между студентите, които са обучавани за работа с технологии и устройства, които ги използват, и тези, които не са преминали такова обучение (Алшехаб 2017).

Според проучване на американската компания Сиесей⁵⁴ от 2020 г. за работата на 7 363 преводачи малко повече от една трета (37%) от тези, които използват машинен превод, смятат, че общото качество е добро. 81% забелязват сериозни вариации в качеството на машинния превод при текстове, необработени предварително от човек (например елиминиране на многозначността, заместване на фразеологизмите и др.). Преводачите, които използват машинен превод, обикновено предпочитат да работят с адаптивни системи, каквато е Лилт⁵⁵ (заради цялостното по-добро качество, гарантирано от персонализирането на съдържанието), вместо такива с необработен машинен превод (71%), каквито са свободно достъпните платформи като Гугъл Транслейт. Преводачите оценяват приноса на речниците към качествените резултати (91% от анкетираните предоставят по-добро качество, когато използват речници, а 76 %, когато използват преводна памет). Интересно е, че почти една четвърт от преводачите (23%) твърдят, че предоставят по-добро качество, използвайки машинен превод – като опровержение на общоприетото мнение, че крайният продукт на машинния превод е по-лош в сравнение с човешкия превод. Преводачите все още оценяват използването на машинния превод като забавяне на процеса на превода в сравнение с работата с другите инструменти, което най-вероятно е обвързано с факта, че резултатът на машинния превод не е достатъчно добър и трябва да се редактира.

Що се отнася до инструментите за компютърноподпомогнат превод, според проучване на Комън сенс адвайзъри (Common Sense Advisory⁵⁶ или CSA) от 2020 г. 66% от 7 363 преводачи използват преводна памет и инструменти за

⁵⁴ Сиесей Рисърч (CSA Research) – независима американска компания за маркетингови проучвания. <https://csa-research.com/>

⁵⁵ Lilt <https://lilt.com/>

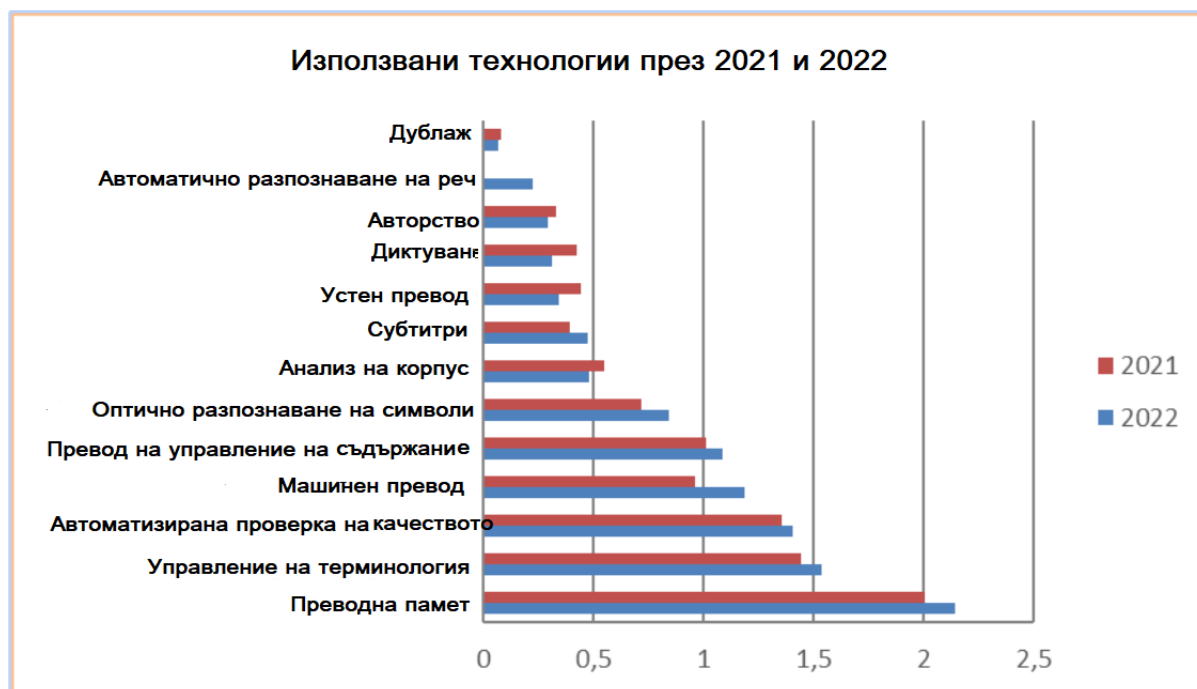
⁵⁶ <https://csa-research.com/>

компютърноподпомогнат превод, 60% използват инструменти за проверка на качеството и 48% използват системи за управление на терминологията. Към последните принадлежат терминологичните бази, вградени в инструментите за компютърноподпомогнат превод. Употребата на този вид програми е наложителна, ако преводачите искат да останат конкурентноспособни на пазара. Според проучване на 300 новопостъпили студенти от университета Ван Ланг (51,7% от тях принадлежат към Факултета по чужди езици) 13,7% от всички студенти съобщават, че използват речници почти всеки ден за проверка и изучаване на нови думи; 7,3% казват, че понякога използват речник; 79% признават, че ги използват само когато имат нужда, например за подготовка за изпити или когато са в специални ситуации (Ан 2021: 35). На въпроса дали предпочитат електронния или хартиения вариант на речниците, повечето от студентите избират електронните речници, защото са удобни и полезни, например при подобряване на произношението. В проучване на Кнежевич, публикувано през същата година, 705 студенти отговарят на въпросник от 51 точки с 4-степенна скала за използване (1 до 2 = ниска честота, 2 до 3 = умерена честота и 3 до 4 = висока честота). Най-ниската стойност (1,51) „се регистрира при използването на специализирани едоезикови речници [...], както и при използването на общи едоезикови речници“ (Кнежевич и др. 2021: 4). Най-високата средна стойност е 3,12 и се отнася до онлайн речниците, „което предполага, че те са най-широко използваният тип речник сред студентите“.

Всичко казано по-горе се потвърждава от годишния доклад *Проучване на европейската езикова индустрия* (European Language Industry Survey или ELIS⁵⁷) за 2022 г. В него са взели участие 1 342 специалисти (разделени в 264 фирми за езикови услуги, 745 независими езикови специалисти, 263 представители на обучаващи институции и студенти и 70 фирми, възлагащи езикови услуги). Според констатациите: „значително по-висок процент от независимите професионалисти съобщават, че технологичното обучение, което получават от създателите на инструменти, подпомагащи превода по различни начини, е достатъчно, за да се справят успешно (48%, в сравнение с 40% през 2021 г.), но все още има 21% сред анкетираните, които не са съгласни с това заключение, което е повече в сравнение с резултата през 2018 г. – 18%“. Машинният превод показва най-висок темп на растеж по отношение на използването си, но все още е далеч зад преводната памет. По-малко от 10% от анкетираните планират да инвестират в

⁵⁷ <https://elis-survey.org/>

закупуването на технологии, които подпомагат превода, с изключение на машинния превод (11%).



Фигура 4: Технологии за превод, използвани през 2021 и 2022 според *Проучване на европейската езикова индустрия*

II.5. Проучване сред преводачите в България

Поради големия брой проучвания, които изследват пазара за преводи и предпочитанията на преводачите на международно ниво, беше направено подобно проучване, което е фокусирано върху България. Тъй като в много от проучванията от този вид участници са студенти в магистърска програма по превод, тази анкета беше споделена само в два канала: с групата във Фейсбук „Помощ при превод“⁵⁸, със 7500 членове, които са професионални преводачи (най-голямата група, посветена на професионалния превод от и на български, другите групи са под 4000 членове); и в уебсайта Proz.com⁵⁹, смятан за най-голямата платформа за професионални преводачи в света. Анкетата е създадена посредством Гугъл форма и е споделена с връзка. Планирано е да не отнема повече от 7 до 10 минути за отговор на въпросите. Състои се от 26 въпроса: 16 са затворени, 6 позволяват повече от един избор и 4 са отворени въпроси, които са незадължителни и позволяват на преводачите да отговорят свободно⁶⁰. Всички

⁵⁸ <https://bg-bg.facebook.com/groups/pomosht.pri.prevod/>

⁵⁹ <https://www.proz.com/>

⁶⁰ <https://docs.google.com/forms/d/1YkIXezcuBVfEId9o7DqP6OrdPqFtZ6nzzZZkSRwMpPw/edit#responses>

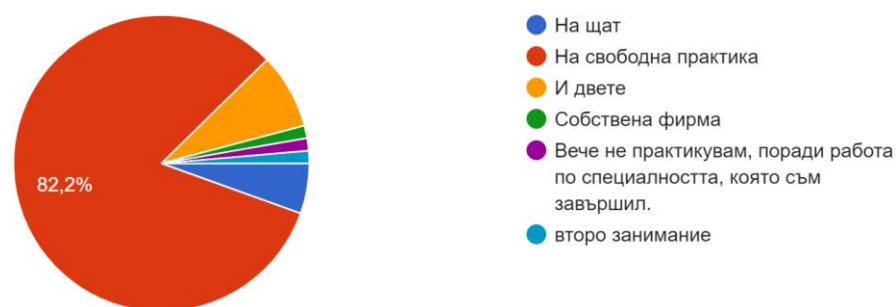
въпроси са фокусирани върху използването на инструментите за компютърноподпомогнат превод, използването на преводна памет, на терминологичните бази и на машинния превод. Въпросите, свързани с инструментите за компютърноподпомогнат превод, имат за цел да проучат дали преводачите в България ги използват, ако да – кои точно инструменти, как ги оценяват, какво смятат, че ще ги подобри.

Анкетата беше проведена в периода 01.02 – 31.03.2022 г. и в нея участваха общо 73 професионални преводачи. Цялата анкета е представена в Приложение 1.

Само четири въпроса не са свързани с технологиите за превод: образованието на преводачите и дали са на свободна практика или служители в преводаческа агенция или друга институция. Преводачите също са помолени да посочат своите изходни/целеви езици.

Въпросите, свързани с образованието на преводачите, начина им на работа и езиците от и на които превеждат, целят да дадат обща картина на ситуацията в България в началото на 2022 г. 89 % от преводачите имат висше образование. На въпроса „Каква е Вашата професионална заетост?“, 82,2 % отговарят „На свободна практика“ (Фигура 5). Според дадените отговори (доколкото изследването може да се приемем за представително) броят на преводачите на свободна практика в България е голям. Тъй като клиентите вероятно достигат до преводачите на свободна практика посредством преводачески агенции, това означава, че пазарът е, от една страна, голям, от друга страна, преводаческите агенции не се ангажират да наемат много преводачи едновременно.

Каква е Вашата професионална заетост?
73 отговора



Фигура 5: Професионалната заетост на участващите в анкетата

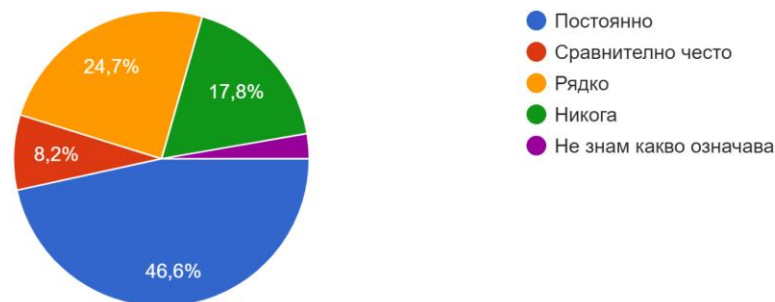
Интересна информация предоставят отговорите на въпросите от какъв и на какъв език превеждат преводачите. Очаквано, езикът, от който се превежда най-много, е английският, след него се нареждат българският, следват немският, руският и френският. От испански, италиански, гръцки, шведски, полски и хърватски също превеждат относително голям брой от преводачите. В обратната посока на първо място е българският, следван от английският, немският, френският, руският, гръцкият. Има езици, които не са посочени от преводачите и в двете посоки: фински, чешки, естонски, унгарски, латвийски, литовски, малтийски, словашки, ирландски, румънски. Все пак преводи от и на чешки, словашки, румънски (а и на останалите от изброените езици) се правят, което показва, че анкетата е достигнала до определен кръг от преводачи, които са активни в социалните мрежи, докато други преводачи, които владеят и използват професионално посочените езици, не са участвали в анкетата.

Над половината от анкетираните (54%) превеждат над 1000 страници годишно, което означава грубо по над 3 страници на ден. В съвременните условия производителността на труда на преводачите се увеличава и три страници превод на ден означава, че вероятно анкетираните се занимават и с други дейности и професионалният превод е допълнителна заетост (което също очертава параметрите на пазара в България).

Най-голям процент от анкетираните са отговорили, че превеждат специализирани текстове (90,4 %), следват тези, които превеждат административни документи (69,9 %), художествена литература (23,3 %) и друг тип документи (31,5 %). С други думи, по-голяма част от групата на респондентите се занимава с документи, които са подходящи за компютърноподпомогнат превод. По тази причина не са изненадващи резултатите, че 46,5 % от преводачите използват системи за компютърноподпомогнат превод постоянно, а 8,2 % – сравнително често. От друга страна обаче, почти половината от анкетираните използват системите за компютърноподпомогнат превод рядко или никога (Фигура 6).

Използвате ли инструменти за компютърноподпомогнат превод?

73 отговора

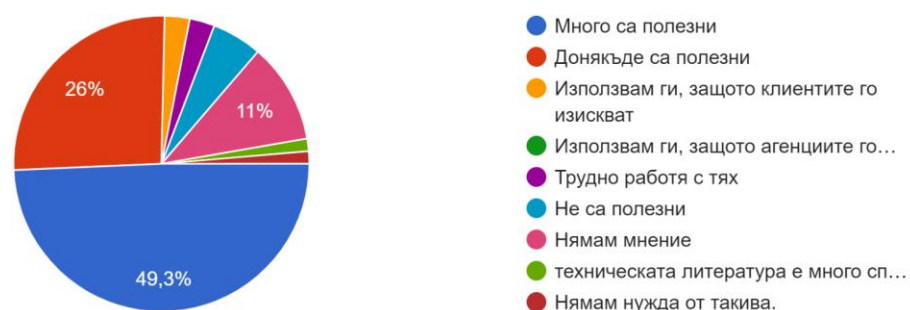


Фигура 6: Честота на използване на инструментите за компютърноподпомогнат превод

Най-популярните системи за компютърноподпомогнат превод очаквано са Традос, Мемсоурс, мемоКю, Уърдфаст, СمارтКАТ. Почти половината от анкетиранияте оценяват системите за компютърноподпомогнат превод като много полезни (49,3 %), а над една четвърт – като донякъде полезни (26 %) (Фигура 7). Около една четвърт от преводачите нямат мнение или смятат, че системите за компютърноподпомогнат превод не са полезни. От тези числа може да се направи изводът, че системите за компютърноподпомогнат превод се използват от българските преводачи.

Как оценявате употребата на инструментите за компютърноподпомогнат превод?

73 отговора



Фигура 7: Оценяване на инструментите за компютърноподпомогнат превод

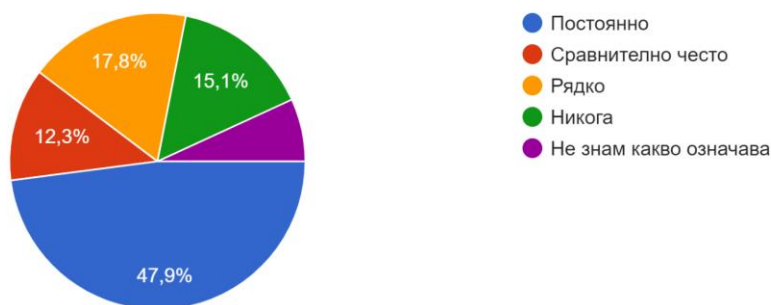
На въпроса „Как според Вас трябва да се развият и подобрят инструментите за компютърноподпомогнат превод?“ са отговорили общо 55 от 73 участници в анкетата. Тук са поместени част от получените отговори: „Да имат адекватни инструкции и на български език. Не всички преводачи ползват английски“; „Обогатяване на речника и

включване на идиоматични изрази, да се работи за по-добра граматика и лексикална прецизност“; „Необходима е по-добра работа с преводачески памет и терминологични бази; възможност за директно търсене на термини напр. в IATE“; „Да се обогати базата данни, за да бъдат повече двойки езици пълноценни помощници при превод“; „По-прецизна проверка на правописа, добър вариант би бил, ако е възможно да се интегрира диктуване на текст“; „Трябва да предоставят подобрени възможности за проверка на качеството“; „На първо място, да могат да се превеждат pdf файлове или да има вграден конвертор“; „Да има по-големи възможности за визуализация, редактиране и форматиране на текста на целевия език и за сравнението му с изходния текст, без да се налага използване на MS Word. Подобряване на възможностите за обмен и добавяне на преводачески памет“; „Усъвършенстване на автоматизирания превод. Осигуряване на по-големи облачни пространства за съхранение, гарантиращи конфиденциалност на информацията“; „Да бъдат по-модуларни, т.е., по-леки и съдържащи само основни функции и останалите им функции да се добавят поотделно като плъгин, както в Градос, която въпреки това е прекалено тежка програма“. Отговорите показват, че (интуитивно или не) преводачите имат реалистични очаквания за развитието на системите за компютърнопомогнат превод, а също така – доста добре очертани изисквания към функционалностите на тези системи, което (без съмнение) се дължи на опита им за работа с тях.

На въпроса дали използват преводна памет, 47,9% от анкетираните са отговорили, че използват постоянно, а 12,3% – сравнително често, това дава над 60% общо в полза на използването на преводна памет (Фигура 8).

Използвате ли преводна памет?

73 отговора

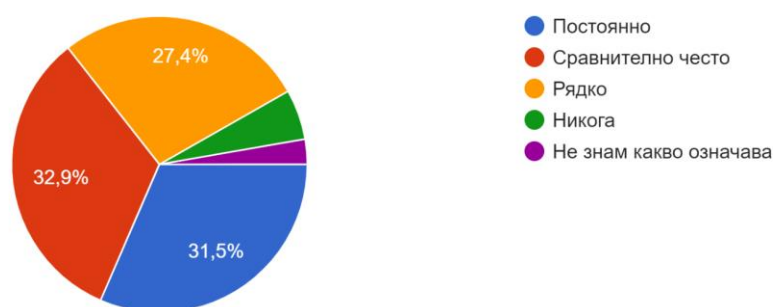


Фигура 8: Отговори на преводачите относно употребата на преводна памет

Отговорите на въпроса дали преводачите използват терминологични бази, са съпоставими с отговорите на въпроса дали използват преводна памет: 31,5% използват терминологични бази постоянно, а 32,9% – сравнително често, отново над 60% (Фигура 9). От това може да се заключи, че терминологичните бази се използват, което препраща към свободните отговори за употребата на системите за компютърнопомогнат превод, според които някои преводачи искат да имат връзка за търсене на термини в терминологични хранилища (като ИАТЕ). Липсата на подходящи терминологични ресурси е още един повод за намирането на лесен, универсален и достъпен начин за създаването на терминологични бази.

Използвате ли терминологични бази?

73 отговора

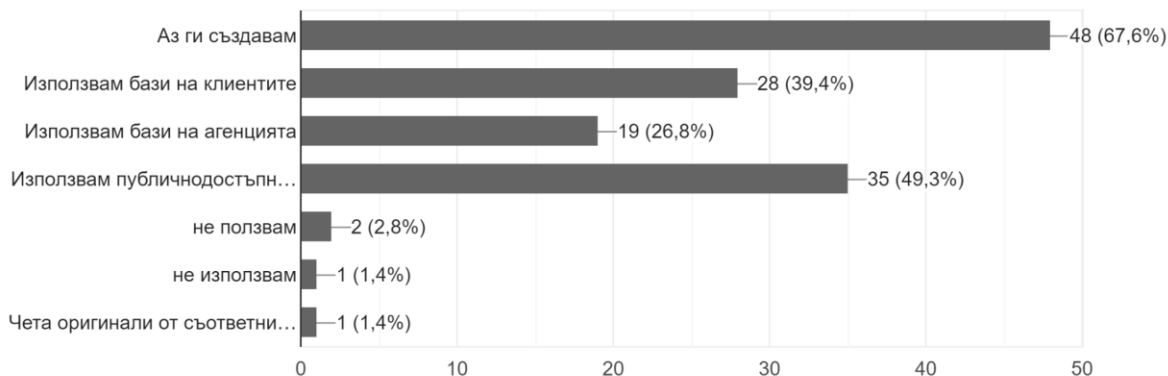


Фигура 9: Отговори на преводачите относно употребата на терминологичните бази

67,8% от използващите терминологични бази ги създават сами, 39,4% използват бази на клиентите, 26,8% – на агенциите, а 49,3% използват публичнодостъпни терминологични бази (сборът е повече от 100%, тъй като отговарящите могат да изберат повече от един отговор: Фигура 10).

Ако използвате терминологични бази, какъв е техният произход?

71 отговора

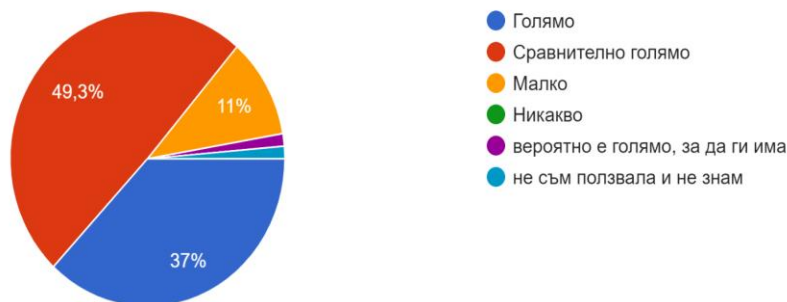


Фигура 10: Произход на терминологичните бази

На въпроса дали използват публичнодостъпни бази, 20,5% от анкетираните отговарят положително, 20,5% ги използват сравнително често, 32,9% – рядко и 26% никога не ги използват. Като най-използваните публичнодостъпни терминологични бази са посочени ИАТЕ и Глосбе. Може да се направи заключението, че сравнително голям процент (41%) от участниците в анкетата използват публичнодостъпни бази, възможно е, ако има по-широка популяризация на предимствата на публичнодостъпните терминологични бази, тяхната употреба да се увеличи. Тук трябва да се отчете и фактът, че в много случаи преводачите се нуждаят от специализирани терминологични бази, които не са отразени в посочените публични източници. Последното е в подкрепа на възможността сами да ги създават.

Преобладава мнението сред участниците, че значението на терминологичните бази за качеството на превода е голямо, на въпроса „Според Вас какво е значението на терминологичните бази за качеството на превода?“ 43% са отговорили сравнително голямо и 37% – голямо (Фигура 11). От тези отговори може да се заключи, че разработването на унифицирана методика за създаването на терминологични бази може да бъде от полза за ежедневната работа на преводачите.

Според Вас какво е значението на терминологичните бази за качеството на превода?
73 отговора



Фигура 11: Значение на терминологичните бази за качеството на превода според преводачите

На въпроса дали използват двуезикови или многоезикови терминологични бази, 20,5% от отговорилите казват, че използват многоезикови, 37% използват двуезикови, 24,7% използват и двата вида бази; 17,8% не използват нито един от двата вида бази. Отговорите показват, че броят на съвременните преводачи, които не използват дву- или многоезикови терминологични бази е относително малък. Обратно, по-голямата част от анкетираните използват двуезикови или многоезикови терминологични бази. Това също потвърждава нуждата от методика, която да позволява създаването както на дву-, така и на многоезикови терминологични бази.

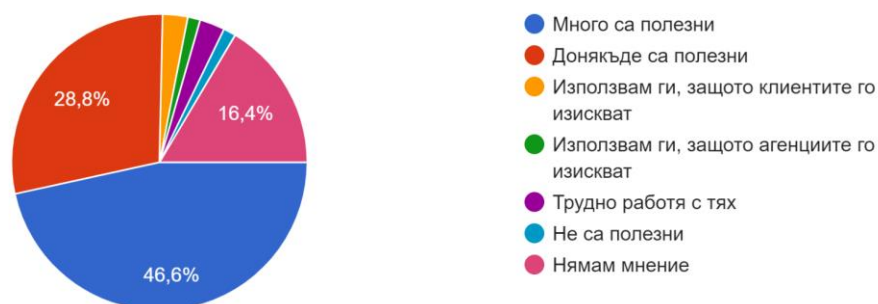
Беше включен въпросът от колко езика се състоят терминологичните бази, които преводачите използват (при условие, че ги включват в работата си). На този незадължителен въпрос отговарят 56 души: 55,4% ги използват за три езика, 8,9% за четири езика, 1,8% за пет езика и 33,9% за повече от пет езика. Отговорът на този въпрос до голяма степен е свързан с използването на публичнодостъпните терминологични бази, защото най-големите и известни са многоезикови.

От друга страна, 34 души от 56-те участници доуточняват в полетата за свободни отговори, че сами създават файловете си с многоезикови терминологични бази. Това се случва чрез запазване на преводите в инструментите за компютърнопомогнат превод или във файл (като Уърд, Ексел), МултиТърм на Традос, Ноутпад и др. Някои използват и речников подход, като си записват преведените термини, в случай че отново се сблъскат с тях.

Работата с терминологични бази е оценена като много полезна от 46,6% от участниците в допитването, а 28,8% ги оценяват като донякъде полезни (Фигура 12).

Как оценявате работата с терминологични бази?

73 отговора



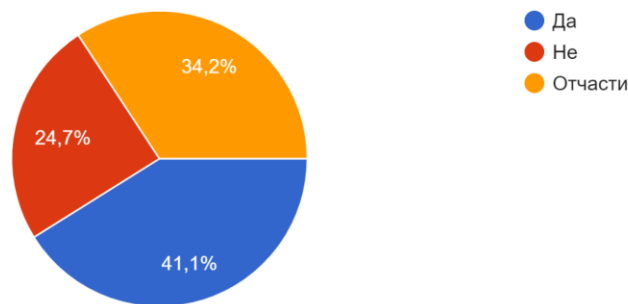
Фигура 12: Как преводачите оценяват работата с терминологичните бази

Голямо разнообразие от отговори има на въпроса „Как проверявате качеството на превода?“. Само 17,8% отговарят, че използват функционалностите за проверка на качеството на инструментите за компютърнопомогнат превод; 9,6% си служат с външни програми, 61,6% (най-голямата група) повторно проверяват текста сами. Останалите отговори съставляват 1,4% и са различни за всеки отговорил: в зависимост от указанията на заявката на превода, консултация със специалисти в областта, комбинация от вградените инструменти за проверка на качеството и повторен преглед на текста, повторно преглеждане в комбинация с външна програма, чрез повторно преглеждане с инструмент, проверяващ за граматически грешки. Анализът показва, че на този етап от работата при създаване на превод съвременните технологии се използват в малка степен и все още преводачите разчитат основно на собствените си умения за корекция и редакция.

Преводачите са попитани също дали използват инструменти за локализация. 9,6% ги използват постоянно, 9,6% сравнително често, 20,5% рядко, 26% никога не ги използват, а 34,2% твърдят, че не познават такива. Почти равномерно се разпределят отговорите на въпроса дали преводачите познават стандартите за превод: 41,1% ги познават, 34,2% отчасти и 24,7% – не (Фигура 13).

Запознати ли сте с ISO стандартите за качество на превода?

73 отговора



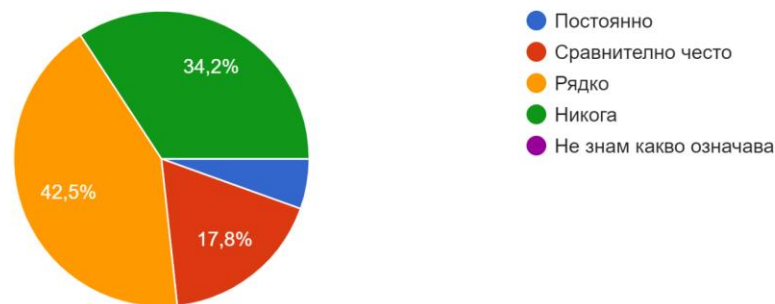
Фигура 13: Познаване на международните стандарти за качество на превода

Интересни са резултатите при отговора на въпроса дали преводачите използват машинен превод – преобладават отговорите, че не се използва или че се използва рядко: 42,5% го използват рядко, а 34,2% – никога. Само 5,5% от преводачите посочват, че използват машинен превод постоянно (Фигура 14). Отговорите може да се дължат на две причини: все още незадоволителното качество на машинния превод от различни езици на български в различни тематични области и на художествени текстове, но и възможно нежелание да се признае, че се използват технологии, които се очаква (ако не да заместят изцяло) до голяма степен да облекчат работата на преводачите.

На въпроса какви приложения използват преводачите, които си служат с машинен превод, са отговорили 55 от 73 участници. На този въпрос беше възможно да се даде повече от един отговор. 40% използват вградените опции в системите за компютърноподпомогнат превод, 65,6% използват свободностъпни програми в интернет. 9,1% също използват програми от интернет, но платени. Останалите отговорили казват, че не използват машинен превод в работата си.

Използвате ли машинен превод?

73 отговора



Фигура 14: Честотата на употреба на машинния превод от страна на българските преводачи

Последният въпрос в анкетата е относно това как според преводачите трябва да се развие и подобри машинният превод. На този въпрос са отговорили 56 от 73 участващи. Отговорите им варират от неутралност до предложения за подобрене и негативно отношение: Като подобрения преводачите предлагат да бъде увеличен обемът на въведените термини, да бъде дадена възможност на специалистите да ги допълват, да може системата за машинен превод да си взаимодейства с преводната памет и терминологичната база на програмите за компютърнопомогнат превод и да заимства термини и фрази оттам, както и да се диверсифицира по области и да се разграничават специфичните термини по области. Възможна причина за изразеното негативно отношение към машинния превод е, че той става все по-добър, което прави грешките все по-малко очевидни и трудни за откриване. Присъстват и коментари, че машинният превод не трябва повече да се подобрява, защото преводачите ще станат излишни. Анкетираните очакват да видят подобрения в качеството на превода на български език.

Анкетата може да се оцени по следния начин. Отговорите на българските преводачи не се различават в значителна степен от другите анкети, които бяха представени. Преводачите се стремят да използват новите технологии като преводна памет и терминологични бази, тъй като те могат да допринесат за подобряване на бързината и донякъде на консистентността на превода. От друга страна, има все още преводачи, които не само не използват нови технологии, но и изразяват притеснението си от тях. Тук трябва да се посочи и липсата на публичнодостъпни терминологични бази

в много области от човешкото познание, както и недостатъчно качественият машинен превод особено от езици, различни от английски.

Обобщение

Подробно бяха представени основните компоненти на системите за компютърноподпомогнат превод: преводна памет, машинен превод, терминологични бази, като беше обърнато внимание на многоезиковите (основно двуезикови) терминологични бази. Прегледът на системите за компютърноподпомогнат превод показва не само напредъка на технологиите в тази област, но и измененията в работните процеси, образованието и обучението на преводачите. Ясна тенденция е постоянното разширяване на спектъра от функционалности на системите за компютърноподпомогнат превод (чрез допълнителни възможности за работа с регулярни изрази, управление на съдържанието за изграждане на уебсайтове, мобилни приложения и други). Еднотипните документи, в които всяко изречение (или част от него) принадлежи към отделни сегменти, запазени в преводната памет, предопределя работата с текстовете със стандартизирано съдържание. Докато системите за компютърноподпомогнат превод и свързаните с тях преводна памет и терминологични бази постепенно навлизат трайно в работата на преводачите, от направения анализ стана ясно, че отношението на преводачите към машинния превод е амбивалентно: за едни той е полезен, макар че трябва да се подобри (особено за някои двойки езици), а за други е неприемлив, включително се разглежда като потенциална конкуренция и заплаха за работата.

Използването на системите за компютърноподпомогнат превод налага промяна в изискванията към образованието и умението на преводачите. Профилът на преводача вече не може да бъде ограничен само до езиковите умения. Преводачите се превръщат във висококвалифицирани технически експерти поради съдържанието, което превеждат, и инструментите и софтуера, които трябва да използват в ежедневната си работа (свързана с познаването на различните системи за компютърноподпомогнат превод, управлението на преводна памет, създаването и редактирането на терминологични бази, последващото редактиране на машинен превод, необходимостта от програмиране, работата със субтитри, локализацията, обработката на изображения, обработката на файлове в различни формати и др.).

Анкетата сред български преводачи, чиято ежедневна работа е свързана с технологиите за компютърноподпомогнат превод, потвърждава тези изводи. От

резултатите става ясно, че (с малки изключения) българските преводачи са добре запознати с инструментите за компютърноподпомогнат превод, макар че достъпността (предимно от финансова гледна точка) се оказва съществен фактор за тяхното използване. Преводачите формулират редица препоръки към функционалностите на системите за компютърноподпомогнат превод, включително тяхното използване да стане все по-опростено. В тази връзка, мнението на българските преводачи оправдава нуждата от разработване на универсална, интуитивна и леснодостъпна методика, чрез която да могат бързо, безплатно и сигурно да създават терминологични бази.

ВТОРА ГЛАВА

Терминологични бази в системите за компютърноподпомогнат превод

Увод	66
I. Термини, терминологични речници и терминологични бази	67
I.1. Термините в контекста на превода	67
I.2. Терминологични речници	70
I.3. Терминологични бази	72
II. Сравнение на терминологичните бази с терминологичните речници	81
III. Връзка на терминологичните бази с преводната памет	84
IV. Създаване и управление на терминологичните бази в системите за компютърноподпомогнат превод	85
IV.1. Притурки към основната програма	87
IV.2. Системи за управление на терминология	88
IV.3. Използване на контролирани езици	91
V. Умения за работа с терминология и терминологични бази	93
Обобщение	95

Настоящата глава описва функционалностите на терминологичните бази в системите за компютърноподпомогнат превод. Направено е сравнение между терминологичните бази и други видове терминологични ресурси, както и на методите и средствата, с които се създават. Разгледани са текстовете за превод в контекста на употребата на терминологичните бази. Основната цел, която се поставя за разрешаване във Втора глава, представлява изясняването на принципите за построяване на терминологичните ресурси и изясняването на приликите и разликите между традиционните терминологични ресурси (които също имат място при превода) и терминологичните бази, които се вграждат в системите за компютърноподпомогнат превод.

По-конкретно, Втора глава е насочена към изпълнението на следните цели:

- Описание на основни понятия, свързани с термините, терминологичните речници и терминологичните бази. Това предполага кратко проследяване на възгледите за

обхвата и описанието на термините в контекста на съвременните подходи при превода, както и дали тези възгледи се различават по отношение на възгледите за традиционните терминологични речници.

- Да се представят на видовете терминологични речници и тяхната структурата в съпоставка със структурата на терминологичните бази. Целта е да се очертаят приликите и разликите, както и преимуществата при използването на терминологични бази в процеса на работа със системи за компютърноподпомогнат превод.
- Да се опише структурата и особеностите на терминологичните бази. Представяне за световноизвестни терминологични бази, които могат да се използват във или извън системите за компютърноподпомогнат превод.
- Да се представи създаването и управлението на терминологичните бази в системите за компютърноподпомогнат превод. Ще бъдат разгледани основните начини за създаване и управление на терминологични бази в системи за компютърноподпомогнат превод. В допълнение ще се разгледат притурки, които могат да се използват заедно със системите за компютърноподпомогнат превод, както и самостоятелно функциониращи системи за работа с терминология, които могат да бъдат използвани паралелно.
- Да се покаже използването на терминологичните бази в зависимост от различните типове текст за превод и приложението на контролирани езици. Тук основно внимание ще се обърне на изискванията за последователност и непротиворечивост при употребата на термини и възможността да се наложат специфични ограничения. Ще бъдат разгледани накратко контролираните езици и тяхното приложение в специфични области.

Накратко ще бъдат посочени новите умения, които се изискват от преводачите за работа с терминология и терминологични бази.

Значението на главата се изразява в това термините в терминологичните бази и самите терминологични бази като структура и предназначение да бъдат поставени в контекста на терминологията като начин за представяне на специфично човешко познание, на традиционните терминологични речници, които не са и няма да изгубят своето значение, както и на широкомащабното описание на термини от различни тематични области, на което сме свидетели в момента.

Увод

Терминологията е ключова част от богатството на един език. Основната ѝ цел е да се елиминира многозначността и да се постигне еднозначно съответствие между понятие и термин. Може да се каже, че е „част от специален речник, който включва думи и фрази, които назовават обекти и понятия от различни сфери на професионалната дейност на човека и не се използват често“ (Бактиорович 2021: 12).

Термините са показател за развитието на науката. Те представляват уникален начин за изказване на дадена идея: по своята същност термините са еднозначни и имат за цел да назоват определено понятие.

Изследването на терминологията като средство за комуникация за специални цели започва през 20-ти век, когато става очевидно, че това е отделна област, изискваща свои методи на описание. Институции и организации по света са разработили и продължават да разработват терминологични ресурси, които имат за цел да подпомогнат комуникацията. Например в страни с повече от един официален език като Канада, Швейцария или Южна Африка често създаването на терминологични ресурси е належащо, за да може да се улесни разпространяването и разбирането на законите, както и правата и задълженията на гражданите на тези страни. Същото се прави и от наднационални организации като Европейския съюз и Обединените нации, чиито терминологични ресурси се използват в подкрепа на международната комуникация и търговия. Неправителствени организации като Световната здравна организация също имат терминологични ресурси и с тях целят да подпомогнат ефективното прилагане на своите програми. Частните компании разработват терминологични ресурси за своите продукти и услуги, имащи за цел повишаване на качеството, подобряване на производителността и запазване на последователната употреба на използваните от тях термини.

Конкретно, що се отнася до превода, приема се, че в два езика „могат да се изразяват едни и същи значения“ (Пим 2009: 82). Дадена дума обаче може да има различни значения на друг език или дадено понятие може да не присъства в езика изобщо, което е и причина връзката между два езика да не е симетрична.

Мария Попова определя като характерни две тенденции за съвременната езикова ситуация: 1. Английският език е глобален език и 2. Създаването на общ фонд от „съотносителни термини и терминоелементи“, използвани във всички езици (Попова 2020: 158). Според нея езиковата глобализация засяга всички равнища на езика или както тя ги нарича, на „терминоелементите“ (корени, префикси и суфикси), термините думи и

термините словосъчетания. Резултат от глобализацията са не само терминологичните системи, но и отделни страни в строежа на даден термин (съответно: значение – понятийна структура, комплекс от номинационни признаци – вътрешна форма, название – външна форма (Попова 2020: 159). М. Попова посочва, че най-често в приемащия език навлиза цялостният английски термин. Пример за това е думата *инфлация* (inflation). Има обаче и случаи, в които се заема само един пласт от общия строеж на термина, неговото значение е пренесено частично. Например думата *пул* (pool) е навлязла в езика ни само със значението ‘обединение от предприятия’. Според М. Попова „термините са най-информативният пласт в лексиката на научния език, реч и текст (Попова 2019: 149).

I. Термини, терминологични речници и терминологични бази

„Политическите, социалните и идейните промени след Втората световна война намират отражение и в областта на терминологията. Някои термини в областта на обществените науки се преосмислят, въвеждат се и много нови термини, други отпадат, тъй като отразяват понятия от отречената буржоазна идеология“ (Бояджиев 2012: 46). Макар че думите обикновено се разбират като поредица от знаци, ограничени с интервали или пунктуация, термините може да включват повече от една дума (например *изменение на климата* или *спортен автомобил*). Терминът *спортен лек автомобил* е известен на английски с абревиатура SUV. Следователно както семантично, така и морфологично термините проявяват определени свойства, които ги отличават от така наречените общоупотребими думи (Уорбъртън 2015: 646).

I.1. Термините в контекста на превода

В третомния „Речник на съвременния български книжовен език“ се дава следната дефиниция за термин: „Дума или израз, които се употребяват в някоя област на науката, техниката, изкуството и означават строго определено понятие“, например технически термини, юридически термин и музикални термини. Сред експертите в дадена област е налице разбирателството да се използват едни и същи термини с дадено конкретно значение. В момента, в който такъв термин излезе извън границите на своята „среда“, това разбирателство се губи (Делавиние 2017: 31). Терминът придобива нови значения и се появява многозначност. Т. Бояджиев констатира, че „около една пета част от думите в българския език са многозначни, като в отделните им типове многозначността е развита в различна степен. Съществителните имена, прилагателните имена и глаголите показват по-интензивни семантични изменения. Всяко пето съществително или прилагателно име и всеки четвърти глагол са многозначни“ (Бояджиев 2011: 81). В свой

експеримент върху аотиран корпус Св. Коева, Д. Благоева и С. Колковска изчисляват, че коефициентът на многозначност на част от речта е 1,51, което означава, че средно всяка втора дума има повече от едно значение (Коева, Благоева и Колковска 2010: 3681). Тук ще бъдат разгледани дефинициите за термин с оглед на употребата на термините в терминологичните бази, където многозначността не е желана (въпреки че може да се има предвид в зависимост от проекта за превод). По наличието на многозначност термините си приличат с думите. Ако са многозначни обаче, термините губят основното си свойство да назовават по уникален начин дадено понятие, т.е. да са еднозначни. Терминологията и преводът са две неразривно свързани области. В тази връзка М. Телен разграничава два вида терминология: едната ориентирана към превода (Translation-oriented), а другата ориентирана към теорията за термините (Theory-oriented). Той анализира взаимодействието между терминологията и превода. Приносът на терминологията към превода е очевиден, но този на превода към терминологията, макар и по-малко очевиден, също съществува (Телен 2015: 379). Тясното взаимодействие между тях може да хвърли повече светлина върху това има ли разлика (и ако има, каква е тя) между термини и общоупотребими думи и между специфичен за конкретна тематична област език и общоупотребим език.

Х. Сайер предлага следната дефиниция за термин: „Термините се различават от думите по това, че притежават специална форма на референция (form of reference), а именно, отнасят се до отделни концептуални единици, свойства, дейности или отношения, които се отнасят до знанието от определена предметна област. За да се направи разлика между специална и обща референция, се установява разграничение между термините, които имат специална референция в рамките на определена научна дисциплина, и думите, които функционират с обща референция в различни предметни области“ (Сайер 2001: 261).

А. Кис дава значително по-съкратена дефиниция: „терминът (авторът използва названието *terminus technicus*) е израз, принадлежащ към техническия език“ (Кис 2005: 106). А. Имре констатира, че различните дефиниции подчертават различни аспекти: някои от тях вземат предвид „формата“ на думата или израза, а други се фокусират върху значението. Терминът трябва да има само едно значение, като синонимните термини се изключват, но най-важното е, че значението на термините е добре установено (без припокриване със значението на други термини), така че да не може да бъде разширено или стеснено (Имре 2013: 127). Както става ясно, многозначността е факт и термините

трябва да се използват по възможност с унифицирано за даден контекст значение, т.е. да не са многозначни.

Тук е мястото да се спомене разграничаването на широко и тясно разбиране за терминологията: при широкото разбиране за терминология се смята, че термин е всяка дума или фраза, изразяваща понятие, която се използва в определен контекст (Тръмбъл и Стивънсън 2002); тясното разбиране приема, че термин е дума или фраза, използвана с точно определено значение в дадена тематична област или дисциплина. Въпреки това термините рядко се използват изолирано от другите езикови средства: тяхното място е в професионалните текстове и трябва да се подчиняват на езиковите правила на езика, на който се прави преводът. Освен това, за да бъдат напълно работещи, термините трябва да назовават съответните понятия по недвусмислен начин (Лукашик 2012: 98).

„Правилните“ (добре подбрани) термини помагат за справянето с предизвикателствата на превода. Като част от всеки проект за превод терминологията или преводачите трябва да идентифицират подходящи еквиваленти на специализираните термини от изходния текст. В повечето случаи няма разлика между терминологията, използвана в едноезикови текстове и преводната терминология. Ако текстовете няма да бъдат използвани на друг език, е много вероятно терминологът или този, който подбира термините, да се ограничи до употребата им за конкретна област. Ако обаче текстовете ще се превеждат, в зависимост от вида текст е възможно да бъде изпратена първа версия на термините към преводача, за да се провери какви еквиваленти би избрал и доколко има риск значението да се измени в целевия език.

Определянето на терминологията, необходима за извършване на даден превод, може да бъде трудна и отнемаша време задача. Следователно терминологията и преводачите съставят речници или терминологични бази от данни (в най-добрия случай използват готови), в които записват резултатите от своите терминологични изследвания за бъдещи справки. Всяка организация използва специфични термини, които не винаги могат да бъдат намерени в стандартните речници. Например, различни термини се използват за обозначаване на една и съща функция в различни браузъри: „любими“, „отметки“ и др. За дадена организация е важно да използва последователно правилните термини във всички свои форми на комуникация (технически ръководства, помощни статии и др.)⁶¹.

⁶¹ <http://www.ugr.es/~mvegaex/SDL%20MultiTerm%202009%20for%20Translators.pdf>

I.2. Терминологични речници

Терминологичните речници се характеризират с експертни дефиниции, думи и значения, които принадлежат към конкретна техническа или научна предметна област (като медицина или математика) (Де Калуве и Ван Сантен 2003: 82). Изборът на термини за речника определя неговия тематичен обхват, а когнитивната му роля се реализира или чрез представянето на семантичните отношения между термини и понятия, или чрез имплицитна семантична информация, предадена с помощта на дефинициите на включените термини. От тази гледна точка терминологичният речник се превръща в средство за трансфер на професионално знание (Лукашик 2012: 100).

Въз основа на такава информация е възможно правилно да се идентифицира даден термин и подходящото му използване в различни езикови контексти. Следователно е необходимо включването на набор от езикови данни (съобразени с нуждите на потенциалните потребители на речника), за да може терминологичният речник да се превърне в справочник с езикова стойност (Лукашик 2012: 101). Формалните характеристики на термина са първият индикатор за неговото естество. След това в даден запис идват различни видове етикети: граматични, синтактични, прагматични, етимологични, ортографски, фонетична транскрипция, колокационни модели, примери за употреба и т.н.

Речниците също са повлияни от развитието на технологиите. Това влияние се отразява най-вече върху вида и достъпността им, и по-конкретно дали са онлайн, или печатни. Според статия на Ню Йорк Таймс⁶² от 2017 г. „продажбите на отпечатани речници остават значителни и представляват средство за печалба за някои издатели“. В статия на Гардиън⁶³ от следващата година обаче се посочва, че продажбата им намалява, докато броят на онлайн приложенията и алтернативните онлайн публикации продължава да расте, например третото издание на Оксфордския речник на английския език⁶⁴ не може да бъде закупено в хартиен вариант. Справка на сайта показва, че такава възможност има само за второто издание⁶⁵. Според подобни данни бъдещето на отпечатаните речници е неясно, което се обяснява с достъпността на същия вид информация онлайн. Печатните речниците се състоят от списък с думи, подредени по азбучен реда. Посочват се частите на речта и в някои случаи честотата на отделните

⁶² <https://www.nytimes.com/2017/02/11/fashion/merriam-webster-dictionary-social-media-politics.html>

⁶³ <https://www.theguardian.com/news/2018/feb/23/oxford-english-dictionary-can-worlds-biggest-dictionary-survive-internet>

⁶⁴ <https://www.oed.com/>

⁶⁵ <https://public.oed.com/help/>

граматични форми. Развитието на математически, предимно статистически, методи за изучаване на езика е свързано със създаването на честотни речници, думите в които получават числов, статистически показател. Чисто практически цели се преследват от правописните речници, указващи „правилния“ (тоест отговарящ на приетата норма) правопис на думите и техните форми или съответно „правилното“ им произношение.

Не трябва да се забравя, че независимо как ще бъдат достъпни (като книга, електронно или като база от данни), както и към коя област принадлежат, термините от терминологичните речници изискват надлежно обмисляне преди употребата им в преводните текстове. Невъзможно е преводачите да бъдат експерти във всички сфери на познанието и трябва да им се предложи достъп до съществуващите терминологични речници, както и актуализация на терминологичните речници, за да се осигури последователност (Янг 2021: 1).

Най-видимата разлика между традиционните речници и тяхната онлайн версия естествено е хартиеният носител, който дава възможност на преводача да го докосне физически. Най-често печатният речник е съставен от експерти с признат авторитет. Дигиталните речници пък имат предимството да бъдат достъпни дистанционно. Друго ясно предимство на дигиталната версия е по-лесният достъп до лексикални ресурси. В традиционния речник лексикографските данни се подреждат най-често по азбучен ред. За разлика от тях, дигиталните речници не са ограничени нито от формални, нито от семантични критерии в зависимост от избрания принцип на организация (Лю 2013: 80). Ж-М. Дьо Шривер споменава цели 118 предимства на електронните речници пред традиционните хартиени издания (Дьо Шривер 2003: 153 – 185), сред които възможността за кръстосани препратки, липсата на ограничения на пространството, достъпността на нови видове информация (линкове, видео и други), пряка връзка с корпуси и възможността за постоянно обновяване. Към тях може да се добави и наличието на речници под формата на мобилни приложения. Разбира се, все още не всички електронни приложения са качествени и с вярно съдържание. От друга страна, е необходимо да се отбележи, че за употребата на електронните ресурси са необходими електричество и интернет, което също е форма на съобразяване с условията.

Към терминологичните речници могат да бъдат включени един или повече от следните компоненти (Попова 2016: 71):

- научната област на термина;
- речникова единица – заглавка;
- лексикално-граматична информация – като част на речта;

- формалноезикова информация – граматични особености;
- функция;
- етимологични данни;
- информация за съчетаемостта;
- прагматичен параметър – особености при употребата и степента на разпространение;
- илюстративен параметър;
- регистрационен параметър (датата на регистрация на дадения термин);
- интерпретационен параметър;
- параметър на понятийно-семантичната системност;
- параметър на лексикалносемантичната системност;
- времеви параметър;
- метадиалектен параметър (термините на дадена научна школа);
- идиолектен параметър (термините на даден учен);
- ареален параметър;
- категориален параметър;
- сведения за източниците.

Съществуват различни видове терминологични речници, формалните характеристики на термина са първият индикатор за неговото естество. Граматичната, синтактичната и етимологичната информация, фонетичната транскрипция, примерите за употреба и т.н., също допринасят за неговото характеризиране. Всичко това е ключово, за да може правилно да се идентифицира терминът като езиков знак на даден език и подходящото му използване в различни езикови контексти. Това на свой ред помага терминологичният речник, в който е поместен терминът, да се превърне в справочник с езикова стойност за специалистите, които имат нужда от него.

I.3. Терминологични бази

Според ISO 30042:2008 терминологична база е „база от данни, която съдържа информация за специализирани езикови понятия и термините, които обозначават тези понятия, заедно с допълнителна информация“. В търговски условия обаче трябва да се отбележи, че терминологичната база не е ограничена до „специализирани езикови понятия“, а може да съдържа всяка лексикална единица, която трябва да се „управлява“, за да постигне качествено и последователно съдържание. Терминологичната база е база от данни с термини, отнасящи се до конкретна област (или проект), която е интегрирана в инструмент за компютърноподпомогнат превод.

Терминологичните бази са централизирани бази от данни, които съдържат специфични за дадена тематична област или даден проект за превод термини и са един от основните компоненти в системите за компютърнопомогнат превод. Те могат да бъдат двуезикови или многоезикови. Всеки езиков термин представлява отделен запис и е придружен със своето еквивалентно значение или значения. Терминологичната база може да съдържа дефиниции или специфични инструкции за използване на преводната терминология. Може да се използва като преводната памет, за търсене на термини, които са подходящи за целите на даден превод. Терминологичната база помага да се гарантира точността и последователността на общата или специализирана терминология в проектите за превод. Ако терминологичната база е добре интегрирана с преводната памет, тя позволява на преводачите да превеждат значително по-бързо.

Обикновено терминологичната база се изгражда от потребителите на инструмента за компютърнопомогнат превод (това не е задължително условие за възлагането на даден проект). Терминологичната база оптимизира, прецизира и персонализира използването на преводната памет. Когато паметта не дава очакваните съвпадения, терминологичната база може да помогне. Терминологичните двойки не се съхраняват автоматично, както се случва при сегментите в паметта, а се създават и определят в началото на даден превод, след което може да се добавят нови записи, като се маркират изходните и целевите двойки термини или фрази.

Докато преводната памет се използва за запаметяване на съдържание на ниво изречение и за подпомагане на процеса на превод с необходимото равнище на последователност, терминологичната база е по-скоро специално „ръководство“, което подпомага преводачите да се справят с терминологията, свързана с областта на превода (Вокалович 2021: 8).

Сами по себе си терминологичните бази са като речници, но начинът по който са създадени и управлявани и по който функционират, ги прави коренно различни. Основната им характеристика е, че са персонални и персонализируеми за всяка институция, частна компания и отделен преводач. Могат да бъдат създавани в неограничен брой и с предпочитано съдържание. Техните създатели избират как да ги конфигурират, за кои езикови комбинации, кои термини да бъдат добавени, премахнати, забранени и как тези термини да бъдат класифицирани. Тези промени могат да бъдат осъществявани във всеки един момент. Могат да бъдат комбинирани няколко двуезикови бази в многоезикови, да бъдат споделяни, съхранявани, разделяни и регрупираны с различен статус в зависимост от нуждите на всеки нов проект.

Понастоящем няма ограничение за количеството и вида на информация във всеки терминологичен запис, тъй като съвременните технологии освобождават лексикографските и терминологичните ресурси от ограниченията за място. По-конкретно, терминологичният ресурс трябва да бъде съобразен със специфичните нужди на потребителите, които работят с многостранни типове специализиран текст (Фабер и Араус 2021: 588).

Най-общо казано, към терминологичните бази могат да бъдат включени един или повече от следните компоненти (Райт 2001: 573):

- дефиниция (описание на значението на термина);
- източник;
- контекст на употребата на термина;
- област, към която принадлежи терминът⁶⁶;
- граматическа информация (глагол, съществително и др.);
- етикет за употреба (например: фигуративен, американски английски, официален и др.);
- автор („създаден от“ – потребителят, който е добавил новия термин; в случай на различни потребители преводачът знае кой термин от кого е създаден / добавен);
- дата на създаване / модификация („създаден / модифициран на...“);
- статус („проверен“, „одобрен“, „предпочитан“ или „забранен, да не се използва“);
- връзки / линкове между отделните термини и дефиниции, както и линкове към външни източници;
- бележки.

В зависимост от това какъв инструмент се използва за създаването на терминологичните бази, тези компоненти се създават автоматично, например информацията за авторство или датата на създаване. Ако преводачът създава сам терминологична база, е необходимо да определи каква информация иска да добави. При програмите за компютърноподпомогнат превод тези полета са налични винаги, стига за тях да е налице конкретната информация.

Докато преводната памет се използва за запаметяване на съдържание на ниво изречение и за подпомагане на процеса на превод с необходимата последователност, терминологичната база е по-скоро специално „ръководство“, което да помогне на

⁶⁶ Докато при речниците биват изредени всички или поне най-популярните значения на даден термин, в терминологичните бази се предпочита еднозначността. Всяка база е създадена за конкретна област, клиент или проект. Не се взема под внимание какво означава даденият термин в други области.

преводачите да се справят с терминологията, свързана с областта на превода (Вокалович 2021: 8).

Базираният на понятия подход към управлението на терминологията изисква цялата терминологична информация, отнасяща се до едно понятие, да се обработва като един терминологичен запис. В система, базирана на понятия, данните са организирани около значението, а не около термина, подобно на организацията на данните в Уърднет (Милър и др. 1990). Всички термини, синоними и преводни еквиваленти, които обозначават едно понятие, както и всички описателни и административни данни, отнасящи се до същото понятие, се съхраняват в един терминологичен запис.

Според К. Уорбъртън стандартен процес на управлението на терминологията в проектите за превод може да изглежда така:

1. Терминът от изходния език бива идентифициран още преди началото на превода. Този етап може да включва в себе си процеса за извличане на термина;
2. Терминологът добавя термина от изходния език към терминологичната база с допълнителна информация, ако е необходима;
3. Определя се езикът или езиците за превод и думата се маркира за превод. Това може да включва използването на специално поле за тази цел;
4. Терминологът или терминологите, работещи с целевия език, биват уведомени по изработения ред;
5. Терминологът или терминологите (особено когато става въпрос за повече от един език) добавят преводите за избрания термин;
6. Преводите стават достъпни за преводачите чрез филтър, който идентифицира проекта, към който принадлежи новодобавеният термин. Това може да включва и експорт/импорт към инструментите за компютърноподпомогнат превод или директна връзка с програмно-приложен интерфейс (API);
7. Съдържанието на проекта (документ, уебсайт и др.) се изпраща за превод;
8. Преводачите могат да добавят нови термини към двуезиковата терминология по време на превода;
9. Новодобавените термини могат директно да се включат в базата (чрез експорт/импорт или директна връзка) или да чакат одобрение от специалиста, отговорен за конкретния проект (Уорбъртън 2021: 186).

Необходимо уточнение е, че критериите за подбор на термини зависят от изискванията на конкретния проект. В медицинската терминология грешният избор на

дума може да има сериозни последици. Поради тази причина по-долу предлагаме следните стъпки за идентифициране на това дали даден термин е подходящ за конкретен проект:

- За да определи кои термини са подходящи, преводачът трябва да знае дали терминът, който иска да използва, принадлежи към книжовния език, или не. Не е изключено при големи компании, разположени в различни градове, да са се развили различни начини да се обозначи един и същи термин, но една от използваните думи да не принадлежи към книжовния език;
- Да се определи дали наистина терминът се използва;
- Да се определи дали терминът се използва с еднакво значение в организацията или проекта, за който е предназначен;
- Да се определи дали няма риск от неправилно използване на термина. По-конкретно, дали принадлежи на техническия език, свързан с проекта за превод, или е част от общоупотребимия език;
- Да се определи дали терминът не води до културни препратки. Това ще доведе до затруднения по време на превода;
- Да се определят синонимите и какви значения носи всеки един от тях.

Благодарение на информационните технологии и интернет в днешно време могат да бъдат намерени голямо количество свободностъпни терминологични ресурси. По долу са изброени някои от тях. Част са собственост на европейски институции, а други – на неправителствени организации.

Общото между тях е, че са популярни сред преводачите и могат да бъдат приети като авторитетни. Л. Бендана и А. Мелби (Бендана и Мелби 2012: 53–54) представят изброените по-долу терминологични бази като примери за авторитетни ресурси. Няколко от тях бяха посочени и от българските преводачи по време на направената анкетата.

- ИАТЕ (Interactive Terminology for Europe или IATE⁶⁷) може да се смята за най-известният терминологичен ресурс за преводачите в България, както бе потвърдено от анкетата сред тях. В платформата към юли 2022 г. могат да бъдат намерени общо 8 047 139 термина. Поддържани са всички официални езици на Европейския съюз. Този терминологичен ресурс предоставя свободен достъп до

⁶⁷ <http://iate.europa.eu>

проверена терминология. Освен това позволява на терминологите от ЕС да получат обратна връзка от външни потребители. Структурата на данните в ИАТЕ се основава на концептуално-ориентиран подход: всеки запис съответства на едно понятие. Програмата дава възможност за търсене в речници и публикации на ЕС. Списъците с термини могат да бъдат изтеглени за използване. Термините са групирани по тяхното значение и всяко понятие в идеалния случай трябва да бъде покрито от един запис. Даден запис е разделен на три нива: ниво, независимо от езика (language-independent level или LIL), езиково ниво (language level или LL) и ниво на термина (term level или TL). На нивото, независимо от езика, са областта, към която принадлежи понятието, информация откъде е взет терминът, дали вече не се използва, или не съществува. Към езиковото ниво принадлежат езиковият код (за български е „bg“), дефиницията към коя част на речта принадлежи терминът, и евентуални бележи, свързани с него. Към нивото „термин“ принадлежат терминът, негови синоними, съкращения, формули (например химически формули), словосъчетания и вариации в правописа.

- Евровок (EuroVoc⁶⁸) е многоезиков и мултидисциплинарен тезаурус, обхващащ терминология от областите на дейност на ЕС. Също като ИАТЕ и този ресурс се предлага на 24-те официални езика на ЕС, както и на езиците на трите страни кандидатки (Албания, Северна Македония и Сърбия). Търсенето на термин става чрез избирането на език от менюто и въвеждане на термина в полето за търсене. Програмата дава възможност за търсене в речници и публикации на ЕС. Списъците с термини могат да бъдат изтеглени за използване. Термините са йерархично организирани, като на първо ниво са областите на употреба, на второ ниво са относително малък брой термини, принадлежащи към всяка област, и на трето ниво са т.нар. микротермини.
- Електропедия (Electropedia⁶⁹) е база от данни на Международната електротехническа комисия (IEC) и предоставя най-големия брой термини и дефиниции в света в областта на електрическата енергия и електрониката, като съдържа повече от 22 000 терминологични записа на английски и френски, организирани по предметни области и с еквивалентни термини на различни други езици: арабски, китайски, чешки, нидерландски, финландски, немски,

⁶⁸ <https://op.europa.eu/en/web/eu-vocabularies>

⁶⁹ <https://www.electropedia.org/iev/iev.nsf/Welcome?OpenForm>

италиански, японски, корейски, монголски, норвежки (букмол и нюношк), полски, португалски, руски, сръбски, словенски, испански и шведски, като броят термини варира според предметната област. Например термини в областта на математиката, и по-конкретно на общите понятия и от линейна алгебра, където са изредени всички налични термини за даден език и потребителят може да разгледа еквивалентите им за някой друг от включените езици.

- Глосари линкс (Glossary Links⁷⁰) е инструмент за търсене в база от данни от повече от 8 000 речника. Всички връзки се актуализират в зависимост от критерии като уместност и надеждност. В него могат да се намерят речници на различна тематика и език. Речниците обхващат широк спектър от области, например правни въпроси, околна среда, медицина, защита на потребителите, човешки права, икономика и други. Връзките са само към заглавните думи на отделните речници. Полето „източник“ дава възможност за търсене по имената на авторите на речниците.
- Голям терминологичен речник (от фр. Le grand dictionnaire terminologique⁷¹ или GDT) е банка от терминологични записи, създадени от Квебекския офис на френски език (от фр. Office québécois de la langue française). Всеки термин се отнася до понятие, свързано със специализирана област на употреба, като термините се представят на френски и английски, а понякога и на други езици. Термините са класифицирани по области на употреба и често се обясняват с дефиниция, бележки или илюстрация. Това прави възможно намирането на френските еквиваленти на английски или испански или проверката на значението на термин, принадлежащ към техническия или научния речник. Както се посочва, макар че в продължение на много години терминологията се фокусира почти изцяло върху езиковото обозначение, тази голяма терминологична база вече включва по-голямо количество визуално съдържание (Боукър 2020: 579 – 583).
- Термиум плус – Банка за терминология и езикови данни на правителството на Канада (Termium Plus: Government of Canada’s Terminology and Linguistic Data Bank⁷²). Втре могат да бъдат намерени термини на английски, френски, испански и португалски със съответните съкращения, дефиниции и примери за употреба в широк спектър от специализирани области. Банката може да служи за

⁷⁰ <https://termcoord.eu/glossarylinks/>

⁷¹ <http://www.oqlf.gouv.qc.ca/>

⁷² www.termiumplus.gc.ca

разбиране на акроним, проверка на официално заглавие и намиране на еквивалент на друг език. Примери за налични термини е лексиконът на Канадските институти за здравни изследвания.

- Многоезиковата терминологична база на ООН (United Nations Multilingual Terminology Database⁷³ или UNTERM) е поддържана съвместно от Секретариата и част от специализираните агенции на системата на ООН, включително Международната морска организация, Организацията на обединените нации за образование, наука и култура, Световната здравна организация и Световната метеорологична организация. Съдържа терминология по теми, свързани с работата на ООН. Информацията се предоставя на шестте официални езика на ООН (арабски, китайски, английски, френски, руски и испански), а има и записи на немски и португалски. За първи път е представена през април 2012 г. като портал за търсене, обединяващ всички терминологични колекции на ООН.
- ТермТерм (TermTerm⁷⁴) е проект, имащ за цел да изясни и стандартизира основните понятия и терминология, използвани в науката за терминологията. Базата от данни предоставя безплатен достъп до: основните понятия на терминологичната работа, термини на няколко езика и определенията, взети от съответните терминологични стандарти. Например, в полето за търсене се въвежда даден термин и се зареждат всички налични преводи за него, както и всички словосъчетания, към които принадлежи, ако има такива.

Разбира се, в интернет могат да се намерят много свободнодостъпни терминологични бази на различна тематика, които са собственост на частни компании, например терминологичният речник на Тилде (Tilde Terminology⁷⁵) или този на Майкрософт. Речникът на Тилде позволява семантично търсене, визуализация на терминологичната мрежа на всеки термин, показва ресурса, в който се намира терминът, и връзката към вписването на този термин. Позволява се също и проследяване на йерархията на термините. Терминологичната колекция⁷⁶ на Майкрософт представлява набор от стандартни технологични термини, използвани в продуктите на компанията.

⁷³ <http://unterm.un.org>

⁷⁴ <https://termterm.kaleidoscope.online/dashboard>

⁷⁵ <https://term-extraction.tilde.com/>

⁷⁶ <https://www.microsoft.com/en-us/language>

Структурата на терминологичните бази е стандартизирана от ISO TC37 чрез ISO 16642: Терминологична рамка за маркиране. Съгласно този стандарт терминологичните записи са структурирани в три йерархични раздела: понятие, език и термин. Информацията на ниво понятие описва понятието като цяло и това описание се споделя от всички термини на всички езици в записа. Цялата информация, отнасяща се до даден език, е организирана в специален езиков раздел, който е разделен на секции за информация за отделните термини. Информацията, която се споделя от всички термини на даден език, се появява на езиково ниво, като специфична за езика дефиниция или коментар, а информацията за конкретен термин се представя на ниво термин като бележка за употреба или контекст. Например, секцията с термини може да съдържа само един термин, езиковата секция може да съдържа множество секции с термини, а секцията с понятия може да съдържа няколко езикови раздела. Всички термини в даден запис са синоними. Този принцип се нарича концептуална ориентация (Уорбъртън 2015: 654).

Браншовите стандарти (ISO 704, ISO 30042, ISO TC37, ISO 16642 и ISO 26162) описват принципите и методите на терминологична работа, универсалния формат и проектирането, внедряването и поддръжката на системите за управление на терминологията.

Законодателната рамка, регулираща авторското право, е различна за всяка държава. Поради тази причина са възприети универсални и обединяващи международни стандарти. В Приложение 7 на своята статия, публикувана в книгата *Наръчник по превод и технологии на Рутледж* (The Routledge Handbook of Translation and Technology⁷⁷) от 2020 г., Сю Елън Райт тематично организира и изрежда стандартите, засягащи терминологията (Райт 2020):

- Теория на терминологията и терминологичните изследвания (ISO 704, ISO 860 и ISO 1087): описват се принципите и методите на терминологичната работа, хармонизирането на понятия и термини, както и терминологичната работа с речник и други компютърни приложения;
- Лексикография и терминография (ISO 1951, ISO 12199, ISO 12615, ISO 12616, ISO 22128, ISO 23185): описват се лексикографските символи, които се използват в речниците; начинът на организация на записите в речниците (по азбучен ред на

77

<https://www.routledge.com/The-Routledge-Handbook-of-Translation-and-Technology/OHagan/p/book/9781138232846>

многоезиковите терминологични и лексикографски данни); библиографските препратки; терминографията, ориентирана към превода, и терминологичната работа в подкрепа на многоезиковата комуникация. Представено е също и ръководство за терминологични продукти и услуги и за оценка и сравнителен анализ на терминологични ресурси;

- Терминологични практики (ISO 10241 и ISO 15188): по техен модел се изготвят и оформят международните терминологични стандарти, в това число общи изисквания и примери за представяне, както и насоки за управление на проекти за стандартизация на терминологията;
- Терминологична политика и езиково планиране (ISO/ NP 29383): този стандарт описва разработването и внедряването на терминологични политики;
- Моделиране на терминологичните понятия (DIN 2331 и ISO 24156): описание на системи от понятия и тяхното представяне, насоки за използване на единен език за моделиране;
- Системи за управление на терминологията (ISO 26162): проектиране, внедряване и поддръжка на системи за управление на терминология.

II. Сравнение на терминологичните бази с терминологичните речници

При сравнение на терминологичните бази и терминологичните речници в интернет могат да се определят следните общи черти между тях:

- Възможност за бързо препращане към други термини чрез връзките между тях, ако има такива;
- Възможност за публикуване на изображения и други помощни средства в записа за всеки термин;
- Когато се сблъскат с проблем с терминологията, преводачите могат да получат бърз достъп до основни терминологични бази, както и до множество речници и глосари за почти всеки език и тематика, ако търсената информация е налична;
- Могат да бъдат двуезикови или многоезикови;
- Позволяват актуализация на информацията. Това е важно, защото така се развиват заедно с развитието на езиците и технологиите. Дори може да се каже, че в определен контекст те са проводник за налагането на нови думи в целевия език (например, когато дадена корпорация решава какви да бъдат преводите за конкретна страна и организира маркетинговата си кампания около тях). Според

американската компания Глобъл ленгуидж монитор (Global Language Monitor⁷⁸ или GLM) на всеки 98 минути се създава нова дума в английския език. Ясно е, че не всички езици приемат такова количество нови думи, но това не означава, че и те не претърпяват неологични промени (например чрез придаване на ново значение на думи като „електронен документ“, „харесвам“, „споделям“ и др.). Преиздаването на традиционните речници прави изключително трудно следенето на всички неологизми. В повечето случаи, когато се появи нова дума, съставителите на даден речник трябва първо да решат дали тя да бъде включена в него. Някои от думите са „модерни“ за сравнително кратък период от време и не са актуални, докато дойде моментът за печат. Подобна е ситуацията и с онлайн версиите. Техните създатели трябва внимателно да обмислят дали даденият термин може да бъде добавен.

Приликите обаче свършват дотук. Общоприето е, че по време на процеса на превод терминологичните бази улесняват, а много често и изобщо елиминират избора на преводача дали да предпочете, или да отхвърли даден термин. Стандартизирана практика на преводачите да бъде изпратена предварително подготвена база с термините, одобрени от клиента, които да бъдат използвани в целевия текст. Ако се прави аналогия с терминологичните речници, даден клиент придобива ролята на съставител, защото той е този, който определя какво да се включи във всяка терминологична база. Нови термини могат да се добавят и от самия преводач, най-често чрез маркиране на желан термин в изходния език и избиране на опцията за добавяне в базата⁷⁹). Според някои специалисти именно заради лесната употреба това действие често води до „замърсяване“, което може да се превърне в проблем за преводача или създателите на терминологичната база⁸⁰: множеството резултати може да не се визуализират едновременно и преводачите може да се затруднят да изберат най-добрия вариант или да не знаят кой е предпочитаният термин за конкретния текст (Имре и Барабас 2020: 229–230). Друг проблем може да бъде конвертирането от един файлов формат в друг,

⁷⁸ <https://languagemonitor.com/>

⁷⁹ Разбира се, точният начин за добавяне е строго индивидуален за всяка програма за компютърнопомогнат превод. Тук се предлага обобщение на повечето инструменти за компютърнопомогнат превод.

⁸⁰ При стандартна работа клиентът, собственик на терминологичната база, решава колко и каква свобода да даде на преводача или редактора. В своите инструкции той може да забрани или разреши дали да има право да се добавят нови термини, или да се променят съществуващи.

при което да бъдат загубени бележки относно употребата на даден термин или наличието на думи, които не са термини и пречат на работата по време на превода.

Друга разлика е достъпността. Терминологичните бази са достъпни само в инструментите за компютърноподпомогнат превод (освен в случаите, когато дадена база или част от нея е експортирана, за да се използва с конкретна цел). В повечето случаи онлайн речниците са със свободен достъп. Пример за адаптивност и улеснение на потребителите са мобилните приложения на най-авторитетните речници в света.

Не на последно място е различен начинът на работа. В предната глава бе описан накратко начинът на работа на терминологичните бази. По време на превод в текстовия редактор на преводача не му се налага да търси термини. Програмата сама ги визуализира.

Както всички други, така и терминологичните речници могат да бъдат разделени на онлайн и печатни. Що се отнася до печатните, въпреки че са изправени пред голяма конкуренция, те далеч не трябва да бъдат отхвърляни. Докато терминологичните бази са индивидуални за всяка компания, проект и преводач, то традиционните терминологични речници и техните онлайн версии носят унификация и утвърждаване на значението на термините. Техните създатели обикновено са специалисти с утвърден авторитет. Според Лю и Дьо Шривер през вековете на речниците се е гледало като на авторитетни издания, често с възхищение и благоговение, а статутът на „речника“ в някои страни е можел да се оприличи с този на Библията (Лю и Дьо Шривер 2014). Последното не може да се каже за терминологичните бази, при които въпреки ясните инструкции за употреба или забрана на ползване на даден термин, не са изключени спорове между клиент и преводач за добавянето или премахването на конкретен термин, неговата употреба или значение. Специалистите в дадена област в много случаи предпочитат да се консултират с авторитетно издание, където могат да проверят правописа и значението на термина, който търсят.

Алтернатива на терминологичните бази са хартиените издания на терминологични речници. Пример за такива, включващи български език, са (не се цели изчерпателност при изброяването):

- Английско-български юридически речник, Христо Данов, издателство „Труд“, 1991;
- Английско-български терминологичен речник по дистанционни изследвания, Румяна Кънчева, АИ „Проф. Марин Дринов“, 2020;

- Английско-български речник по кардиология, Валентина Минчева, АИ „Проф. Марин Дринов“, 1991;
- Английско-български аграрен, бизнес и лесотехнически речник, Цветелина Цакова, АИ „Проф. Марин Дринов“, 2012;
- Английско-български морски речник с колокации, Вяра Петкова, издателство „Стено“, 2006;
- Многоезичен тематичен речник на европейската интеграция, издателство „Колибри“ от 2007.

Особеното при тези ресурси е, че всеки речник представлява различна тематична област и има различен обхват. Дори и в случаите, когато няколко терминологични речника представляват една дисциплина, съдържанието им и достоверността им се сравняват трудно. Причините за това също са много разнообразни. Времевият диапазон в годините на публикуване може да бъде ограничаващ при използването им. Ограниченията в тиража също могат да бъдат пречка до достъпността на тези ресурси. Именно заради последната причина двуезиковите ресурси в интернет (включващи български език) са добра алтернатива, например следните:

- Английско-български онлайн речник⁸¹;
- Английско-български онлайн речник на PONS⁸²;
- SA речник⁸³.

Проблемите с тях произтичат (с малки изключения) основно от липсата на информация за тяхната надеждност и достоверност. В повечето случаи не е ясно дали са създадени от експерти. Не е ясно също дали включените лексикални единици отговарят на критериите за термини и критериите за подбор. Разликата в обхвата и подбора на термините, както и специалистите, които са ги подбирали, са друг много важен фактор.

III. Връзка на терминологичните бази с преводната памет

Както беше споменато в предходната глава, преводната памет е вид инструмент, който помага на преводача да не превежда повторно сегментите, които преди това е превеждал (или други са превеждали). Преводната памет обикновено е интегрирана в системите за компютърнопомогнат превод. Всяко преведено изречение се съхранява в нея. Когато преводачът срещне нов сегмент, който съответства на друг от базата,

⁸¹ <https://www.rechnik-bg.com/>

⁸² <https://bg.pons.com/%D0%BF%D1%80%D0%B5%D0%B2%D0%BE%D0%B4>

⁸³ <http://notrial.bg/software/windows/sa-dictionary/>

готовият превод се извлича и се предлага за повторна употреба. По този начин няма нужда да се превеждат сегментите, които веднъж са били преведени (Тайбех 2008: 97), ако са преведени добре или ако преводът е подходящ за конкретната задача. Преводната памет може да бъде полезна за обогатяването на терминологичната база. Частичните съвпадения могат да помогнат при намирането на терминология, подобна на тази, използвана в терминологичната база. Ако например терминът „locate“ е в терминологичната база от данни и термините „located“ или „location“ трябва да бъдат преведени в изходния текст, терминът „locate“ ще бъде изходен за тях. Тогава преводачът може да въведе правилната форма.

Процесът по създаване на терминологични бази с помощта на преводна памет не е автоматизиран. Докато добавянето на всички сегменти (с цялото им съдържание) в паметта е задължителна стъпка, за да може да бъде завършен един превод в която и да е от системите за компютърнопомогнат превод, добавянето на термини е изцяло ръчен процес, който се основава на преценката или инструкциите на преводача. Самото добавяне е възможно, докато преводачът работи с текстовия редактор. Новите термини и техните преводи могат да се добавят към една или повече терминологични бази. Най-често терминът е дума или израз, който трябва да бъде преведен по един и същи начин в целия проект. При добавянето на термин директно от редактора, програмата го възприема като нов запис в терминологичната база. Някои терминологични бази имат допълнителното предимство да предлагат автоматични предложения за термин (automatic term suggestions). При тях програмата идентифицира възможните нови термини и ги предлага за добавяне на преводача.

За добра практика се възприема термините (съществителни, глаголи, прилагателни) да се добавят в основната си форма, т.е. в единствено число и с малки букви, дори и в текста да не се срещат така. Някои терминологични бази включват и настройки за засичане на малки и главни букви, както и настройки за съгласуване (concordance settings) и подобряване на предложенията за термини, което на свой ред намалява броя на фалшивите положителни резултати по време на проверката за качество.

IV. Създаване и управление на терминологичните бази в системите за компютърнопомогнат превод

Всяка съвременна система за компютърнопомогнат превод включва наличието на функции за управление на терминология. Точно както функцията на преводната памет, функцията за терминология следи дали в сегмента за превод има

съвпадение. Когато го открие, функцията предлага съответното визуализиране. Повечето системи прилагат и някакъв вид частично разпознаване на терминологията, като се отчитат и морфологични промени. Както при преводната памет по-голямото количество термини не винаги е по-добро решение: терминологичната база също трябва да се отнася (доколкото е възможно) до дадена тематична област, клиент или проект. Поради това е обичайна практика да се съставят множество временни бази, които могат да бъдат държани отделно за определени цели (и периодично да се събират в голяма обща терминологична база).

Начините за създаване на терминологична база варират според използваната програма, но общото между тях е, че всяка програма за компютърноподпомогнат превод има настройки, които позволяват на потребителите да създават колкото бази пожелаят. В Традос, например, създаването на нова терминологична база е стъпка от създаването на нов проект. В програмата преводачът има възможност да преведе отделен файл или да създаде проект. Преди да започне самата работа по превод, преводачът или избира една от вече съществуващите бази, или създава нова. Това става в нов прозорец, където стъпка по стъпка преводачът избира дали да добави файлове, онлайн бази и т.н. Допълнителни настройки дават възможността за бързо импортиране (Fast Import), където файлът за импортиране е напълно съвместим с формата Мултитерм ексемел (MultiTerm XML). По този начин се спестява времето за проверка на записите във файла за импортиране. Функцията за извършване на пълна реорганизация след импортиране (Perform full reorganisation after import) се използва при актуализирането на съществуващи терминологични бази с нови или променени термини. При този метод на работа не е изключено да се получат грешки след импортирането като нежелани дубликати или частичното търсене да не работи правилно. Тази опция предотвратява необходимостта от пълна ръчна реорганизация след импортиране⁸⁴.

Що се отнася до мемоКю⁸⁵, за да създаде нова терминологична база в програмата, потребителят трябва да отиде в прозореца за ресурси на програмата, където да намери бутона за създаване на нова база. След това следват настройките за език, клиент, име на базата, както и изборът на допълнителни настройки (например свойствата за търсене по подразбиране за нови термини или нуждата от одобрение при добавянето на нови думи).

⁸⁴ <https://gateway.sdl.com/articles/SolutionArticles/000001192>

⁸⁵ <https://docs.memoq.com/current/en/Plaaces/new-term-base.html>

IV.1. Притурки към основната програма

Освен стандартните функционалности, свързани с терминологичните бази в системите за компютърноподпомогнат превод, преводачите имат възможност да инсталират допълнителни помощни средства. Те не са част от основния софтуер, не всички програми за компютърноподпомогнат превод ги поддържат и много от тях са платени. Повечето се свързват към основната програма чрез плъгин, но някои могат да работят и самостоятелно. Пример за такива притурки са⁸⁶:

- Лоджитърм (LogiTerm)⁸⁷ е средство за свързване с терминологични ресурси. Позволява широк спектър от удобни за потребителя функции, като се започне от неговия интерфейс за търсене. Лоджитърм дава достъп до терминология, корпуси и референтни бази от данни, както и незабавен достъп до 40 от най-популярните източници на терминология в света. Пример за такива е Терминава (Terminava⁸⁸) – терминологичните бази на доставчика на граждански аеронавигационни услуги, терминологичната база на Университета в Отава. След стартиране на търсенето в една от базите, потребителят може да го продължи в друга.
- Джуреми (Juremy)⁸⁹ е онлайн инструмент за двуезиково търсене в базите данни Евролекс и ИАТЕ за 24-те езика на ЕС. Предназначен е за работния процес на преводачи, юристи лингвисти и устни преводачи, работещи с технически или свързани с ЕС текстове. Освен търсенето на точна фраза Джуреми предлага интелигентно търсене, като автоматично довършва фрази, съвпадения на корена на думата или автоматично търсене без ударение. Може също така да коригира често неправилно подравнените абзаци в документите на Евролекс, като допълнително намалява времето, изразходвано за проучване. Тъй като Джуреми се фокусира върху изследванията на терминологията на ЕС, инструментът може да осигури допълнителна специфична функционалност, фокусирана върху превода на ЕС, като сортиране и филтриране на метаданни на документи.
- Ердабълюес Апстор (RWS AppStore)⁹⁰ е хранилище, където преводачите могат да търсят различни допълнителни притурки и помощни средства.

⁸⁶ Всички примери за такива програми са взети от списъка на Нимзи: <https://www.nimdzi.com/tbs/>

⁸⁷ <https://terminotix.com/index.asp?content=category&cat=4&lang=en>

⁸⁸ <http://www1.navcanada.ca/logiterm/addon/terminav/termino.php>

⁸⁹ <https://juremy.com/about>

⁹⁰ <https://appstore.rws.com/>

Комбинацията от терминология и технологии не е нова. Констатира го Т. Кабре още в края на 90-те години на миналия век: информационните технологии и терминологията взаимодействат помежду си и се възползват от най-новите постижения в двете области; терминологите използват компютърните и информационните технологии в своите изследвания, а резултатите от терминологичните изследвания като речници, глосари и термини се използват в специализираната комуникация между експерти. Т. Кабре представя тази интерактивна връзка между информационните технологии и терминологията от диахронна гледна точка и разграничава четири етапа в нейната еволюция още в края на 90-те години (Кабре 1998: 160). Първият етап от въздействието на технологиите върху терминологията е представен от използването в терминологията на много прости компютърни програми като текстообработващи програми, програми за правопис и др. Вторият етап включва по-сложни приложения като бази от данни, електронни речници и инструменти за компютърнопомогнат превод. Следващият етап на еволюцията, описан от авторката, споменава работата по интерминологията (interminology) при обработката на естествените езици чрез приложения като лематизатори, конкордансъри и софтуер за статистически анализ. В описанието на последния етап на еволюцията е посочено използването на експертните системи, алгоритми и изкуствен интелект в специализиран езиков анализ (Кабре 1998: 161). Терминологите се подпомагат от „интелигентни“ инструменти като софтуер за извличане на термини, автоматични индексирани системи и програми за автоматично генериране на текст. През последното десетилетие различни автори изтъкват важни характеристики на това как модерните технологии като виртуална реалност, електронно обучение и облачно базирана терминология (cloud-based terminology) се прилагат към терминологичната работа. Може да се каже, че те илюстрират нов етап от еволюцията на терминологията, а именно виртуализацията на терминологията (Варга 2013: 28 – 29).

IV.2. Системи за управление на терминология

Системите за управление на терминология (Terminology Management Systems или TMS) са софтуерни програми, специално създадени за управление на терминология. Те са предназначени за събиране, поддържане и достъп до терминологични данни. Използват се от преводачи, терминологи, технически асистенти и различни други потребители. Могат да бъдат както независими от системите за компютърнопомогнат превод, така и да са част от тях. Когато първите терминологични бази са били разработени през 70-те години на миналия век, такива системи не са съществували и затова организациите, създаващи тези терминологични

бази, разработват свои собствени системи. През 80-те и 90-те години системите за управление на терминология започват да се появяват като част от програмите за компютърноподпомаган превод. Трябва да се отбележи, че компонентите на инструментите за компютърноподпомогнат превод обикновено нямат функции, необходими за разширени приложения на терминологични данни, тъй като са проектирани специално за използване от преводачи и писатели. Не може да се очаква, че системите за управление на терминология, предназначени за една цел, са подходящи за разработване и управление на терминологични ресурси за други цели. Създаването на конкретна програма е обвързано с тези, които ще я използват. Според К. Уорбъртън не трябва да има небалансирана структура, при която има повече налични полета за описание на така наречения „основен“ термин, отколкото за други термини (Уорбъртън 2015: 657). Подобна небалансирана структура се наблюдава в някои системи за управление на превода, където има редица полета за документиране на първия термин (като част на речта и дефиниция), но по-малко полета за друг вид информация (като синоними и абривиатура). Например вместо: част на речта, дефиниция, синоним и съкращение, трябва да има: част от речта, дефиниция, вид на термина, където полето позволява стойности като съкращение и акроним. Синоним не трябва да е необходима стойност, тъй като всички термини в даден запис са синоними един на друг по подразбиране (Уорбъртън 2015: 658).

Софтуерът за управление на терминологията се отнася до приложения, предназначени да обработват терминологията, необходима за текущ превод или вътрешни речници на клиента. Те трябва да се различават от пълноценните системи за управление на бази от данни, които позволяват на потребителите да създават общи или специализирани терминологични бази от данни.

Управлението на терминологията трябва да се прилага не само при самия превод, което често се случва и може да се третира като закъсняло, но още на предварителния етап по подготовката и организацията на превода (Гилбърт 2022).

Термините могат да бъдат импортирани в автоматичен режим (без човешка намеса) или полуавтоматичен режим (като преводачът избира най-подходящия термин, като приема или отхвърля това, което предлага системата). Така наречените системи за управление на терминологията често просто свързват приложения за управление на терминология или „автоматични речници“ с текстов редактор (Гаудек 2007: 270 – 271).

Годишният доклад на Нимзи⁹¹ за 2021 отбеляза, че секторът на системите за управление на терминология процъфтява от 2010 г., като всяка година се реализират десетки нови програми. Тази тенденция продължава и досега. Създават се нови компании, които разработват свои технологии, а вече утвърдените в областта надграждат своите платформи и системи за компютърноподпомогнат превод с цел да предложат иновативни решения, базирани на изкуствен интелект.

Първата група, описана в доклада, са програми, посветени на локализацията. Тяхната структура се различава от инструментите за компютърноподпомогнат превод. Чрез тях се автоматизира преводът на уебсъдържание, бекенд, мобилни приложения и файлово съдържание. Втората група са системи за управление на терминология, предназначени за юридически лица. Например, използването на стари юридически термини в нови документи може да предизвика объркване и липса на яснота. В допълнение много компании превеждат съдържанието си на няколко езика. Основните предимства на управлението на корпоративната терминология са по-добрата последователност на текстовете на компанията, спестяване на време за реализация на пазара, проверка на корпоративната терминология (което на свой ред може да доведе до намаляването на разходите, благодарение на намаленото време за създаване и проверка на терминологията). Основният принцип на третата група инструменти, предназначени за управление на терминология, е сканирането на всички уебстраници на даден сайт, разпознаването на преводимото съдържание (не всички компоненти от интернет страниците позволяват да бъдат превеждани) и извличането на това съдържание във формат, който след това да може да се локализира. Последната и най-многобройна група са системите за управление на терминология, които са част от системите за компютърноподпомогнат превод. Към тях се причисляват програми като мемоКю, Розета, Мемсорс и много други.

Трябва да се отбележи, че проблем може да представляват следните компоненти: достъпът до термините, кой ги проверява, както и различните формати на програмите. Една примерна ситуация е файл с термини, който се обменя многократно, след което може да не е възможно да се проследи кой какво е добавил. Друг проблем могат да бъдат файловете формати, ако даден файл се използва в различни инструменти. Най-често доставчиците получават терминологията на своите клиенти в Ексел. На доставчика се налага да конвертира файла към програмата, която използва. Ако клиентът иска да

⁹¹ <https://www.nimdzi.com/language-technology-atlas/#the-tms-arena>

добави нови термини, агенцията трябва да конвертира файла обратно към Ексел. Ако създателят на файла не е запознат с терминологията, той може да добави термини, които не са термини, а обикновени думи.

Поради развитието на технологиите може да се очаква, че десктоп версиите на програмите за компютърноподпомогнат превод нямат бъдеще в преводаческата област. Причините за това са няколко: едната е поддръжката (това са програми, които се изтеглят и инсталират на конкретен компютър, те трябва да бъдат поддържани чрез периодично обновяване на версиите). Другата е сложността им (ако не са снабдени с потребителски интерфейс, това създава нуждата да се отдели време, за да се научат потребителите как да ги използват). Интеграцията със системите за компютърноподпомогнат превод и обработката на екселски таблици (обикновено системите за компютърноподпомогнат превод не са пригодени да обработват Ексел) е следващата причина. Уеббазираните версии на програмите за обработка на терминология разрешават частично проблемите, създадени от десктоп версиите. При тях е необходима единствено интернет връзка, за да могат да бъдат достъпни. Естествено, те също могат да създават трудности. Например, трудност може да представлява интеграцията към системи за компютърноподпомогнат превод. Друг фактор е цената на такива програми. Също така много от тези програми за терминология не са предвидени за многоезикови бази. В повечето от тях има само изходен и целеви език, а подобна подредба не е подходяща, когато става въпрос за повече от един език. За разрешаването на всички проблеми, описани по-горе, Д. Гилбърт констатира, че се налага допълнителна стандартизация. Тъх форматът е изключително полезен, но не е достатъчен. Д. Гилбърт съветва всички производители на системи за компютърноподпомогнат превод да използват стандартизиран програмно-приложен интерфейс (API) за интеграция на терминология (Гилбърт 2022). Така инструментите за създаване на терминология, тези за управление и тези за съхранение могат да се използват доста по-лесно.

IV.3. Използване на контролирани езици

Както пише Джордж Стайнър, „преводът формално и прагматически се съдържа във всеки акт на комуникация, в създаването и възприемането на всевъзможните форми на значението, както в най-широкия семиотичен смисъл, така и в по-специфичните словесни обмени“ (Стайнър 2014: 18). В специализираните текстове последователността при използването на терминология е наложителна в целия текст, както и в различни документи от една и съща област. Това се изисква не само заради четимостта и непротиворечивостта (consistency), но и заради последствията при грешка:

неправилният превод например в областта на медицината или в правен договор би могъл да има сериозни последици за заинтересованите страни (Рутиер 2020: 41).

Много често изискванията на частните компании налагат промени както в работните процеси на доставчиците на езикови услуги, така и в значенията и употребата на термините. Правилното и последователно използване на подходящи термини е задължително условие за всички компании, които искат да бъдат последователни в начина, по който представят своите продукти и услуги. Това важи и за големите технологични гиганти като Фейсбук. През 2019 г. се изгражда речник на маркетинговите термини за Фейсбук. Вместо постепенно да се създава речник въз основа на термини, предоставени от създателите на съдържание, се създава основен речник на маркетинговите термини, който да може да се използва веднага, ускорявайки възприемането на тези термини. Основният речник съдържа записи на 33 езика. Следващата стъпка е интегрирането на речника в инструментариум, който да позволява автоматична обратна връзка при използването на терминологията. Крайната цел е създателите на маркетингово съдържание да имат възможност да проверяват термините на всеки етап от процеса на създаване на съдържание и да получават автоматична обратна връзка за правилното използване на термините: сигнал „не използвайте термина“ или сигнал, когато в текста се появи неодобрен вариант на термина (т.е. неправилно изписване, главна буква или тире)⁹².

Тук е мястото да се споменат контролираните езици. Те са предназначени да подобрят комуникацията между хората, особено за тези, които не говорят съответния език. Контролираните езици се използват за човешка комуникация, за техническа документация, за машинен превод и за семантични системи. Има специфични изисквания за дизайна на контролирания език и дори в рамките на една област често съществуват конкуриращи се формално еквивалентни езикови конструкции и не винаги е ясно коя от тях се прилага най-добре. Без съмнение са необходими сравнителни оценки в тази област, за да се определи кои конструкции са най-подходящи за конкретна потребителска група. Въпреки това контролираните езици се използват успешно в много индустрии през последните 20 години като метод за подобряване на четимостта на техническите документи и за улесняване на превода им на езика на техните клиенти. В такива ситуации авторската подкрепа е абсолютно необходима за създаването на текстове и документи, тъй като хората, които ги пишат, могат най-добре да преценят

⁹² <https://slator.com/how-facebook-scaled-terminology-management/>

дали едно изречение или параграф отговарят на правилата на контролирания език и дали даден израз принадлежи към одобрения комуникационен корпоративен стил, или не (Швитер 2015: 462).

„Контролираният речник е организираното подреждане на думи и фрази, използвани за индексване на съдържание и/или за извличане на съдържание чрез сърфиране или търсене. Към контролирания речник могат да бъдат добавени предпочитани термини и техни варианти, както и да има дефиниран обхват или да се описва конкретен домейн“⁹³. Контролираният речник е не само инструмент за подпомагане на съхранението и извличането на информация, но също така е част от естествения език, който се използва за комуникация както между хората, така и с автоматизираните технологии (Бейкър 2017: 5). Предимството на контролираните езици е, че чрез тях авторите имат възможност да създават документи, които са лесно четими и по-последователни по отношение на използвания речник и стил. Недостатък от използването им е, че творческата свобода намалява, но всъщност тя не е препоръчителна в специализирани и технически текстове.

V. Умения за работа с терминология и терминологични бази

Терминологичните бази в инструментите на компютърноподпомогнат превод предоставят функционалност, при която, ако изречение, което трябва да се преведе, съдържа термин, който е в базата, съответният запис автоматично се показва на преводача. Термините на изходния език могат да бъдат снабдени с информация за употреба, например кои термини да се избягват. Извън инструментите за компютърноподпомогнат превод потребителят решава дали и кога да потърси необходимата му информацията. Този подход се прилага в потребителски сценарии като контролирано авторство (Уарбъртън 2015: 655 – 656). Така обяснен, начинът на работа с терминологията по време на превод може да изглежда доста лесен. Въпреки това знанията, свързани с терминология (terminology skills) могат да се смятат за отделна компетентност (separate competence) на преводачите.

За да вършат работата си, терминологите използват терминологични бази и корпуси в комбинацията с други инструменти. Терминологите могат да бъдат преводачи, експерти по локализация, компютърни лингвисти и ръководители на проекти. Терминологичната работа е интердисциплинарна дейност, засягаща области

⁹³ https://www.getty.edu/research/publications/electronic_publications/intro_controlled_vocab/what.pdf

като лингвистика, информатика, когнитивна наука и документация. Работата на професионалните терминологи позволява комуникация на специализирани езици, превод на специализирани текстове и безпроблемно предаване на знания. Профилът на професионалния терминолог⁹⁴ обхваща знания и умения, свързани с познаването на принципите на терминологията (теоретични и практически), създаването и локализирането на терминологични ресурси за конкретни цели и групи, отлични езикови познания (на използваните езици), изследователски умения и способност за идентифициране на важна информация, умения за използване на терминологични бази, както и отличното владение на системите за управление на терминология, инструментите за компютърноподпомогнат превод, софтуера за машинен превод и инструментите за извличане на термини⁹⁵. Трябва също да имат познания за онтологии и семантични мрежи. Владенето на различни браузъри и опитът при дефинирането на критерии за изследване и управленските и комуникационни умения, също са умения, които терминологът трябва да притежава. Специализирани умения като трансфер на знания и преподаване, познаване на специални предметни области, способност за планиране и управление на процеси и проекти, управление на промените, способността за вземане и прилагане на стратегически решения в координация с други звена в организацията⁹⁶ също са необходими.

Дейностите, свързани с терминологията, не трябва да бъдат подценявани. Още по време на лекционните курсове по превод студентите трябва да имат възможност да се запознаят с терминологични източници, онлайн речници, енциклопедии, едоезични, многоезични, паралелни и съпоставими корпуси и да придобият умения, които ще им позволяват да използват тези ресурси ефективно (Сикора 2018: 18 – 19).

Все по-голямо количество университети предлагат възможности за специализация и допълнителна квалификация. Примери за такива са: аспекти на превода и устния превод, свързани с терминологията⁹⁷, терминологията и професионални нужди (където се разглежда ролята на терминологията в документацията, лексикографията, езиковата нормализация, езиковото обучение и лингвистичното инженерство)⁹⁸,

⁹⁴ <https://termcoord.eu/why-terminology/31318-2/>

⁹⁵ <http://clearwordstranslations.com/language/en/7-terminology-management-best-practices/>

⁹⁶ https://terminorgs.net/downloads/Terminology_skills.pdf

⁹⁷ Магистърски курс по терминология на Люксембургския университет https://www.en.uni.lu/studies/fhse/master_in_learning_and_communication_in_multilingual_and_multicultural_contexts/programme

⁹⁸ Онлайн програма за терминология към Института за приложна лингвистика, изследователски център към Университета Помпео Фабра <https://www.upf.edu/en/web/terminologiaonline>

развитието на основни умения за превод и постигане на умения в разпознаването на регистри на различни езици⁹⁹, разработването на изследвания в области като техническия, литературния и медийния превод, конферентния превод и управлението на терминологията¹⁰⁰.

Освен академичните институции има редица частни учреждения, които също предлагат обучения и специализирани курсове за терминология. Пример е програмата за сертифициране за превод за управление на терминология (Terminology Management Translation Certificate) на Термнет¹⁰¹. Организирането на и участието в конференции е друг начин за придобиване на специализирани умения в областта на терминологията. Един такъв пример е Трансфеър (TransFair¹⁰²), където се срещат представители в областта на специализирания превод и на висшите учебни заведения с цел да подготвят студентите за професионалния им живот.

Обобщение

В тази част на дисертацията се предлага описание на основните понятия, свързани с термините, терминологичните речници и терминологичните бази. Като най-адекватна от гледна точка на разработването на терминологични бази в дисертацията се приема дефиниция, според която „терминът е израз, принадлежащ към техническия език“.

Терминологичните бази са един от основните компоненти в системите за компютърноподпомогнат превод. Въпреки че наподобяват стандартните терминологични речници, терминологичните бази се отличават със значително повече възможности (нямат ограничения по отношение на обема и вида на съдържанието, могат да се съставят от всеки, който работи с инструментите за компютърноподпомогнат превод). Представени са приликите и разликите между традиционните терминологични речници и терминологичните бази, като се акцентира върху предимствата на терминологичните бази особено когато се работи със системи за компютърноподпомогнат превод. В допълнение са описани накратко известни

⁹⁹ Магистърска програма по превод и терминологични изследвания в Малтийския университет <https://www.um.edu.mt/courses/overview/PMTTFTT-2021-2-O>

¹⁰⁰ Център за изследвания на превода към Виенския университет <https://transvienna.univie.ac.at/en/about-us/>

¹⁰¹ www.termnet.eu

¹⁰² <https://etransfair.eu/>

терминологични бази, които могат да се използват във или извън системите за компютърноподпомогнат превод.

Показани са начините за създаване и управление на терминологични бази в системи за компютърноподпомогнат превод. Наред с това съществуват и притурки за работа с термини, които могат да се използват комбинирано със системите за компютърноподпомогнат превод. Използват се и отделни програми за работа с терминология, които работят независимо от системите за компютърноподпомогнат превод. При наличието на многообразни възможности за създаване и управление на терминологични бази все още съществена трудност създава конструирането на съдържателната страна на терминологичните бази, тъй като този процес изисква наличието на достоверна информация за термините, тяхното значение и правилния им превод. Последното е свързано и с използването на контролирани езици като възможност за поддържане на последователност и непротиворечивост при употребата на термини при превод.

ТРЕТА ГЛАВА

Полуавтоматично създаване на многоезикови терминологични бази

Увод	98
I. Създаване на терминологични ресурси в областта на компютърната терминология	
I.1 Особености на компютърната терминология	102
I.2 Речници с компютърна терминология	104
I.3 Автоматично извличане на термини: методи и добри практики	109
I.4 Инструменти за автоматично извличане на термини	113
I.5 Сравнение на инструментите за автоматично извличане на термини	114
II. Методика за полуавтоматично създаване на терминологична база в областта на компютърната терминология	124
II.1. Двуетиков корпус за извличане на термини	125
II.2. Извличане на термините от корпуса	126
II.3. Ръчна проверка и техническо оформление на терминологичната база	133
III. Описание на характеристиките на създадената терминологична база в областта на компютърната терминология	135
Обобщение	137

Настоящата глава представя методите за създаване на многоезикови терминологични ресурси, програмите и инструментите за извличане на термини и прави преглед на техническите им характеристики. Предложена е алтернативна методика за създаването им, като се характеризират нейните предимства и недостатъци.

По-конкретно, Трета глава е насочена към изпълнението на следните цели:

- Да се представят накратко особеностите на компютърната терминология, като стремежът е към обобщение от гледна точка на включването на компютърните термини в терминологични ресурси.
- Да се опишат начините за създаване на терминологични ресурси, като се започне от традиционните отпечатани терминологични речници, премине се през онлайн терминологичните речници и се достигне до автоматичното извличане на

термини и представянето на съществуващите инструменти за тази цел. Описанието на инструментите ще бъде съпътствано от сравнение на техните функционалности в рамките на определен период от време.

- Да се създаде методика за полуавтоматично създаване на терминологична база в областта на компютърната терминология и да се опишат конкретните етапи от работата по създаването на двуетикова терминологична база в областта на компютърната терминология (за английски и български език).
- Да се представят характеристиките на създадената двуетикова терминологична база.
- Да се предложи унифициран начин на работа при създаването на терминологични бази, основаващ се от най-добрите практики и вземащ под внимание възможностите и начина на работа на преводачите.

Увод

Популяризацията на технологиите Уеб 3.0, както и възможността за свободно създаване на съдържание спомагат за развитието на многоезичието в интернет (Лий 2016: 119).

Преводът от един език на друг е ключов за разпространението на съдържание на различни езици. Технологичният напредък води до създаването на нова терминология и на специализирани термини, които могат да бъдат специфични за определени области. Съдържание под формата на текстове, изображения, аудио и видео не винаги има едно и също значение или възприятие в различните езици и култури. Прост пример е позоваването на разстоянията в километри или в мили, което се възприема по различен начин от различна аудитория. Предлагат се и редица примери за неуместна употреба на термини в български, която по-скоро може да подведе, а не да подпомогне за разбирането: *select all* и *select none*, преведени като *изберете всички* и *изберете никой*, *preview* – като *визуализиране* и др. (Тодоранова и Тодоранова 2017: 213.)

Също така непрекъснато се появяват нови инструменти и програми за езикова обработка, включително за превод и компютърноподпомогнат превод. Информацията от форумите в областта на превода показва, че потребителите намират за много трудно да бъдат в крак с най-новите промени в технологиите. Както е посочено във въведението, изборът на правилния инструмент не е лесен за потребителите. През годините са направени множество публикации за терминологията и нейното използване при превод и са представени различни сравнителни прегледи на терминологичните инструменти.

Поради развиващата се среда, много инструменти, обсъждани в тези документи, са остарели или са престанали да съществуват.

Изборът на преводач, на технология за превод, обмислянето на нюансите в преводния текст и на представянето на информацията по най-добрия начин са много важни. Голяма част от документацията е написана на специализиран език, който включва терминология. Добре дефинираната терминология може да помогне на хората от различни области да общуват по-ефективно, както и да осигури последователен превод.

Известно е, че езикът се развива непрекъснато, като се влияе от политически, културни и икономически фактори. По същия начин и специализираната терминология в различни области се развива непрекъснато. П. Фабер констатира, че „ако се направи проверка на честотата на употреба на думите в публикациите, свързани с терминология, думата *динамика* ще бъде високо в списъка с ключови думи. Това е съвсем естествено, тъй като специализираният език е динамичен и представянето му също трябва да бъде такова“ (Фабер 2011: 9). Познанията ни за света около нас се задълбочават с напредъка на новите технологии. Развитието на специализираните области е свързано със създаването на нови термини и с промяна на съществуващите.

Една област, която привлича значително вниманието, е автоматичното извличане на термини (Килгариф и др. 2014; Коста и др. 2016). Общата цел на извличането на термини е да се идентифицира основният терминологичен речник на дадена специализирана област. Ако определянето на терминологията за дадена област се извършва ръчно, това е процес, който изисква сериозни усилия от много хора. По тази причина все повече се използва автоматично извличане на термини, което може да бъде последвано от ръчна проверка и корекция. Най-ранните методи за автоматично извличане на термини използват само езикова информация като морфо-синтактични модели, по-късно се разработват все по-сложни статистически методи, при които се прилагат различни видове честотни и колокационни алгоритми. На определен етап от развитието на автоматичното извличане на термини се появяват хибридни системи, съчетаващи и двата вида информация: езикова и статистическа.

Освен автоматичното разпознаване на термини изследователите се насочват към (автоматичното) идентифициране на информация, която е ценна за терминологите, например контекста. Корпусите са източниците, които предоставят огромно количество примери за това. Тяхното използване обаче може да бъде както предимство, така и затруднение при съставянето на терминологични речници на базата на контекста. От

една страна, да се разполага с повече информация за даден термин и неговата употреба, е полезно за разбирането на значението и по-ефективната комуникация, но от друга страна, ръчното преглеждане на големи количества текст може да бъде бавно и трудоемко.

Използването на корпусите като емпирично доказателство за употребата на термини е важно и за да се преодолеят затрудненията при ръчната обработка, като все повече навлиза автоматичното извличане на термини. Технологиите за автоматично извличане на термини не само позволяват на терминологите да идентифицират повече термини, отколкото би било възможно ръчно, но и използването на тези технологии повишава съответствието между термините в терминологичната база и в корпуса, чието съдържание трябва да отразяват. Също така автоматичното извличане на термини частично намалява субективността, защото така се включват термини, които действително се употребяват в корпусите.

Решението кои термини да се включат в терминологичната база, е основно предизвикателство и К. Уорбъртън препоръчва то да се основава на корпусни доказателства. Процесът по идентифициране на термините, включването им в терминологична база и добавянето на информация варира в зависимост от това дали подходът, който е избран, е тематичен (според този подход се приема, че термините и техните основни понятия са част от система, и следователно тяхното съществуване и значение могат да бъдат потвърдени само в сравнение с други термини в същата система), или така нареченият ad-hoc подход, при който се решава конкретен проблем, отнасящ се обикновено само до едно понятие и неговото назоваване (Уорбъртън 2015: 652).

Езиковите ресурси, подпомагащи превода, са различни за различните езици и тяхната достъпност е сравнително ограничена за така наречените „малки езици“. Глобализационните процеси се движат в две основни направления: от една страна, те са наднационални (пример за това е все по-големият брой говорещи английски език¹⁰³), а от друга, налагат адаптирането към местните езици и култури¹⁰⁴ (негативните последици от глобалните тенденции пораждаат обратен ефект за предпочитане на „своето“ пред „чуждото“). Дж. Хаус отбелязва, че количеството английски думи в блокове или телевизионни реклами на други езици е забележително, но не може да се каже, че

¹⁰³ Данните са според Етимолог: <https://www.ethnologue.com/guides/ethnologue200> [20.07.2022]

¹⁰⁴ Пример за това е създаването на специални процеси като локализацията.

запълва лексикална празнина в приемащия език (и следователно може да се приеме за ненужно). В същото време редица думи в приемащия език „стратегически“ се заменят с английски, за да се постигне определен ефект (Хаус 2015: 99). „Когато определен език стане доминиращ в дадена държава, това има последствия за останалите езици, които се говорят в тази държава: как те запазват своята идентичност. Когато един език стане глобален, подобни последици засягат всички езици“ (Кристал 2004: 41). Според същия автор трябва да се приеме, че „идентичността на даден език ще се променя – както винаги е било в миналото“ (Кристал 2004: 100). Езикът на следващото поколение никога не е същият като езика на предишното. Приемането на този неизбежен факт е съществена част от всеки реалистичен подход към многоезичието (Кристал 2004: 101).

Езиковите ресурси за малки езици се намират трудно. Например, за езици като уелски, фарьорски или галицийски езиковите ресурси и технологии са малко или съвсем недостатъчни. Следователно прехвърлянето на езикова информация или знания от един език към друг е важен изследователски проблем. При по-голяма част от междуезиковия трансфер (*cross-lingual transfer*) се използва английският като изходен език. Има две причини за това: наличието на значителен брой ресурси на английски език и наличието на паралелни данни за (и преводи между) английски и повечето други езици (Съгард и др. 2015: 1713). Конкретно за български език в свое проучване от 2012 г. Д. Благоева, С. Коева и В. Мурдаров констатира отчетливото интернационализиране на българската лексика, най-вече под влиянието на английския език. „В редица терминологични области специализираната лексика се изгражда основно под влиянието на английския език – компютърни технологии и интернет (*файл, сайт*), финанси, икономика и бизнес (*дилър, брокер*), съвременна музика (*диджей, техно, клип*), спорт (*джогинг, бодибилдинг*). Навлизането на английски заемки се наблюдава и при общоупотребимата лексика, например *тостер, стикер, бодигард*, както и в младежкия жаргон“ (Благоева, Коева и Мурдаров 2012: 14 – 15). Авторите констатира още, че „чрез усъвършенстване на машинния превод с помощта на езиковите технологии биха могли да се превъзмогнат езиковите бариери, но само за тези езици, които съумеят да оцелеят в дигиталния свят.“ (Благоева, Коева и Мурдаров 2012: 2). Към 2022 година изводът, който може да се направи, е, че все още има съществена разлика между езиковите технологии за големите езици като английски, немски, френски, испански и за такива като български език. Автоматичната обработка за големите езици е с много по-добро качество в сравнение с тази за български език. Езиковите технологии бележат значителен напредък през последните десет години. Разликата обаче между езиците с развити езикови технологии

и езиците със слабо развити езикови технологии се запазва и през 2022 година (Коева и Стефанова 2022: 2).

I. Създаване на терминологични ресурси в областта на компютърната терминология

Определянето на „правилната“ терминология е свързано с разработването на терминологични ресурси и с тяхното управление. То е съществена част от процеса на превод и локализация. Преводачите използват терминология, за да направят преводите си по-точни и последователни. Подходящата и актуална терминология подобрява комуникацията и следователно подобрява ефективността и помага за избягване на културни и/или езикови недоразумения.

I.1 Особенности на компютърната терминология

Компютърната терминология е част от ежедневието на по-голяма част от хората или поне що се отнася до най-разпространените термини (*компютър, таблет, смартфон* и под.). Доброто разбиране на терминологията и жаргона, свързани с компютрите, помага за по-голяма ефективност при взаимодействието с други технологии. Отличителна черта на компютърната терминология е постоянният напредък и развитието на технологиите, което изисква и осъвременяване и допълване на използваните термини (Хойзъл и др., 2015). Навлизането, разпространението и употребата на компютърна терминология не е нов феномен. Още в началото на новото хилядолетие Л. Кирова констатира, че: „Компютърната лексика има невероятно развитие за период само от двайсетина години. Колкото повече нараства броят на потребителите на компютърна техника, толкова по-определящо става тяхното участие във формирането на езика за работа с нея“ (Кирова 2002). Д. Николова също отбелязва, че „актуалността на темата се определя от бурното развитие на компютърните технологии като знаково постижение на науката и техниката, глобализацията на обществото, разширяването на межкултурните и междуезиковите контакти, които водят до интензивното развитие на компютърната терминология във всеки един от съвременните езици“ (Николова 2018). Авторката констатира още, че скоростта, с която се променят почти всички области от човешката дейност, оказва влияние върху обществото, а познаването на компютърната терминология помага за ускоряване при решаването на редица проблеми, свързани с теоретичните и практическите приложения на изследваната терминология (Николова 2018: 5).

Компютърната терминология навлиза много бързо в езика и постоянно се развива заради бързото развитие на информационните и комуникационните технологии и заради появата на нови и подобряването на вече съществуващи технологии. Това прави областта актуална за изследване.

По своята структура термините могат да бъдат отделна графична дума (*прозорец, мишка*), но също така могат да бъдат сложни думи (една графична цялост, образувана от две или повече изходни думи) (*киберсигурност, киберпрестъпност*) или съставни думи (*твърд диск, операционна система*), които по своята структура приличат на свободните словосъчетания, но обозначават само едно понятие както отделните думи (Коева 2021: 55). Както се посочва, използването на голямо количество термини, които по своята структура могат да се класифицират като сложни или съставни думи, се дължи на „необходимостта понятието да се назове максимално прозрачно и изчерпателно, вътрешната форма да е видима и разбираема“ (Зидарова 2018: 58).

При термините, които са съставни думи, най-често срещаната синтактична структура е структурата от подчинено прилагателно и главно съществително (Коева 2007: 61) (например *риболовен сезон*), сравнително често се срещат и структурите от главно съществително, предлог и съществително, образувани подчинена предложна група (например *подобряване на почвата*), всяко едно от съществителните – главно или подчинено може да се пояснява от прилагателни (например *електронен трансфер на фондове*), в някои случаи прилагателните може да са повече от едно (например *европейска парична система*), срещат се и словосъчетания от две съществителни, в които второто е главна част (например *тенис корт*).

Термините също така могат да бъдат класифицирани като неутрални от гледна точка на стилистичната си употреба (например *търсеица система*), професионализми или такива, които се споделят от определена група специалисти в дадена област (например *метатърсачка*), и професионален жаргон (например *паяк*) (Кирова 2004).

Абревиатурите се срещат често в (компютърната терминология). Както се посочва: „заимстваните от английски абревиатури започват да функционират като пълнозначни думи и все по-често се налага да влизат в тълковните речници“ (Кирова 2004). Посочва се, че се наблюдават различни случаи на предаване на чужда абревиатура на български език – в някои случаи се възприема съкращението (*джиесем, есемес*), а в други може да се използва както абревиатурата, така и разгърнатото название (*сиди, компактдиск*) (Колковска 2010: 2 – 3).

I.2 Речници с компютърна терминология

В мрежата и книжарниците могат да бъдат намерени редица речници на английски и български в областта на компютърната терминология. Примери за такива са:

- *Компютърен речник A-Z*¹⁰⁵ (хартиено издание от 2003 г., заради остарялата информация не може да се използва като актуален източник);
- *Английско-български компютърен речник: A-Z*¹⁰⁶ (хартиено издание от 2017 г., вероятно би могло да се използва за проверка на отделни термини, но хартиеното издание го прави трудно за създаването на терминологична база);
- *English-Bulgarian Computer Terms*¹⁰⁷ (база от данни, включваща 729 термина на английски и преводите им на български. Базата е подходяща за употреба, но преводачите трябва да имат информация за нея, за да може да я използват);
- *Преводно-тълковен речник на някои компютърни термини*¹⁰⁸ (сайт, в който не са посочени източниците на информация и няма цитирана литература; не е подходящ за създаването на терминологична база, защото преводачите няма как да знаят дали термините наистина са част от компютърната терминология; също така, макар че английските термини са снабдени с превод на български, дефиниция на български и някои примери, описанието не винаги е издържано от лингвистична гледна точка, а броят на включените термини не е голям);
- *Речник на основните термини в компютърната система*¹⁰⁹ (сайт с термини от компютърната област; както и при този по-горе, няма цитирани източници и съответно – данни за достоверността на информацията; броят на включените термини не е голям);
- Николова Д. *Словарь компьютерных терминов / Речник на компютърните термини*. Шумен: Химера, 2016. 258 с. ISBN 978-619-7218-27-5. Речникът представя популярни термини в сферата на информационните технологии: устройство на компютъра, оборудване; компютърни технологии; програмни продукти; работа с компютър; интернет; съдържа определения и обяснения на някои професионални термини; английските термини са представени с руските и

¹⁰⁵ <https://m.helikon.bg/43115-Компютърен-речник-A-Z.html> [20.07.2022]

¹⁰⁶ <https://www.book.store.bg/p5397/anglijsko-bylgarski-kompiutyren-rechnik-a-z.html> [20.07.2022]

¹⁰⁷ <https://elrc-share.eu/repository/browse/english-bulgarian-computer-terms-processed/f9d2a4e44afb11e9a7e100155d02670615329808b1384ab388941fcd4fba5029/> [20.07.2022]

¹⁰⁸ <http://kafence.com/Преводно-тълковен-речник-на-някои-компютърни-термини-201.html> [20.07.2022]

¹⁰⁹ <https://shogo.eu/blog/2008/10/30/речник-на-основните-термини-в-компютър/> [20.07.2022]

българските си еквиваленти, съответно 1094, 1216 и 1192 за всеки един от трите езика; отбелязани са и жаргонните единици – доста голям брой: 606 за руски и 420 за български. Втората част се състои от азбучни указатели, в които са изброени руските и българските термини и жаргонните единици от първата част, позволяващи на практика да се използва триезиков (английско – руско – български) речник; неудобство е хартиеният вариант на речника, както и фактът, че разбирането за компютърната терминология е много широко).

Нормално е речниците да имат различен обхват, различна структура, както и за едни и същи термини да бъдат дадени различни дефиниции и др. Отпечатаните речници с ясно показано авторство са по-авторитетен от гледна точка на качеството на включената информация. От друга страна, дори да се преиздават, хартиените речници остаряват много бързо. Проблемът с речниците в интернет е несигурността в тяхното качество, липсата на информация как са подбрани единиците, настина ли принадлежат към компютърната терминология, правилно ли са представени техните значения.

Поради тези причини събирането на едноезикови речници на английски и български в областта на компютърната терминология също може да помогне при създаването на терминологични бази. Примери за речници на английски в областта на компютърната терминология:

- *Компютърни термини в речника КОЛИНС (COLLINS Computer terms)*¹¹⁰ (речникът се актуализира редовно, включени са компютърни термини от няколко области: термини, използвани при работата с компютър, части на компютъра, видове компютри; към термините е приложена дефиниция, например: backup ‘a file or set of files copied for security purposes’);
- *Компютърна терминология на АйБиЕм (Computer Terminology of IBM)*¹¹¹ (последната актуализация на този речник е била на 03.03.2021 г.; речникът включва термини за АйБиЕм и други продукти; разработен е от Международната организация по стандартизация и Международната електротехническа комисия (ISO/IEC JTC1/SC1); речникът съдържа дефиниции, информация за част на речта и е с добро качество, но не всички термини, които са включени в него, могат строго да се дефинират като принадлежащи към компютърната терминология, например: *активен прозорец (active window)* същ. „Прозорец, в който може да се

¹¹⁰ <https://www.collinsdictionary.com/word-lists/computer-terms-used-in-computing> [20.07.2022]

¹¹¹ <https://www.ibm.com/ibm/history/documents/pdf/glossary.pdf> [20.07.2022]

въвежда информация от клавиатурата. Отличава се с уникалния цвят на заглавната лента и рамката на прозореца“ (“The window that can receive input from the keyboard. It is distinguishable by the unique color of its title bar and window border”) е компютърен термин за разлика от *administrivia* „Всякакъв вид бюрократична документация, отнасяща се до IBM или свързана с друго, която възпрепятства постигането на нечии цели“. ‘Any kind of bureaucratic red tape or paperwork, IBM or not, that hinders the accomplishment of one’s objectives or goals’.

- *Компютърни термини, речник и глосар* (Computer terms, dictionary, and glossary)¹¹² (сайтът е основан през 1998 г. от Нейтън Ембъртън и предлага безплатна техническа поддръжка и информация; списъкът с над 15 000 компютърни единици и дефиниции позволява използването на поле за търсене, преглеждане по букви или проследяване на връзките; тъй като количеството на термините е голямо, те са разпределени в множество подкатегории и някои от тях в много широк смисъл се отнасят към компютърната терминология, например *дейлимоушън* (dailymotion) „Дейлимоушън е уебсайт за споделяне на видеоклипове, който работи от 2005 г. и през 2014 г. събира 2,5 милиарда гледания на видеоклипове всеки месец. Сайтът позволява на потребителите да публикуват лични видеоклипове, които могат да бъдат намерени чрез търсене по ключови думи или тагове. Видеоклиповете могат да бъдат достъпни и чрез разглеждане на създадени от потребителите групи от видеоклипове“ ‘Dailymotion is a video sharing website that began operation in 2005 and as of 2014 receives 2.5 billion video views every month. The site allows users to upload personal videos, which viewers can find by searching for keywords or tags. Videos may also be located by browsing user-created groups of videos’;
- *Дефиниции на компютърни и интернет термини* (Definitions of Computer and Internet Terms)¹¹³ (Всички дефиниции в речника са написани от Пер Кристенсон, който има над дванадесет години опит в областта на информационните технологии; речникът е онлайн от 1999 г., а дефинициите са актуализирани и преработени няколко пъти оттогава; предлагат се примери и дефиниции, като дефинициите в някои случаи имат енциклопедичен характер, например: Капча (Captcha) Онлайн формулярите могат да включват капча, за да се предотврати

¹¹² <https://www.computerhope.com/jargon.htm> [20.07.2022]

¹¹³ <https://pc.net/glossary/> [20.07.2022]

подаването на данни от автоматизирани програми. „Капча е тест, който определя дали потребителят е човек, или робот. Обикновено включва изображение на изкривен текст и поле за писане, в което потребителят трябва да въведе текста. Капча може да се намира в края на формуляри към уебсайтове и трябва да се попълни, за да може формулярът да бъде изпратен...“¹¹⁴. Online forms may include captchas to prevent automated programs from submitting data. ‘A captcha is a challenge-response test that determines whether a user is human or an automated bot. A typical captcha includes an image of distorted text and a form field for the user to enter the text. Captchas are commonly found at the end of website forms, and must be filled out in order for the form to be submitted...’);

- *Глосар с компютърна терминология (Glossary of computer terminology)*¹¹⁴ (сайтът е разработен специално за възрастни хора и има за цел да улесни по-възрастното население, като обяснява значението на термините; по тези причини броят на включените термини е относително малък, а дефинициите нямат структурата на обичайните лексикографски дефиниции, а по-скоро представляват кратки описания, показващи функцията, предназначението и пр.);
- *От А-Я за технологичните термини (A-Z of Technology Terms)*¹¹⁵ (съдържа неголям брой термини, които, макар че са дефинирани като технологични, могат да се определят и като компютърни, и техните дефиниции).

Макар че това представяне вероятно не обхваща всички налични речници, представящи компютърна терминология на английски език, се очертават следните тенденции: речниците, създадени от големи издателства или организации, съдържат богата и достоверна информация; наред с това съществуват множество източници в интернет, които предлагат информация, но само специалисти могат да удостоверят в някои случаи дали термините и дефинициите са подходящи.

За български също са достъпни речници за компютърната терминология, които обаче са електронни:

- *Компютърни термини, които трябва да знаем*¹¹⁶ (кратък списък с термини, представящ най-често използваните названия, свързани с компютрите; самите термини в повечето случаи са представени на английски, което не показва как

¹¹⁴ <https://www.ageuk.org.uk/wp-assets/globalassets/leeds/computing-glossary-pdf-version.pdf> [20.07.2022]

¹¹⁵ https://www.ourcommunity.com.au/tech/tech_article.jsp?articleId=74 [20.07.2022]

¹¹⁶ <https://urocibg.eu/компютърни-термини-които-трябва-да-зн/> [20.07.2022]

трябва да бъдат изписани правилно на български, а дефинициите са от енциклопедичен тип);

- *Речник на компютърните термини и акроними*¹¹⁷ (съдържа над 250 термина и техните дефиниции, но освен малкия брой, други недостатъци, които могат да бъдат споменати, са: представяне на част от термините само на английски и (известна) неточност на дефинициите, например: блутут (Bluetooth) ‘спецификация, която позволява мобилните телефони, компютри и PDA да се свързват безжично през къси разстояния’). Използваната дума „спецификация“ не дава ясна информация, че става въпрос обикновено за лична мрежа.
- *Компютърни термини и понятия*¹¹⁸ (портал за Андроид с новини за устройства с тази операционна система; речникът съдържа допълнителна информация и вероятно по тази причина някои термини не са формулирани добре, а дефинициите приличат на получени посредством автоматичен превод, например Основна (вътрешна, произволен достъп) памет (RAM) на компютъра ‘подредена последователност (масив) от байтове или машинни думи (клетки на паметта)’;
- *Речник на компютърни термини по азбучен ред. Основни компютърни термини. Програмата и нейното инсталиране на компютър*¹¹⁹ (речникът се намира в сайт с новини за информационни технологии, отново някои термини и дефиниции не са предадени подходящо на български език, а също така се наблюдават и повторения на грешките от други източници, като не е ясно кой е първоизточникът, например: основна (вътрешна, произволен достъп) памет (RAM) на компютъра ‘подредена последователност (масив) от байтове или машинни думи (клетки на паметта)’.

Речниците с компютърна терминология на български език се отличават с малък брой термини, липса на представителност и авторство, некоректно представен в много случаи от гледна точка на българския правопис термин или липса на представяне на български език и дефиниции, които също може да не представят адекватно описаното понятие, а в някои случаи има ясни индикации за машинен превод.

Въпреки че някои от изброените речници са полезни, те могат да затруднят преводачите при тяхната работа поради различния си обхват, различната структура,

¹¹⁷ <https://bg.chalized.com/речник-на-компютърните-термини-и/> [20.07.2022]

¹¹⁸ <https://droidd.ru/bg/kompyuternye-terminy-i-ponyatiya-slovar-kompyuternyh-terminov.html> [20.07.2022]

¹¹⁹ <https://roidmod.ru/bg/lighting/slovar-kompyuternyh-terminov-po-alfavitu-osnovnye-kompyuternye-terminy-programma-i-e-ustanovka-n/> [20.07.2022]

както и най-вече за това, че за едни и същи термини са дадени различни дефиниции (дори преводачите да са наясно за качеството на даден речник, който използват). Още един фактор, който прави съществуващите речници не винаги подходящи за създаването на терминологични бази, е специализираната насоченост на конкретния превод. Такива са например специализираните преводи в медицината или изкуствения интелект, за чиято изработка са използвани компютърни технологии. Преводачът, който първи се сблъска с тях, има на разположение само съществуващи ресурси, където термините, свързани с новите технологии, или още не са включени, или ако са включени, може да са с различно значение. Технологиите напредват изключително бързо, което прави доста трудно отпечатването или публикуването в интернет на актуални речници, представящи компютърната терминология. Дори актуализирането на онлайн версиите, макар да е значително по-лесно спрямо отпечатването на издания, е трудно за поддръжка.

Тук се представят само речници за български и/или английски, а задачите за превод далеч не се ограничават до тези два езика. С други думи, заключението е, че няма (а и може би не е възможно да има) речници, представящи адекватно компютърната терминология на два или повече езика, които отразяват използването им в тясно специализирани и най-вече нови области. преводачите трябва да се насочат към използването на новите технологии за генерирането на терминологични бази, които са подходящи за дадена специфична задача за превод.

I.3 Автоматично извличане на термини: методи и добри практики

След като за даден език не съществува необходимият езиков ресурс, то той трябва да бъде създаден. Един начин за създаването на списъци с термини е автоматичното извличане на термини. Методите за автоматично извличане на термини варират. Най-простата възможност е експерти терминологи да подберат нужните термини от съществуваща документация. Друга възможност е при вече налична терминологична база в Ексел, файлът да бъде конвертиран според различните изисквания на системите за компютърнопомогнат превод. Наличието на вече съществуваща преводна памет и възможността за съотнасяне между изречения в текстовете могат да бъдат изброени като други възможни подходи в създаването на терминологични бази.

Автоматичното извличане на термини става от езикови корпуси, които трябва да представляват дадена тематична област. Търсенето и колекционирането на подходящи текстове може да не е лесна задача, защото от тях зависи и качеството на извлечените термини. За нуждите на автоматичното извличане на термини също така автоматично или полуавтоматично се събират големи количества документи. Както беше посочено, за

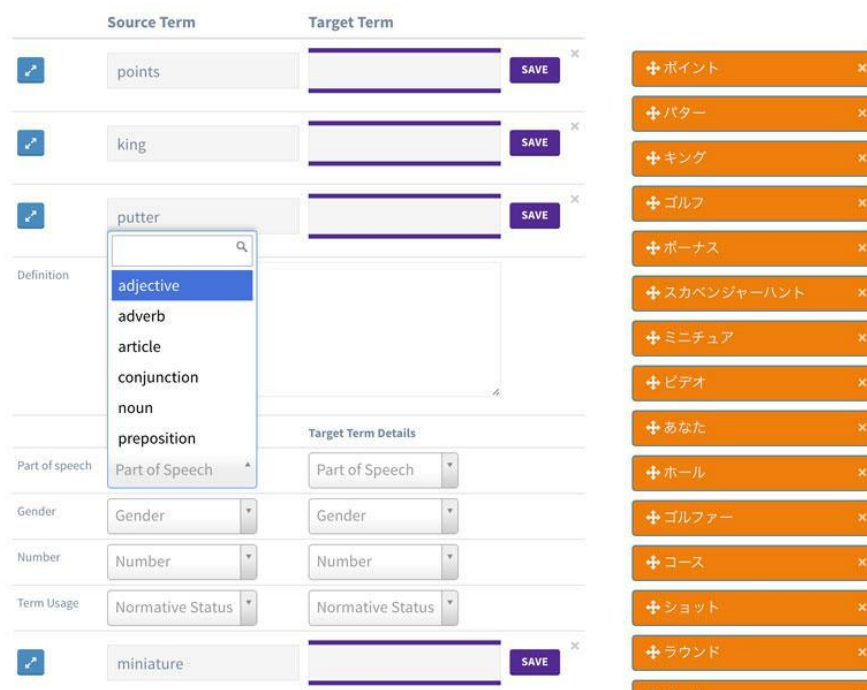
някои езици достъпните ресурси в електронна форма може да не са в големи количества. Някои методи за извличане на термини изискват и голям референтен корпус, който да илюстрира общоупотребимата лексика и сравнението с който да позволява адекватната идентификация на термини. Извличането на термини в два и повече езика е свързано с подравняването на паралелни изречения, думи и фрази в паралелни корпуси.

Въпросът за автоматичното извличане на термини не е нов. Още през 1996 г. К. Кагеура и Б. Умино правят преглед на принципите и методите за автоматично разпознаване на термини (Кагеура и Умино 1996). Те показват две основни тенденции в извличането на термини: едната е извличането на информация (information retrieval), а другата е автоматичното разпознаване на термините (automatic term recognition). А. Шаятович и др. разделят методите за извличане на термини на: честотни методи (които се основават на предположението, че по-високата честота на кандидата за термин предполага и по-голяма вероятност той да е действително такъв); контекстни методи (основават се на контекста, в който се появява терминът кандидат); тематично моделиране (тематичната информация може да се използва вместо честотата) и повторно ранкиране (re-ranking) (методите от тази група използват други методи за извличане на термини като характеристики и се стремят да оценят важността на всяка характеристика) (А. Шаятович и др. 2019: 150).

Някои инструменти използват статистически подход към текста, така че те просто търсят текстовите повторения. По този начин обаче има риск да бъдат извлечени невалидни кандидати за термини (известни са като „шум“, noise). Друг подход при извличането на термини е лингвистичният, според който програмата търси структури като „съществително + съществително“ или „прилагателно + съществително“ и т. н.

Н. Астраханцев разработва ATR4S софтуер с отворен код, който включва повече от 15 метода за автоматично разпознаване на термини и прилага целия процес от предварителна обработка на текстови документи до събиране на кандидати за термини, оценяване на кандидати за термини и класирането им. ATR4S извлича последователни n-грами (поредица от n на брой токъни), по подразбиране от 1 до 4, като кандидати за термини, след което програмата комбинира лема, например „обработка на_информация“, които принадлежат към един и същ кандидат за термин. Последната стъпка е филтрирането на кандидатите за термини, които се появяват по-рядко от предварително зададен брой (по подразбиране – 2), за да се намали шумът преди обработката на данните (Астраханцев 2016).

Друг пример за извличане на термини е инструментът на Тием таун¹²⁰ (TM-Town), платформа за сътрудничество между преводачи и клиенти (Фигура 15). Програмата автоматично извлича ключови термини от изходните и целевите текстове. Тези термини са оценяват според алгоритъм, който изчислява уместността на всеки термин в документа. Това става чрез статистически алгоритми за обработка на естествения език, които анализират съдържанието на документите и идентифицират най-подходящите ключови думи. По-високият търм ранк резултат (Term Rank) показва по-висока важност на термина в документите. След извличането на основните термини от целевия текст потребителят съпоставя изходния термин със съответния му превод. Възможно е също и добавянето на допълнителни данни към терминологичната база (например част на речта на термина, друга граматична информация и др.). Програмата поддържа български език.



Фигура 15: Интерфейсът на Тием таун

През последните години навлизат технологии, основаващи се на изкуствен интелект и машинно самообучение. М. Хосари и др. в своя статия от 2019 г. предлагат нова система, която се занимава с проследяването на развитието на новите технологии и различните области на изкуствения интелект. Системата се прилага за автоматично откриване на текстове, в които се представят нови технологии, и за извличането на

¹²⁰ <https://www.tm-town.com/> [20.07.2022]

термини, използвани за описание на тези нови технологии. Предложеният от тях инструмент се основава на двуетапен каскаден модел (cascading model). Първият етап на този модел класифицира дали изречението съдържа термин, или не, а вторият идентифицира дума или фраза в изречението, свързана с технология (Хосари и др. 2019). За да създадат модела, авторите събират етикетирани набор от данни, където технологичните термини (от една или повече думи) са ръчно анотирани (в приблизително 500 статии от различни онлайн източници).

Въпреки че автоматичното извличане на термини се радва на интерес от научните среди от дълго време, трудностите при тази задача не са намалели. През 2020 година А. Р. Терин и др. представят ТърмЕвал 2020 (TermEval) – споделена задача, предоставяща на изследователите възможността да работят върху автоматичното извличане на термини, като използват едни и същи данни. Предлага се корпус, който съдържа специализирани подкорпуси на три езика (английски, френски и нидерландски) в четири тематични области (корупция, конна езда, сърдечни заболявания и вятърна енергия), които са ръчно анотирани за съответните термини (Терин и др. 2020: 84). Целта е чрез предоставянето на качествено анотирани и разнообразни данни да се даде на изследователската общност, изучаваща автоматичното извличане на термини, възможност за идентифициране на силните и слабите страни на различни съвременни методи в тази област. Резултатите илюстрират как някои методи достигат по-висока точност и как отделните системи извличат различни видове термини, как някои са изключително добри в намирането на редки термини или по-малко се влияят от дължината на термините (Терин и др. 2020: 85).

При повечето системи за извличане на термини от два езика се идентифицират термините кандидати в изходния език въз основа на предварително дефинирани модели. След това се избират еквивалентни термини кандидати в превода на целевия език. Един от алтернативните подходи е генерирането на кандидат термини директно от подравнените думи и фрази в корпуса. Чрез честотна информация (frequency information) и n-грами се определя кои думи или съставни думи са термини. При този подход не се вземат под внимание зависими от езика предварително дефинирани модели, а се свързват фрази, основаващи се на лексикалните съответствия и синтактичното сходство (Лефевер 2009: 496).

Необходимо уточнение е, че въпреки че автоматичното извличане на термини е изключително полезно, то не винаги е подходящо да се използва от преводачите, тъй като е направено от машина, която не „разбира“ човешкия език.

I.4 Инструменти за автоматично извличане на термини

Извличането на термини може да бъде направено както чрез външни за системите за компютърноподпомогнат превод програми, така и чрез вградени в тях приложения, като качеството и нивото на успеваемост варират (и в зависимост от езика). Целта на инструментите за извличане на терминология (независимо от техния вид) е да помогнат на потребителите да изграждат терминологични ресурси по (полу-)автоматичен начин. Нуждата от такива средства се дължи най-вече на нарастващата потребност от управление на информацията при превод, което изисква автоматизирана помощ при изпълнение на задачи, свързани с терминология (Зарецкая и др. 2015). Според същото проучване други предимства от използването на инструменти за извличането на терминология са спестяването на време и възможността за определяне на различни критерии за търсене.

В някои инструменти (например Мултитърм¹²¹ и Кютърм¹²²) е възможно да се види езиковата и статистическата информация за термина, както и контекстът, в който се използва. Не всички инструменти разполагат с такъв набор от функции. Освен по своята функционалност инструментите се различават и по средата на работа: могат да бъдат разделени на самостоятелни (десктоп версии, за които е необходим инсталационен процес), уеббазирани (работят с браузър) и вградени в инструментите за компютърноподпомогнат превод.

Методите за извличане на термини могат да бъде използвани с текстов корпус, който да бъде представителен за областта, към която принадлежи използваната терминология. Това не винаги е възможно особено когато става въпрос за малък език с ограничена наличност на достъпна информация в електронна форма. Например, когато една компания прави опис на своята вътрешна терминология, текстовият корпус представлява колекцията от документи на компанията. Когато обаче се анализира цялостната терминология за дадена област, като „морска биология“ или „аеронавтика“, съставянето на корпус задължително трябва да включва текстове от тази области (Хейлърн и Хертъг 2015: 205).

По-долу е представен инструмент за извличане на терминология, който е свободен достъпен:

¹²¹ <https://www.trados.com/products/multiterm-desktop/>

¹²² <https://docs.memoq.com/current/en/memoQWeb-help/mqw-qterm.html>

- Файлфилтърс (fivefilters.org¹²³) предлага функционалност за извличане на терминология. Това става чрез поставяне на текста за обработване или на интернет връзката към него в определеното за това място. Възможно е извличането на най-много 100 термина в пет различни формата. Потребителят има възможност да определи от колко думи да се състои даден термин, както и дали софтуерът да проверява за малки и главни букви. Този инструмент поддържа и български.

Посоченият по-горе инструмент е само част от прегледаните безплатни инструменти от този вид. Количеството им е значително по-голямо, но са ограничени откъм поддръжка на български език. Понеже българският език е малък (по отношение на достъпните електронни документи и програми за обработка на езика), много от инструментите за извличане на термини не го поддържат. Това прави такива програми неизползваеми от българските преводачи.

I.5 Сравнение на инструментите за автоматично извличане на термини

В проучване от 2016 А. Зарецкая и др. (2016: 18) проверяват доколко инструментите за извличане на термини наистина са полезни. Те правят съпоставка между техническите им параметри.

Избират да анализират следните самостоятелни програми, които определят като „най-популярния тип софтуер“ към момента:

- Мултитърм екстракт (MultiTerm Extract¹²⁴) е компонент на ЕсДиЕл мултитърм (SDL MultiTerm) – търговски инструмент за управление на терминология, който предоставя възможността за съхраняване, извличане и управление на многоезикова терминология. Съществува и като самостоятелно приложение и може да се интегрира в Традос. Това е един от малкото инструменти, създадени специално за използване от преводачи, и е може би най-известният инструмент за извличане на терминология в преводаческата индустрия. Мултитерм екстракт поддържа всички езици, поддържани от Уиндоус 8.1/10.
- Симпъл екстрактор (Simple Extractor) е търговски инструмент за извличане на термини, разработен от Дейл софтуер (DAIL Software S.L.¹²⁵). Това е джава приложение, което извлича автоматично най-често срещаните термини (думи и съставни думи) на английски, португалски, испански, френски и руски. Предлага

¹²³ <https://www.fivefilters.org/term-extraction/> [20.07.2022]

¹²⁴ <https://appstore.sdl.com/language/app/sdl-multiterm-extract/1165/> [20.07.2022]

¹²⁵ <https://www.dail.es/en/artificial-intelligence/> [20.07.2022]

опция за зареждане на списъци и разширена функционалност за търсене, която позволява търсене във вече извлечения списък с термини, всички контексти, в които се появява определен термин, и редактиране на намерените термини.

- Търмсют (TermSuite¹²⁶) е инструмент за извличане на термини с отворен код, написан на джава. Той е разработен в рамките на проекта ТиТиСи (ТТС: Извличане на терминология, инструменти за превод и съпоставими корпуси), чиято цел е да се създаде инструмент, способен да извлича двуезикова терминология от съпоставими корпуси. Потребителят може да избира от няколко възможности за подравняване, максимален брой кандидати за превод за даден изходен термин, използване на мерки за сходство за сравняване на контекста на термина в изходния и целевите езици, наред с други разширени настройки. Не поддържа български.
- Скеч енджин (Sketch Engine¹²⁷) е онлайн инструмент, създаден от Лексикал компютинг (Lexical Computing Ltd¹²⁸) за изграждане, обработка и управление на корпуси, което включва и извличане на терминология. Достъпът до него е възможен при платен търговски или академичен лиценз. Този инструмент осигурява възможност както едноезиково, така и многоезиково извличане на термини. Предлагат се различни възможности за търсене. Резултатът може да бъде изтеглен в .tbx или .cvs формат. При многоезиково извличане на термини трябва да се предостави .tmx файл с паралелен корпус, подравнен на ниво изречение или абзац. Терминологията първо се извлича за всеки език поотделно. При втората стъпка системата търси двойки кандидати, които най-често се локализируют в паралелните документи. Поддържа български.
- Транслейтед (Translated srl¹²⁹) предлага услуги за превод посредством уеббазиран инструмент. Резултатът от системата включва 20-те най-добри термина, подредени по тяхната релевантност. Освен това термините са представени като хипервръзки към съответните резултати от търсене в Гугъл. Под списъка с термини инструментът показва всички термини в техния контекст в рамките на изречение. Програмата е платена. Не поддържа български.

¹²⁶ <http://termsuite.github.io/> [20.07.2022]

¹²⁷ <https://www.sketchengine.eu/> [20.07.2022]

¹²⁸ <https://www.lexicalcomputing.com/> [20.07.2022]

¹²⁹ <https://translatedlabs.com/terminology-extraction>

- Терминус (Terminus¹³⁰) е уеббазирано приложение за управление на корпуси и терминология и може да бъде достъпно чрез покупка на лиценз. Целта на този инструмент е да интегрира пълния процес на терминографска работа (търсене в корпус, компилация и анализ, извличане на термини, речник и управление на проекти, създаване и поддръжка на бази от данни и на речник). Кандидатите за термини се определят посредством извличане на n-грами, извличане на биграми с асоциативни мерки, ключови думи и ръчна проверка на съответните резултати.
- Киа (Kea¹³¹) е алгоритъм за автоматично извличане на словосъчетания (ключови фрази) от текста, които го характеризират. Алгоритъмът идентифицира кандидатите за ключови фрази, изчислявайки стойностите на характеристиките на всеки кандидат и използва алгоритъм за машинно обучение, за да предвиди кои кандидати са добри ключови фрази. Въпреки че алгоритъмът е езиковонезависим, се използва програма за определяне на основата на думите и списък със стоп думи, които са специфични за всеки отделен език.
- Рейнбоу (Rainbow¹³²) е инструмент за извличане на терминология с отворен код, написан на джава, базиран на статистически методи за автоматично извличане на термини от множество файлове и формати. Не поддържа български.
- Джава аутоматик търм екстракшън тулkit (Java Automatic Term Extraction Toolkit или JATE¹³³) е инструмент за разработване на и експериментиране с алгоритми за автоматично извличане/разпознаване на термини. Не поддържа български.

Като цяло, инструментите за извличане на термини обещава големи ползи. Със сигурност са удобни за използване програмите, които разполагат с потребителски интерфейс и са достъпни онлайн. Много от тях са платени. По отношение на български език, той се поддържа от Мултiterм, Скеч енджин и Тим таун.

Поради големия брой инструменти за извличане на термини може да има затруднения при оценката на приликите и разликите между тях: дали наистина са полезни и ако да, доколко. На тези въпроси се опитват да отговорят А. Зарецкая и др. 2016, като разглеждат характеристиките на най-известните програми за обработка на терминология, изброени по-горе. Безспорно предимство на програмите, освен че се

¹³⁰ <http://terminus.iula.upf.edu/cgi-bin/terminus2.0/terminus.pl?Int=En>
131

<https://dl.acm.org/doi/abs/10.1145/313238.313437#:~:text=Kea%20is%20an%20algorithm%20for,%2C%20robust%2C%20and%20publicly%20available.> [20.07.2022]

¹³² <https://okapiframework.org/wiki/index.php/Rainbow> [20.07.2022]

¹³³ <https://javalibs.com/artifact/uk.ac.shef.dcs/jate> [20.07.2022]

спестява време, е възможността да се задават различни критерии за търсене, което позволява адаптирането на заявката за търсене към конкретна задача. По този начин потребителят може да получи различен вид информация и да филтрира резултатите в зависимост от това какво търси. Подходът при сравнението се основава на анкета сред професионални преводачи, посредством която се отчитат предпочитанията на преводачите към функционалностите, които се използват за извличане на термини и терминологични бази, и предпочитанията на преводачите се сравняват с поддържаните от най-популярните инструменти функционалности. По-конкретно, сравняват се следните функционалности:

Двезиково извличане (Bilingual extraction) е функционалност, която се отнася до извличане на термини от изходен и преводен текст едновременно.

Сравняване на контекста в изходния и целевия език (Source and target context comparison) е функционалност, която проследява предходния и следходния сегмент.

Проверка на термин (Terms validation) е процес, чрез който се проверява дали дадена дума или израз е верен.

Двезиково попълване на речник (Bilingual dictionaries compilation) е процес, чрез който се създава двезиков речник.

Контекстно извличане на термини (Context extraction) е функционалност, позволяваща извличане на термини, употребени в определен контекст.

Поддръжка на различни файлови формати (Support various file formats) е функционалност, която позволява използването на различни формати. Към тази функционалност се причислява също възможността за импорт на терминологични бази, създадени с други инструменти.

Честотно подреждане на термини (Rank terms by frequency) позволява термините да бъдат подредени според честотата им.

Поддръжка на различни езици (Support for many languages) е функционалност, която позволява поддръжката на различни езици.

Определяне на минимален брой срещания на термин (Specify the minimum number of occurrences) дава възможност да се зададе допълнителен критерий при търсенето във или при създаването на терминологична база, според който конкретен термин трябва да е наличен минимум определен брой пъти.

Показване на лингвистична информация (Show linguistic information) е функционалност, която позволява към всеки термин да се визуализира лингвистична информация като част на речта и др.

Определяне на минимален и максимален брой преводи (Specify the maximum number of translations) е опция, която позволява да се определи минималният и максималният брой възможни преводи, които да се съотнесат с даден термин.

Опция за изключване (Stopword list option) позволява конкретни често срещани думи (например предлози) да се изключват по преценка на преводача или на работещия със системата.

Избор на минимален и максимален брой думи в състава на термин (Choose the minimum and maximum number of words per term) позволява потребителят на програмата да определи от колко думи максимум и минимум може да се състои даден термин.

Статистика (Term statistics) е статистическата информация, отразяваща броя на срещане на термините.

Таблицата по-долу възпроизвежда наблюденията на А. Зарецкая и др. (2016: 5)¹³⁴.

	ЕсДи Ел Мулт итър м	Сим пъл Екст ракт ор	Тър мС юит	Ске ч Енд жин	Тра нсл ейт ед	Тер мин ус	Киа	Рей нбо л	ЯТЕ
Двуетиково извличане	х		х	х					
Срвяняване на контекста в изходния и целевия език	х								
Проверка на термин	х	х		х		х	х	х	х
Двуетиково попълване на речник	х		х						
Контекстно извличане на термини	х	х	х		х	х	х	х	х
Поддръжка на различни файлови формати	х	х	х	х		х	х	х	х

¹³⁴ <https://wlv.openrepository.com/handle/2436/622550> [20.07.2022]

Честотно подреждане на термини	x	x	x	x		x	x	x	x
Поддръжка на различни езици	x		x	x		x	x	x	x
Определяне на минимален брой срещания на термин	x	x		x		x	x	x	x
Показване на лингвистична информация	x		x			x			
Определяне на минимален и максимален брой преводи			x						
Опция за изключване	x	x			x		x	x	x
Избор на минимален и максимален брой думи в състава на термин	x	x					x	x	x
Статистика	x	x	x	x	x	x	x	x	x

Таблица 2: Сравнение на техническите характеристики на програмите за извличане на термини, направено през 2016 година от А. Зарецкая и др.

По модела на направеното проучване от А. Зарецкая и др. през 2022 г. бяха анализирани същите инструменти, за да се провери дали има нови функционалности, и ако има, какви са промените през последните шест години. Новите функционалности на програмите за извличане на терминология отговарят на технологичното развитие на инструментите за компютърноподпомогнат превод. Такива са:

Многоезиково извличане на термини (Multilingual term extraction) е функционалност, позволяваща извличането на термини на повече от два език.

Проверка на съответствието между текста и терминологичната база (Consistency check between the text and the termbase) помага за проследяване дали ако текстът е преводен, е използвана терминологията от съответната терминологична база.

Възможност за ръчно извличане на термини (Manual terms extraction) позволява допълнително ръчно извличане на термини.

Настройки към проект (Project settings) е функционалност, позволяваща към всеки проект за превод да бъдат зададени различни настройки.

Извличане на отчет за качество (QA report) е доклад, който програмата генерира за преведения текст.

Едноезиково извличане (Monolingual extraction) е функционалност, която може да извлича термини само от изходен или от преводен тест, т.е. само от един език.

Работа с външен софтуер (Third-party software access) позволява програмата за управление на превода да се свърже с външни програми.

Определяне на максимален брой извлечени термини (Select Maximum number of extracted terms) е функционалност с която може да се зададе точен брой термини, които да бъдат извлечени.

Двуетиково подравняване (Bilingual alignment) представлява съотнасяне на най-добрия превод на термин от изходния език на целевия език.

Възможност за търсене с регулярни изрази (Regex search) позволява търсене чрез регулярни изрази.

Концептуална структура (Conceptual structure) е функционалност за създаване на концептуално дърво, за да се структурират термините на даден речник.

Със знака Y са маркирани функционалностите, които дадена програма не е имала през 2016 година, но вече са налични през 2022 година (Таблица 3).

	ЕсДи Ел Мулт итерм	Сим пъл Екст ракт ор	Тер мс юи т	Скеч Енд жин	Тра нсл ейт ед	Тер мин ус	Киа	Рей нбо л	ИА ТЕ
Сравняване на контекста в изходния и целевия език			x						

Проверка на термин	x	x		x		x	x	x	x
Двуетиково попълване на речник	x		x						
Контекстно извличане	x	x	x	Y	x	x	x	x	x
Поддръжка на различни файлови формати	x	x	x	x		x	x	x	x
Честотно подреждане на термините	x	x	x	x		x	x	x	x
Поддръжка на различни езици	x		x	x		x	x	x	x
Определяне на минимален брой появявания за даден термин	x	x		x		x	x	x	x
Показване на лингвистична информация	x		x	Y		x			
Определяне на минимален и максимален брой преводи	Y		x					Y	
Опция за изключване	x	x		Y	x		x	x	x
Избор на минимален и максимален брой думи в състава на термин	x	x			Y		x	x	x
Статистика	x	x	x	x	x	x	x	x	x
Многуетиково извличане	x		x						

на термини									
Проверка на съответствието между текста и терминологичната база	x								
Възможност за ръчно извличане на термини	x								
Настройки към проект*	x					x		x	
Извличане на отчет за качество	x							x	
Едноезиково извличане	x		x	x					x
Работа с външен софтуер	x		x	x	x			x	x
Определяне на максимален брой извлечени термини	x		x					x	
Двуетиково подравняване			x	x					
Възможност за търсене с регулярни изрази				x				x	x
Концептуална структура						x			

Таблица 3: Сравнение на техническите характеристики на програмите за извличане на термини, направено през 2022 година

* Настройки в зависимост от проекта зависят от това дали инструментът за извличане на термини е част от инструмент за компютърноподпомогнат превод, или е самостоятелна програма, както и от това дали става въпрос за извършване на дейности, които не са свързани с превод, например проверка на качеството или създаване на речник.

В своята книга от 2021 г. К. Уорбъртън констатира налагането на интегрираните опции за извличане на термини: „Въпреки че интегрираните инструменти не обслужват всички нужди на потребителите си, много от тях не осъзнават това и избират функционалности в рамките на вече закупения инструмент за компютърноподпомогнат превод само за да открият по-късно, че тяхната терминологична база не е оптимизирана за други цели извън превода“. К. Уорбъртън определя като най-популярни следните инструменти за извличане на термини, интегрирани в системи за компютърноподпомогнат превод (Уорбъртън 2021: 163):

- Мултитерм (на Традос);
- Търмстар¹³⁵ (на Стар Транзит¹³⁶);
- Кростърм¹³⁷ (на Акрос¹³⁸);
- Кютърм¹³⁹ (на мемоКю);
- Търмбейс (на Мултитранс).

В същата книга авторката анализира методите за създаване на терминологични бази, като съветва да се използват най-голям брой филтри (Уорбъртън 2021: 184), основаващи се на различните компоненти на базата, за да може да се извлича релевантното за конкретен потребител или проект съдържание. Филтрите за търсене се използват от терминологите при създаването на бизнес анализи с цел отчитане на напредъка на проекта, както и за идентифицирането на области от терминологичната база, които се нуждаят от допълнителна работа. Причина за използването на филтри е улесняването на работата на преводача. Например, филтър за търсене, който идентифицира всички термини, които нямат изображение или превод на един от езиците (ако базата е многоезикова). При нужда могат да бъдат приложени няколко ограничения при визуализирането на съществуващите термини.

Въпреки наличието на значително количество инструменти за разпознаване, управление и извличане на терминология, най-добрият и сигурен начин за гарантиране на създаването на качествена терминологична база е човешката дейност. Ръчното извличане на термини има предимството, че терминологът има достъп до оригиналните текстове (както на изходния, така и на целевия език) и може да разглежда термините в

¹³⁵ <https://www.star-group.net/en/products/terminology-management.html> [20.07.2022]

¹³⁶ <https://www.star-group.net/en/home.html> [20.07.2022]

¹³⁷ <https://www.across.net/en/knowledge/expert-features/how-to-use-crossterm-instances-efficiently> [20.07.2022]

¹³⁸ <https://www.across.net/en> [20.07.2022]

¹³⁹ <https://www.memoq.com/extensions> [20.07.2022]

техния контекст. По този начин може да прецени дали са термини в действителност, или не. Разбира се, недостатъкът на този метод на работа е, че ръчното извличане може да отнеме много време, което е и причината за разработването на всички описани по-горе инструменти и програми. Друг потенциален недостатък е субективността. Само индивидуалната оценка на терминолога или преводача е определяща за това какво е термин. Въпреки това е необходимо да се отбележи, че ръчната проверка се извършва и след използването на инструмент за автоматично извличане. Автоматично генерираният списък с потенциални термини по подразбиране се нуждае от ръчна проверка, защото все още никоя машина не може да прецени достатъчно точно (както човек) дали извлечените думи или групи от думи могат да се смятат за термини.

II. Методика за полуавтоматично създаване на терминологична база в областта на компютърната терминология

Както стана ясно от описаното дотук, съществуват голям брой инструменти за извличането на терминология. Те варират според лицензионния си модел (безплатни или платени), броя на поддържани езици и вида на програмата (онлайн или десктоп приложение). Български не се поддържа от голяма част от тези програми. Положението е още по-сложно, когато стане въпрос за някоя специализирана област: наличните езикови ресурси за български също не са много. Затова е избрана езиковата комбинация английски-български. Наличието на езикови ресурси за английски език е значително по-голямо от тези за български. Избраната област за изследване е компютърната терминология. Технологиите постоянно напредват и е неизбежно да има нови термини и промяна в значението на вече съществуващи. Ограничението при събирането на паралелни текстове и липсата на инструменти за български език са причина да се предложи методика за полуавтоматично създаване на терминологична база в областта на компютърната терминология. Целта е да се предложи работеща процедура, която да може да бъде използвана от преводачите в България.

Необходимо уточнение е, че тази процедура не претендира за всеобхватност и изчерпателност, но цели, доколкото е възможно, да предложи работеща методика, която да улесни преводачите по време на тяхната работа. Обикновено на преводачите се възлагат различни проекти, в които тематиката варира (или ако е сходна, може да варира между различните клиенти), при това проектите трябва да бъдат приключени в определен срок. Отговорите в анкетата, представена в Първа глава, посочват (с няколко изключения), че терминологични бази се създават ръчно, обикновено при конкретен

превод, Всичко това води до заключението, че дефинирането на процедура, която може да бъде следвана при създаването на терминологични бази, може да бъде от полза при ежедневната работа на преводачите. В тази връзка оптимално би било разработването на програма, специално предвидена за български терминологични ресурси, която да бъде със свободен достъп (или символичен абонамент) и интуитивен интерфейс. Подобна програма ще ограничи, ако не елиминира изцяло все още неизбежната ръчна работа при създаването на терминологични бази. Към момента опити в тази насока са направени от Секцията по компютърна лингвистика към Института за български език. Тяхната програма позволява извличането на термини от документи за български и английски¹⁴⁰.

II.1. Двезиков корпус за извличане на термини

Д. Костадинова¹⁴¹ представя корпус от 159 термина, насочени към компютърната терминология. Въпреки че е обещаващ, корпусът не е подходящ за настоящото изследване, защото е прекалено малък.

Опус (Opus)¹⁴² е колекция с публичен достъп от преводни текстове на много езици. Текстовете са анотирани и подравнени по изречения. Цялата предварителна обработка е извършена автоматично и не са въвеждани ръчни корекции. За български и английски в интернет са налични и други паралелни корпуси, но компютърната терминология като област ограничава възможностите.

Както вече стана ясно, първата стъпка в създаването на една терминологична база е намирането на достоверен източник. Предвид описаните по-горе ограничения и особености при извличането на термини, езиковият корпус е най-подходящият ресурс за подбор на термини.

За целите на тази дисертация бе избран паралелният английско-български корпус в областта на компютърната терминология КюТиЛиип (QTLeap V1.2)¹⁴³, който съдържа 4 000 двойки въпроси и отговори в областта на отстраняването на проблеми, свързани с информационните технологии (както хардуерни, така и софтуерни). Материалът за корпуса е събран чрез услуга, използваща чат. Записани са само взаимодействия, съставени от един въпрос и съответния отговор. Лицензът за употреба на корпуса е позволява академична и некомерсиална употреба, изисква цитиране на

¹⁴⁰ <https://dcl.bas.bg/termextraction/>

¹⁴¹ <https://core.ac.uk/download/pdf/326573787.pdf> [20.07.2022]

¹⁴² <http://opus.nlpl.eu/> [20.07.2022]

¹⁴³ <http://metashare.ilsp.gr:8080/repository/browse/qt leap-corpus-v12/0176c39ae9cd11e4a2aa782bcb074135ba7d767f645a48dca1d50ee3c9504253/> [20.07.2022]

авторството и същия начин на споделяне (CC-BY-NC-SA). Създаден е в периода 2014 – 2015 година с европейско финансиране. Редактиран е за последно през 2016 година. Корпусът е избран, защото е двуезиков и отговаря на подбраната терминологична област. Трябва да се отбележи, че има няколко недостатъка: корпусът е малък и текстовете и термините в него не са разнообразни. Трябва също да се отбележи, че броят на специализираните термини не е голям, а се срещат по-скоро термини, които чрез разпространяването на технологиите в ежедневието, са навлезли в употреба.

II.2. Извличане на термините от корпуса

Задача на настоящото изследване е да се опита, доколкото е възможно, да се предостави работеща методика за извличане на терминологична база за нуждите на преводачите. Избрана е гледната точка на преводачите, защото средствата за компютърноподпомогнат превод са фокусът на изследването, както и защото, както стана ясно от цитираните изследвания, ресурсите за малък език като българския са крайно недостатъчни. В отговорите си в анкетата преводачите описват създаването на терминологични бази посредством екселски таблици, подравняване, записване в тетрадки, търсене на нужните термини по време на превод. Някои от тях използват вградените функционалности на програмите за компютърноподпомогнат превод. Настоящото изследване ще се опита да помогне с решаването на повдигнатите проблеми, свързани със създаването на терминологични бази.

Сред относително голямото количество от инструменти за извличане на терминология, основно разграничение, което е свързано с употребата им, е дали са платени или безплатни.

Ключов компонент за това дали дадена програма за извличане на термини ще намери приложение, е дали поддържа български, което обаче към момента (особено за свободностъпни инструменти) не е често срещано. Програми като Мултитърм поддържат огромно количество езици, но са платени и тяхното използване е пряко обвързано с лицензите за Традос. По този начин, използването на Мултитърм е ограничено до тези преводачи, които разполагат с лиценз.

Другият ключов компонент е наличието на ресурси за български. Не винаги има налични корпуси с нужната тематика. След намирането на подходящ за преводния проект корпус, следващата стъпка е извличането на термини от него. Умението да се борави с термините също е ключов компонент.

От анкетата към българските преводачи става ясно, че методите, които те използват за създаването на терминологични бази варират: повечето от отговорилите

използват вградените функции на програмите за компютърноподпомогнат превод (добавяне на нови термини по време на превод във вече съществуващите бази, някои използват Мултитърм или Кютерм), други използват програми на Майкрософт като Уърд и Ексел, а трети пишат в тетрадки и добавят нови термини само когато се сблъскат с тях по време на превод. Може да се заключи, че преводачите в България предпочитат да боравят с инструменти, които познават, до които имат достъп и за които не губят време, за да ги разучават.

Поради всичко описано до този момент, както и предвид останалите отговори от анкетата, в този труд ще предложим следния общ подход за извличане на термини и създаване на терминологични бази: преводачът намира източниците, от които има нужда, извлича необходимите термини, конвертира ги в желанния формат и използва новосъздадената терминологична база.

Необходимо уточнение е, че преводаческите проекти варират по обем, изисквания и срокове. Определянето на универсален метод за работа не е част от целите на този труд.

За да бъде избран конкретен инструмент за извличане на термини, той трябва да отговаря на следните изисквания:

- да поддържа български език и другия език от/на който ще се превежда (задължително);
- да бъде безплатен (препоръчително);
- да бъде самостоятелна програма или вградена функционалност на (безплатна) система за компютърноподпомогнат превод (задължително за хора без допълнителни умения за програмиране);
- да бъде с интуитивен интерфейс (препоръчително);
- да притежава настройки за избор на минимален и максимален брой думи, от които да се състои даден термин (препоръчително);
- да предлага извлечените термини във формат, който позволява лесното им конвертиране. (препоръчително).

Много малко инструменти за извличане на термини предлагат поддръжка на български. Изискването инструментите да са безплатни, е, защото много от програмите за компютърноподпомогнат превод са платени, а техният лиценз осигурява и възможността за употреба на съответните функционалности за извличане на термини. Не всеки преводач има достъп до инструментите за компютърноподпомогнат превод. Интуитивният интерфейс е наложителен, за да може и неспециалист да работи с

избраната програма. Възможността за конвертиране между различни файлови формати е второстепенна, защото най-популярните инструменти за компютърнопомогнат превод могат да работят с различни формати.

Въпреки наличието на огромен брой инструменти за извличане на терминология, намирането на такъв, който да отговаря на посочените изисквания, е доста трудна задача.

След направеното проучване се установи, че инструмент, който отговаря на посочените критерии, е Файффилтърс. По-долу са показани резултатите от извлечените термини с Файффилтърс. Програмата се справя с извличането на английски термини от корпуса. За български предлага различни резултати според това дали е избрана настройката за термини, които се състоят от една дума, или от повече думи (например три). Във втората колона от Фигура 16 са използвани настройки за извличане по подразбиране, както при английските термини и са разпознати са само продуктите имена, написани с латински букви.

Английски			Български			Български с ограничен брой думи за термин		
Term	Occurrence	Word count	Term	Occurrence	Word count	Term	Occurrence	Word count
computer	197	1	В LibreOffice	7	2	Google	8	1
change	174	1	Google	8	1	Chrome	5	1
laptop	91	1	text box	5	2	Windows	4	1
use	89	1	Когато отварям Chrome	2	3	pin	3	1
Facebook	85	1	Имам рутер Linksys	2	3	add-ons	3	1
Skype	70	1	Какво е VPN	2	3	computer	3	1
network	65	1	Какво е PHP	2	3	AutoRecover	2	1
way	62	1	Какво е Linux	2	3	filesystem	2	1
password	55	1	блокирано от Kaspersky	1	3	Transfer	2	1
PC	55	1	Какво е Panda	1	3	XP	2	1

Фигура 16: Резултати от извличането на термини с Файффилтърс

Както се вижда от Фигура 16, въпреки че програмата е езиково независима, резултатите за български не са задоволителни. Друг експеримент с програмата включва редуциране на максималния брой думи в състава на даден термин до 1. Както се вижда от трета колона на фигура 16, програмата разпознава само имената на продукти и

Предвид описаното по-горе, бе променен подходът за извличане на термини, като описаните критерии бяха насочени към работа с комбинация от програми, а не с отделен инструмент. За целта като първа стъпка бе избрана програма, поддържаща извличане на термини от документи на английски език. Термостат Уеб 3.0¹⁴⁵ (TermoStat Web 3.0) е програма за автоматично извличане на термини от няколко езика, включително от английски. Програмата извлича единични термини и термини съставни думи, показва тяхната честота, степента на специфичност и формите на думите, които се срещат в текста. Изображенията по-долу представят интерфейса на програмата, както и извлечените от корпуса термини. Изредени са предложенията за термин, честотата им в корпуса, специфичността им и техните варианти.

Програмата е разработена през 2003 година от Патрик Друан и се отличава с хибридна техника за извличане на термини от технически корпуси. Основната цел е намаляването на шума в списъка с кандидати за термини (Друан 2003). Независимо че програмата е разработена преди почти 20 години, голямо количество учени я използват в своите експерименти: например Ортего Антон и Фернандес (2015)¹⁴⁶ за експертно представяне на знание в двуезикови електронни речници; Р. М. Уюн (2018)¹⁴⁷ за ключовите думи в романите на Халед Хосейни „Ловецът на хвърчила“ и „И планината отекна“; К. Варгас Касерес (2020)¹⁴⁸ за предварителен анализ на корпус с текстове на студенти. Друга свободнодостъпна самостоятелна програма с уебинтерфейс за извличане на термини от английски с голяма степен на точност не беше открита.

¹⁴⁵ <http://termostat.ling.umontreal.ca/interfaceTermostat.php> [20.07.2022]

¹⁴⁶ <https://dialnet.unirioja.es/servlet/articulo?codigo=5126623> [20.07.2022]

¹⁴⁷ <https://repository.unair.ac.id/79103/> [20.07.2022]

¹⁴⁸ https://d1wqtxts1xzle7.cloudfront.net/70692341/libro_de_resumenes_actas_iii_cilcc_2020_y_v_wopatec_2020_virtual-with-cover-page-v2.pdf?Expires=1658270074&Signature=epNBqAYOglkY0TfJ4ySfryKOLqa09gqBLWYVRMnECocZrRPye mrY8epi0gg2zjzFx0KWm9X0O3Kup-KvkrOain8dHWamewPKcEfrVzk8oJfMNJj-d5FeYGLB2StJ2109F22lxWgajQfF2A5FV7xucJFwxNqeRBqUwaX2v6JU1XaJWqzZ34hFuvxyyAuCkFJ6b6I pJwLCgo68XolaW8E-CdP5gh4S2IK7e1HaubmBveoF66o3HZCc1TTugsfXCbGQHvAzf7KYSFxzBYR6IkM-QFzINE1c5MbqIJrCToVEyORQqVFjaDT6CaWJy25mDIGU5ZXrXqSWf6UMIhsNpkksA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA#page=183 [20.07.2022]

Corpus >> qlcpcorpus_q_en_v1.2				
Résultats				
Liste des termes viki Aide Déconnexion				
Usage Statistiques Structuration Diagrammes				
Candidat de regroupement	Fréquence	Score (Spécificité)	Variantes orthographiques	Matrice
facebook	136	145.96	facebook	Nom
laptop	123	136.52	laptop laptops	Nom
internet	110	130.55	internet	Nom
iphone	91	119.17	iphone	Nom
password	90	114.7	password passwords	Nom
wireless	93	113.3	wireless	Nom
router	80	111.65	router routers	Nom
ipad	74	107.33	ipad	Nom
email	73	106.59	email	Nom
file	182	104.81	file files	Nom
computer	258	99.66	computer computers	Nom
pc	109	96.86	pc pcs	Nom
google	58	94.84	google	Nom
folder	58	85.6	folder folders	Nom
itunes	43	81.41	itunes	Nom
wireless network	42	80.44	wireless network wireless networks	Nom Nom
hard drive	37	75.37	hard drive hard drives	Adjectif Nom
powerpoint	36	74.32	powerpoint	Nom
website	34	72.16	website	Nom activate Windows
wifi	33	71.06	wifi	Nom to Settings to activate Windows.
antivirus	32	69.94	antivirus	Nom
usb	31	68.81	usb	Nom

Фигура 18: Част от извлечените термини от корпуса с програмата Термостат

Както се вижда от Фигура 18, програмата извлича термините, показва тяхната честота, специфичност, определя към коя част на речта принадлежат и извлича техните форми, които се срещат в текста. Заложените критерии при извличането са: да се извлекат само съществителни (в това число и съставни), да се извлече максималният брой разпознати от програмата термини. Терминът с най-голяма честота (136) е Фейсбук (впоследствие обаче търговските марки бяха премахнати ръчно при създаването на терминологичната база). Общият брой термини извлечени от програмата е 618, като преобладават единичните съществителни (378 на брой), съставните термини от две думи (223 на брой) и съставни съществителни с три думи (15 на брой). Резултатът е достъпен в Приложение 2 на този труд.

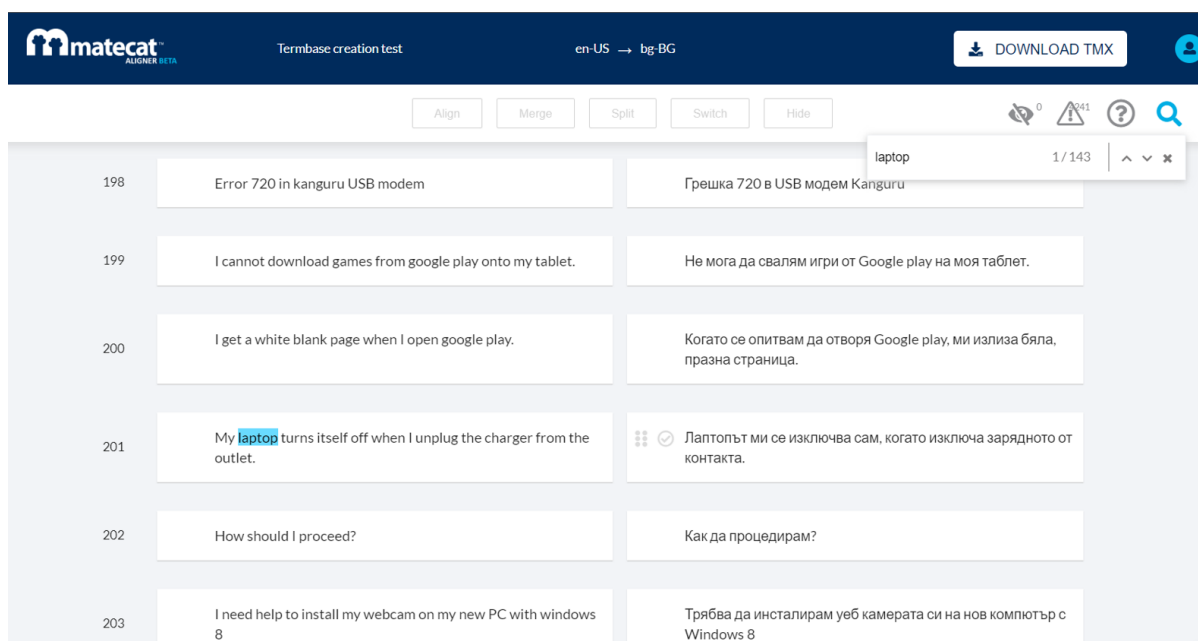
Разбира се, в списъка с извлечени английски термини присъстват и думи, които не са термини, и това наложи да бъдат отстранени ръчно. Например такива са *wont*, *much ram*, *doesn't*, *someone*, *much space* и *old file*.

След извличането на английските термини, са извлечени преводите им от използвания корпус. За целта английският и българския текст са подравнени с Мейткат (MateCat), програма, която също отговаря на заложените критерии за избора на подходящ инструмент.

Подравняването на еквивалентни езикови единици в повечето случаи е вградена функционалност в системите за компютърнопомогнат превод и се използва за

създаването на преводна памет. Изходният и целевият текст се съпоставят, за да могат да се създадат преводните еквиваленти, които имат (пълно) съответствие (описание се предлага в Първа глава). Тук обаче подравняването се използва за създаването на терминологична база. Съществува и възможността за едновременно двуезиково извличане, но предвид описаните ограничения, не е възможно да се използва в настоящия експеримент. Към функционалностите за разработване на програма, насочена към създаването на терминологични ресурси за български, може да се добави и възможността за едновременно извличане на двуезикова терминология. Необходимо уточнение е, че такава възможност има. Тя е разработена от латвийската компания Тилде¹⁴⁹. За да се използва обаче, са необходими умения за програмиране, което ограничава възможността за по-широко приложение¹⁵⁰.

След като текстовете бяха визуализирани в програмата Мейткат, така че изреченията да се едно до друго, английските термини се търсят като ключова дума, за да може да се види техният български еквивалент¹⁵¹. Запазва се сегментирането на ниво изречение (Фигура 19).



Фигура 19: Търсене на еквивалентни термини на български

¹⁴⁹ <https://www.tilde.com/>

¹⁵⁰ <https://github.com/tilde-nlp/taws>

¹⁵¹ Подравняването може да се използва за всякакъв вид търсене на преводни еквиваленти с различна цел, например стилистична.

След намирането на еквивалентните термини за български, резултатите са поместени в екселска таблица. Въпреки че Термостат позволява извличането на файлове в .txt формат, екселската таблица е избрана поради своята достъпност и затова че файловите формати могат да бъдат директно импортирани във всяка програма за компютърноподпомогнат превод. Също така, от анкетата към преводачите в България става ясно, че те я предпочитат. Броят на извлечените английски термини съответства за този на българските. Съкращенията в английски език са предадени непроменени на български.

II.3. Ръчна проверка и техническо оформление на терминологичната база

В настоящия експеримент беше направена ръчна проверка и нормализация на терминологичната база. Според изискванията на клиента или техническите възможности на програмата за компютърноподпомогнат превод форматите на терминологичната база може да варират, например .txt, .csv, .tbx, .sdltb. Мемсорс, например, позволява импортването на терминологични бази само от формати .xlsx, .tbx .tbx.xml (от Традос)¹⁵².

В мрежата могат да бъдат намерени и редица самостоятелни средства, с които извлечената терминологична база да се преобразува от един формат в друг. Пример за такива са: ТиБиЕкс Конверт (TBX Convert¹⁵³), Глосари Конвъртър (Glossary Converter¹⁵⁴), Ексбенч (ApSIC Xbench¹⁵⁵), Голдпан Тиём/Тибиекс Едитър (Goldpan TMX/TBX Editor¹⁵⁶).

За настоящия експеримент бе използвана десктоп програмата Олифант (Olifant¹⁵⁷), разработена от Панджеаник (Pangeanic).

След конвертирането в желания формат (в този случай полученият файл бе конвертиран в tmx), терминологичната база може да бъде импортирана в проекта за превод. Като инструмент за компютърноподпомогнат превод бе избрана програмата Мейткат (MateCat¹⁵⁸) и в нея беше създаден тестов проект за превод.

Процесът по добавяне на терминологичен ресурс е подобен във всички програми за компютърноподпомогнат превод.

¹⁵² https://www.memsource.com/uploads/2021/05/05/memsource_importing_term_bases.pdf [20.07.2022]

¹⁵³ <https://www.tbxconvert.gevterm.net/glossary/index.html> [20.07.2022]

¹⁵⁴ <https://appstore.sdl.com/language/app/glossary-converter/195/> [20.07.2022]

¹⁵⁵ <https://www.xbench.net/index.php/download> [20.07.2022]

¹⁵⁶ <https://logrusglobal.com/goldpan.html> [20.07.2022]

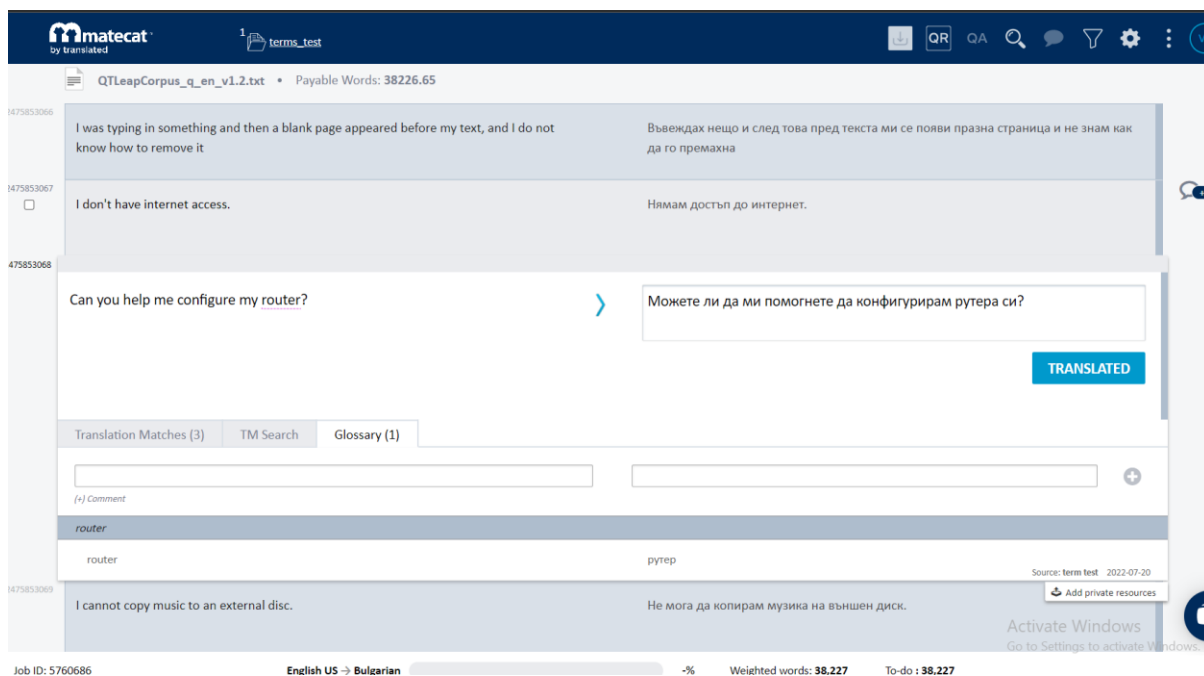
¹⁵⁷ <http://okapi.sourceforge.net/downloads.html> [20.07.2022]

¹⁵⁸ <https://www.matecat.com/> [20.07.2022]

След импортирането терминологичните ресурси се визуализират веднага, макар че за някои програми може да се наложи проектът да се затвори и пак да се отвори.

Необходими уточнения са, че:

- Конвертирането от един формат в друг за създаване на терминологична база не винаги е необходимо. Повечето програми за компютърнопомогнат превод поддържат голямо количество формати, най-вече Ексел. Достатъчно е документът да е оформен както трябва (в две съседни колони, с езиковия код начело на всяка колона и съответно термините на изходния и на целевия език;
- Предложеният метод не е оптимален и не решава въпросите, свързани с трудностите за преводачите. Показва само един от възможните пътища за решаване на проблема в условията на липса на достатъчно ресурси и програми за обработка на български език. Цел за последващи изследвания е разработването на инструмент, който да може да отговори още по-адекватно на нуждите на преводачите.



Фигура 20: Визуализиране на разпознат термин и неговия еквивалент за български в програмата Мейткат

III. Описание на характеристиките на създадената терминологична база от данни в областта на компютърната терминология

Ключово за качеството на една терминологична база е дали подобрите ресурси отговарят на нуждите на проекта за превод. Подбраната терминология трябва да е потвърдена от специалисти, извлечена от достоверни корпуси (или от документация, свързана с проекта за превод при липса на други ресурси). Към основния термин на изходния език трябва да са прибавени необходимите преводи за целевите езици.

Новосъздадената терминологична база съдържа извлечените от корпуса термини на английски език и техните еквиваленти на български език. Добавени са три допълнителни полета (за синоними, бележки и части на речта), но друг вид информация като дефиниции, изображения, препратки между термините или към външни източници не се предоставя. Към момента на един изходен термин съответства само един термин от целевия език.

От морфологична гледна точка в базата присъстват прости и съставни думи и всички термини са съществителни. Извлечените термини варират с максимална честота в паралелния корпус от 136 до минимална честота 3.73.

При прегледа на извлечените резултати, може да се направи следният езиков анализ: в английски език има сложни думи, които се предават в български като две отделни думи (например *cellphone* – *мобилен телефон*); съкращенията от английски или се предават по същия начин (например *wifi* – *wifi*¹⁵⁹), или са предадени с цялото им изписване на български (например *pc* – *настолен компютър*). Забелязват се сложни думи на английски, които се предават на български като съставни (напр. *antivirus* – *антивирусна програма*). От български към английски не се забелязват много такива примери (*url* – *URL адрес*). Срещат се термини, образувани чрез калкиране, например *playlist* – *плейлист*, както и запазване на съкращения като *ram* – *рам*.

При автоматичното извличане на термини като допълнителна информация към терминологичната база могат да се включат метаданни. Често срещана грешка при употребата на метаданни в базата е добавянето на допълнителна информация в скоби до термина. Добра практика е това да не се прави, защото може да се създаде проблем при правилната работа на предложенията на инструмента за компютърнопомогнат

¹⁵⁹ Необходимо уточнение е, че са възможни правописни грешки и в изходния език, които после се прехвърлят в целевия. Има риск да повлияят на цялостното качество на превода на текста.

превод. Друга добра практика е поставянето на бележки в конкретното поле, създадено за тази цел.

Необходимо уточнение е, че в новосъздадената база не са предложени синоними към термините, както и че няма дефиниции или обяснение за частите на речта. Причината за това е, че не е добра практика при превод на термини да се използват синоними освен ако няма конкретна причина или инструкция за това. Например терминът *рутер* има синоним в български думата *маршрутизатор*, терминът *изтегляне* има синоним *сваляне*, но в зависимост от предназначението на превода може да се използва по-често употребявания термин или терминът, който се предпочита в книжовната реч.. Въпреки че синонимите показват богатството и развитието на езика, те не са взети под внимание при създаване на терминологична база, тъй като заради стандартизираните и еднотипните текстове е задължително спазването и последователната употреба на едни и същи термини за конкретен проект и текст.

В зависимост от терминологичната база са възможни много полета, но в повечето случаи се използват само няколко. Полето за дефиниция е важен елемент особено когато трябва да се уточни за какъв термин става въпрос. Последното е изключително важно особено в случаи на омографи. При наличие на синоними и акроними, те най-често се поставят заедно с основния термин, като биват разделени с наклонена черта, запетая или точка и запетая. Подобно представяне би било пречка за работата на системата за компютърноподпомогнат превод, както и на инструментите за проверка на качеството. Всички синоними, назоваващи едно и също понятие, трябва да бъдат събрани в същия запис като отделни термини, като към всеки от тях трябва да бъдат добавени конкретни разяснения за нюансите в значението и употребата в целевия език. Може да се добави допълнителна информация за това кой термин е предпочитан, кой забранен, кой принадлежи към конкретен проект и др. В случаите, когато даден термин бива отхвърлен, добра практика е той да не се изтрива, а да се маркира като забранен. Това е много полезно при проверка на качеството на превода, защото програмата веднага може да засече дали забранени термини са използвани по време на превода. Друго предимство е, че след като даден термин вече е в базата и е маркиран като забранен, няма опасност той да бъде добавен повторно и съответно използван погрешка от преводача. Както при отхвърлянето на термини тази практика дава възможност за по-лесно проследяване при проверката на качеството на преведения текст.

Обобщение

В контекста на липса на терминологични бази, не само висококачествени и изчерпателни, но и приемливи като качество и с малък обем стават важни езиковите ресурси, които могат да се използват, за да се създадат (полуавтоматично или автоматично) терминологични бази. В тази част на дисертацията се показва, че за български език езиковите ресурси са недостатъчни: трудно се намират едоезикови и паралелни корпуси, принадлежащи към различни тематични области, които да се използват за извличане на термини; няма терминологични речници, създадени от експерти, за много от научните и технически области; не са налични двуезикови терминологични речници за различни двойки езици, отразяващи научното и техническо знание. На фона на съществуващия недостиг на подходящи езикови ресурси се наблюдава и липса на технологични средства (свободни и достъпни за всички), с помощта на които може да се извличат термини за един или повече езици. Всичко това поставя пред преводачите необходимостта сами да намерят подходящи начини за създаване на терминологични бази, които да им служат при конкретен превод.

В дисертацията се предлага методика за създаване на терминологични ресурси, която се основава на безплатни и независими инструменти за извличане на термини, като ги комбинира с подходящи корпуси. За илюстрация е избрана тематична област (свързана с компютърната терминология) и се представят особеностите на термините, които са специфични за тази област. Проследяват се начините за създаване на терминологични ресурси, за да се покаже, че полуавтоматичното извличане на терминологични бази е най-подходящият подход в условията на недостатъчни терминологични ресурси, едоезикови и двуезикови, както и на подходящи програми за извличане на терминология.

Методиката за полуавтоматично създаване на терминологична база в областта на компютърната терминология (английско-българска) се състои от следните компоненти: избор на двуезиков паралелен корпус от изследваната тематична област, избор и работа с програма за извличане на термини от изходния език (в случая английски), ръчно извличане на съответните български термини от подравнените изречения, проверка и техническо оформяне на базата. Накрая се описват характеристиките на създадената двуезикова терминологична база от данни, за да се покаже, че изходният двуезиков ресурс оказва сериозно влияние на качеството и пълнотата на създадената терминологична база, което, от своя страна, може да повлияе на превода. Доколкото основните етапи на работа включват подбор на паралелни езикови корпуси и

автоматично извличане на термини, може да се каже, че се предлага унифициран начин на работа при създаването на терминологични бази от данни. Направеното широко проучване гарантира, че са представени и сравнени най-добрите към момента инструменти за създаване на терминологични бази.

Предложената в тази глава методика за разработване на терминологични бази не предлага универсално решение за подобряване и улесняване на работата на преводачите във връзка с уместното използване на терминология. Тя има своите ограничения (ръчния подбор на корпуси и ръчното търсене на преводни еквиваленти) и не претендира за изчерпателност особено ако се сравни с възможностите на програмите за извличане на термини за други езици. Въпреки това е първи опит в тази посока с цел да се подпомогне разработването на специализиран софтуер за извличане на термини и създаването на двуезикови терминологични бази, в които единият език е български.

ЧЕТВЪРТА ГЛАВА

Измерване на качеството в системите за компютърноподпомогнат превод и техните компоненти

Увод	140
I. Измерване на точността на превода със системите за компютърноподпомогнат превод	142
I.1. Измерване на точността на превода без терминологични бази	143
I.2. Измерване на точността на превода с терминологични бази	144
II. Сравнение на ефективността на програмите за оценка на превода	148
III. Класификация на грешките при превод	153
IV. Сравнение на резултатите от автоматичните средства за измерване на качеството на превода	156
V. Международни стандарти за качеството на превода	162
Обобщение	168

Настоящата глава представя методите за измерване на качеството на превода в системите за компютърноподпомогнат превод. Разгледани са международните стандарти, регулиращи този вид дейност. Направен е преглед на функционалностите за проверка на качеството, вградени в системите за компютърноподпомогнат превод. Разгледани са и техническите параметри на независимите програми, използвани за проверка на качеството. Предложена е нова методика за измерване на превода в системите за компютърноподпомогнат превод.

В Четвърта глава се поставят за изпълнение следните цели:

- Да се представи измерването на точността на превода със системите за компютърноподпомогнат превод, съответно без или с използването на терминологични бази.
- Да се представи измерването на точността на превода с функционалности, вградени във или външни за инструментите за компютърноподпомогнат превод.
- Да се направи сравнение на ефективността на програмите за оценка на превода по предварително зададена методика.

- Да се предложи класификация на грешките при превод със системи за компютърнопомогнат превод с оглед на тяхната значимост на съдържанието на текста.
- Сравнение на резултатите от автоматичните средства за измерване на качеството по отношение на направената класификация на грешките при превод.
- Представяне на международни стандарти, метрики и модели за качеството на превода.

Увод

Измерването на качеството на превода е изключително трудна задача, защото няма само един идеален превод за даден текст. Например, юридическият превод има много по-различни изисквания по отношение на точността и спазването на специфични законодателни норми в сравнение с рекламен текст или ръководство за употреба. Спецификите на преводния текст зависят не само от преводача, от текста, който ще бъде преведен, или от субективната му интерпретация, но и от причините за превода, потенциалните читатели и публикационната и маркетингова политика, т.е. от фактори, които надграждат превода като езикова процедура. Тук спадат и социалните фактори, както и социокултурните, политическите или идеологическите ограничения, които могат да имат значително влияние върху превода (Хаус 2015: 153).

Според С. Колина „Липсата на консенсус за това как може да се определи качеството на един превод, произтича до голяма степен от разногласията относно концепцията за превод и от противоречивия и относителен характер на качеството, което в крайна сметка се оценява въз основа на социални, исторически и културни ценности и приоритети“ (Колина 2020: 458). Според К. Чуню и Уонг Так-минг, оценката на качеството на превода изисква разбирането на съдържанието на текста да се използва за определяне на различните видове отношения на еквивалентност и идентифицирането на грешки в превода. Тези когнитивни процеси обаче се различават по своите цели: за разбирането на текста е необходимо да се изгради интегрирано представяне на текста, докато идентифицирането на грешките показва какви са проблемите в текста и се опитва да намери решения за тяхното отстраняване (Чуню и Уонг Так-минг 2015: 220 – 221). Голяма част от критиките, отправени към различните подходи за оценка, се отнасят до тяхната зависимост от разбирането за „еквивалентност“, което е също толкова трудно за определяне, колкото и самият превод. Още в края на 90-те години Х. Хьониг констатира, че приемането на еквивалентността на изходния и преводния текст като априорна мярка

за качество е неправилно, защото, докато еквивалентността е компонент на превода, видът и степента на еквивалентност обикновено варират в зависимост от условията, при които се извършва преводът (Хьониг 1997: 9). Моделите за оценка на качеството на превода, базирани на еквивалентност, имат ограничаващ характер, тъй като предполагат, че еквивалентността е крайна цел на добрия превод. Задължително условие е спазването на нормите за правопис и пунктуация на целевия език, но не трябва да се забравя, че всеки отделен превод носи със себе си и допълнителни изисквания. В повечето случаи преводите се извършват с конкретна цел, различна от тази на оригинала, например за кого е предназначен преводът, свързани са с използването на определени термини за сметка на други, с избягването на дадени словосъчетания, възможни ограничения от целевия език и т.н. Тези допълнително заложи изисквания или „критерии“ за качество на превода изискват да се оцени дали преводът отговаря на тях към момента на измерване, или не.

Друга трудност е, че оценката за качеството на превода се извършва от хора, като по този начин субективният компонент на „човешкия фактор“ е още по-силно изразен (Зехналова и др. 2013: 43). Така оценката на превода не разчита на конкретна теория, а на лични възгледи, придобити чрез обучение и професионални контакти (Колина 2020: 458). Различни проучвания потвърждават наличието на субективни елементи при човешката оценка на превода, а основният проблем, свързан с оценяването на превода, несъмнено е субективността на оценяването. Този въпрос е третиран както от специалистите в областта, така и в академичните среди. Според Дж. Хаус „може да се каже, че измерването на качеството на превода е в основата на всяка теория на превода“ (Хаус 2015: 1).

Качеството на превода може да се разглежда и като степента, до която преводът следва договорените спецификации (Дурбан и Мелби 2008¹⁶⁰). Този подход всъщност следва препоръките на международните стандарти за качество. При стандартните работни процеси, проверката на качество е една от последните стъпки. Затова качеството на превода трябва да бъде дефинирано като относителна (а не абсолютна) адекватност по отношение на рамка, предварително договорена между възложителя и преводача. Тъй като целевият текст не може да бъде абсолютно еквивалентен на изходния текст поради естеството на човешките езици и насочеността на превода към конкретна ситуация, цел

¹⁶⁰ <https://www.communicaidinc.com/a-10-strategic-translation.php>

и аудитория, оценката на качеството на превода трябва да се ориентира по същия начин: за конкретна ситуация, конкретна цел и конкретна аудитория.

Друг фактор, който заслужава внимание е машинният превод. Навлизането на невронния машинен превод позволява подобряване на качеството на машинния превод и ефективното му използване в програмите за компютърноподпомогнат превод. Това на свой ред налага промяна и в работните процеси за проверка на качеството.

За да се намери решение на описаните дотук проблеми и за да се отговори на нуждите за по-голяма скорост и качество при проверката на превода, през последното десетилетие се разработват различни софтуерни програми. Тъй като броят им е голям, е полезно да се разграничат на вградени в системите за компютърноподпомогнат превод програми, самостоятелни програми и програми в облак. Самостоятелните програми за проверка на качеството на превода представляват най-широката група и са най-използваните. Основното им предимство е, че предлагат значително повече функционалности и могат да работят с голям брой файлови формати, докато останалите две групи са ограничени от възможностите на програмите, към които принадлежат.

I. Измерване на точността на превода със системите за компютърноподпомогнат превод

Разликите в граматическите структури на изходния и целевия език често водят до известна промяна в съдържанието при превод. Тази промяна може да бъде добавяне към целевия текст на информация, която не е изразена в изходния текст (и обратно). Това може да се случи, когато целевият език притежава граматична категория, която липсва на изходния език. М. Бейкър посочва, че при превод от английски или френски на индиански език като яна или навахо, би трябвало да се добави информация относно формата на предметите, споменати в текста. Аналогично при превод към езика амеша, трябва да се посочи дали някой, споменат в текста, е мъртъв или жив. Подробности, които не присъстват в изходния текст, но трябва да бъдат посочени на целевия език, могат да представляват сериозна трудност за преводача (Бейкър 1992: 99). Определянето на качеството в професионалния превод е широко обсъждан проблем, който и до днес не е решен, отчасти поради широкия спектър от възможни подходи. Предвид трудностите при формулирането на изискванията към качество на превода, качеството на превода трябва да бъде дефинирано от многостранна и всеобхватна перспектива (Матео 2016: 38).

Типът грешки, които трябва да се проследяват, са грешки в езиковия регистър, пунктуацията, правописа, граматиката, неправилно преведен термин, липсващо или добавено съдържание, грешки в цифрите, неправилно предаване на интернет връзките (най-често линкове, имейли и интернет сайтове), неспазване на таговете (чрез тях и изискванията за форматиране), неправилно превеждане на буквено-цифрови комбинации (като търговски марки и игри на думи). В допълнение, други ключови проверки са съответствието или липсата на правилна употреба на терминологията.

Съществуват инструменти за проверка на превода, които изискват известни познания при дефиниране на функционалностите за проверка, а други позволяват на потребителите лесно да създават свои собствени правила за проверка. Съществуващите инструменти за осигуряване на качеството на превода могат да бъдат класифицирани според набора им от функционалности по подразбиране. Някои от тях предлагат решения, при които потребителят може да променя конфигурацията, като избира или не дадени възможности за проверка (Макушина 2007: 5). Друг интересен факт е тенденцията на консолидация: инструментите за проверка на качеството следват същия еволюционен път като програмите за конвертиране на файлове, приложенията за преброяване на думи, инструментите за подравняване и софтуера за извличане на терминология.

1.1. Измерване на точността на превода без терминологични бази

Всеки инструмент за компютърноподпомогнат превод разполага с функционалност за проверка на качеството. Те са настроени по подразбиране и потребителят решава дали и как да ги използва (дали да добави допълнителни проверки към основните, или да ги деактивира изцяло и др.). МемоКю¹⁶¹ например може автоматично да провери използваните термини, форматирането и дължината на целевия сегмент спрямо изходния сегмент, последователността между сегментите за превод и преводната памет, дали направените промени в източника са приложени към целевия текст, дали цифрите, включени в изходния текст, присъстват в целевия, дали препинателните знаци, кавичките, интервалите, главните букви, специалните символи и вградени тагове се използват правилно в целевия текст. Програмата може да провери също и дължината на преведените сегменти в брой знаци и да използва регулярни изрази за откриване на проблеми, които не попадат в изброените (регулярните изрази трябва да бъдат написани от преводачите). Мемсорс предлага същите функционалности (без

¹⁶¹ <https://docs.memoq.com/current/en/Places/edit-qa-settings.html>

възможността за търсене чрез регулярни изрази), но функционалностите са разделени в четири основни категории и към тези категории е добавена и крайна персонализирана проверка. Четирите категории са лингвистика, терминология, форматиране и тагове и последователност при работния процес, като всяка от тях има подкатегории, съответстващи на подробна проверка за грешки и функции (Четинер 2021: 346).

Преводачът не може да приключи превода, без преди това да е направил проверка на качеството (в зависимост от настройките и проекта за превод) и разрешил евентуалните грешки и несъответствия. Контролът на качеството се извършва на първо място от човека, който е извършил превода, – самия преводач, а след това от редактора, възложителя или от други оператори (Гаудек 2007: 74).

1.2. Измерване на точността на превода с терминологични бази

В общия случай придържането към терминологията, включена в терминологичната база, е задължително. Тук е мястото да се споменат и така наречените DNTs или „Do Not Translate“ списъци – поредица от ключови термини, които трябва да останат непроменени в целевия език. Такива са например търговските марки (Кока-Кола, Фанта и др.) или имената на герои от видеоигри, които вече са познати и нямат нужда от адаптиране. Тази практика, колкото и полезна да изглежда, може да породи сериозно объркване и противоречия при използване на термините в дадена област. В много случаи експертите не могат да постигнат консенсус относно употребата и значението на даден термин. Така се появяват проекти за уеднаквяване на терминологията, например проектът „Терминология без граници“¹⁶² на Генерална дирекция „Писмени преводи“ към Европейската комисия.

Трябва да се отбележи, че досега няма компютърна програма, която да замени професионалния редактор. Езиковите технологии може само да улеснят работата на редакторите. Ако програмата за проверка на качеството на превода не е открила никакви грешки, това не гарантира, че преводът е (напълно) правилен. Не всички грешки обаче винаги са истински „грешки“. Възможно е сегментите, които софтуерът разпознава като съдържащи грешка, всъщност да са правилни. Преводачът (или редакторът) трябва да вземе решение за всеки отделен случай. Тъй като грешките в целевия език могат да окажат значително влияние върху качеството на превода, тяхната оценка включва нарушения в структурата на целевия език и фразеологията, които трябва да се

¹⁶² <https://termcoord.eu/>

разглеждат като „грешки в превода“, т.е. общи проблеми, които се наблюдават при създаването на текстове (Хансен 2010: 386).

За да се провери измерването на точността на превода с терминологични бази, както и за да се демонстрират възможностите за проверка на качеството в системите за компютърноподпомогнат превод (с фокус върху българския език), са проведени два експеримента: за първия бяха избрани 10 специализирани текста (с обем от около 15 000 думи) от регламентите на ЕС¹⁶³, преведени от експерти от английски на български, за превода на които се използва вече готова база; за втория бяха използвани текстове от корпуса (около 20 000 думи приблизително), използван за създаване на терминологичната база, описана в Трета глава.

Що се отнася до първия експеримент, английските текстове и българските им преводи са подравнени с инструмента Традос алайнер (Trados aligner), а терминологията е извлечена ръчно (общо 50 термина). След това термините са представени в основна форма както в изходния, така и в целевия език, защото лематизацията е метод, който се използва при създаването на речници и терминологични бази. Към тестовия корпус са въведени следните допълнителни грешки:

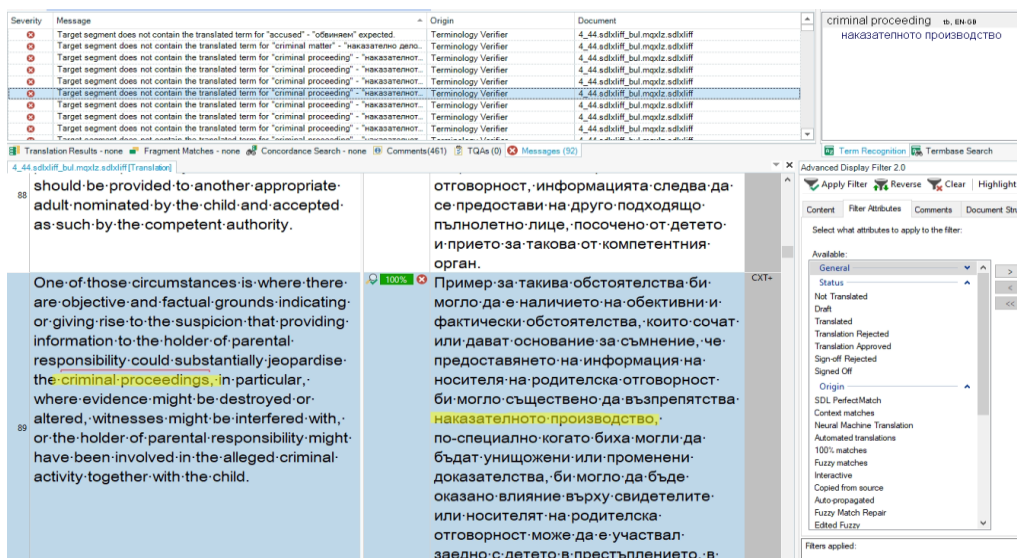
- изтриване на термин (като симулация на случая, в който даден термин не е преведен от преводача);
- частично представяне на термин (когато терминът е съставна дума), състоящо се в изтриването на една от думите, част от термина;
- непоследователен превод (използване на различни термини за предаване на едно и също понятие).

Всички въведени грешки са възможни в реална ситуация. Преводачът може да не следи терминологичната база или да пропусне част от определен сегмент. За проверка на изкуствено въведените грешки при употребата на терминология бяха използвани Традос и мемоКю (най-популярните програми за компютърноподпомогнат превод). За улеснение, всички останали настройки за проверката на качеството и на двете програми са изключени, с изключение на проверката на терминологията. С настройките си по подразбиране програмите засичат само термина в основната му форма. Експериментът е извършен от двама преводачи с Традос и мемоКю на различни компютри.

По-долу са показани резултатите от извлечените отчети за качество от тестваните програми. Въведените грешки се откриват с някои изключения. При Традос, например,

¹⁶³ <https://eur-lex.europa.eu/homepage.html>

в изходния език програмата разпознава само „наказателно производство“ и игнорира множественото число на термина, което е и очаквано поведение¹⁶⁴. Същата ситуация се повтаря и в целевия език.



Фигура 21: Резултати от проверката с Традос

Примерът от Традос е част от следния сегмент:

Изходен език: *...this Directive aims to strengthen the trust of Member States in each other's criminal justice systems and thus to improve mutual recognition of decisions in criminal matters.*

Целеви език: *...настоящата директива има за цел да засили взаимното доверие на държавите членки в техните системи за наказателно правосъдие и по този начин да спомогне за постигане на по-широко взаимно признаване на решенията по наказателни производства.*

Както се вижда, „наказателно дело“ се появява в съобщението на програмата: „Преводът на изходния термин „наказателно дело“ липсва в целевия сегмент. Възможни термини: наказателно дело“, докато преводът присъства ясно в целевия текст. Не се разпознават само формата за множествено число.

При следващия пример:

¹⁶⁴ Това е често срещан проблем и при други езици като шведски и фински.

Изходен език: *The results of the medical examination shall be taken into account when determining the capacity of the child to...*

Целеви език: *Резултатите от медицинския преглед се вземат предвид при определяне на способността на детето да...*

Съобщението, изведено от мемоКю, е: „Преводът на изходния термин „physical integrity“ липсва в целевия сегмент“. Програмата не може да оцени, че терминът вече е използван правилно.

Вторият експеримент имаше за цел техническа проверка на методиката за създаване и проверка на терминологична база за компютърна терминология, за да се провери дали може да се използва от системите за компютърнопомогнат превод без технически затруднения. Проверката на термините е по модела на първия експеримент и обхваща същите изкуствено въведени грешки като изтриване на термин, частично изтриване на една от думите, част от термина и последователно използване на различни термини за предаване на едно и също понятие. Термините бяха коректно визуализирани и резултатът беше същият както при предходния експеримент. Програмата засича липсата на термините или частичното им присъствие във всички изкуствено въведени грешки. Повтаря се обаче ситуацията, описана по-горе, с неразпознаване на множественото число на термините, което показва, че всяка форма трябва да присъства като запис в терминологичната база.

Document	Row	Code	Description	Ignore
QLeapCorpus_q_en_v1.2.txt	5	03091	Translation of source term "wireless network" missing from the target. Possible terms: безжична мрежа	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	5	03091	Translation of source term "network" missing from the target. Possible terms: мрежа	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	9	03091	Translation of source term "mobile phone" missing from the target. Possible terms: мобилен телефон	<input checked="" type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	16	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	18	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	19	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	20	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	21	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	21	03091	Translation of source term "update" missing from the target. Possible terms: актуализация	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	22	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	22	03091	Translation of source term "mobile phone" missing from the target. Possible terms: мобилен телефон	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	23	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	24	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	25	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	26	03110	Extra space at the end of the segment	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	27	03091	Translation of source term "icon" missing from the target. Possible terms: икона	<input type="checkbox"/>
QLeapCorpus_q_en_v1.2.txt	29	03110	Extra space at the end of the segment	<input type="checkbox"/>

Фигура 22: Резултати от проверката на новосъздадената база с МемоКю

Проведените експерименти проверяват функционалностите за проверка на качеството в системите за компютърнопомогнат превод за български език, както и поведението на новосъздадената терминологична база. Чрез изкуственото добавяне на грешки, срещани в реална работна среда, бе възможно да се покаже, че терминологичната база може успешно да бъде използвана.

II. Сравнение на ефективността на програмите за оценка на превода

Програмите за проверка на качеството, външни на инструментите за компютърнопомогнат превод изпълняват същата функция като вградените, с тази разлика, че разполагат с много по-широк спектър от допълнителни функционалности. Следните програми са сред най-разпространените: Кюей дистилър¹⁶⁵ (QA Distiller), Ексбенч¹⁶⁶ (Xbench), Верифика (Verifika), Еръспай (ErrorSpy) и Лингвистик тулбок или Елтиби¹⁶⁷ (Linguistic ToolBox или Ltb). В началото на своето развитие тези програми са били групирани в две категории: граматика и форматиране. Граматиката е свързана с правописа, пунктуацията и степента на владеене на езика, а форматирането – с откриване на ненужни двойни интервали и др. За да може да бъдат систематизирани откритите грешки (и ако се изисква от потребителя), грешките се включват в отчет, който позволява удобна корекция без използване на външен софтуер. Основното предимство на тези програми е, че възможните грешки, свързани с терминологията и съгласуването, се отбелязват по различен начин. Някои от тях предлагат възможност за създаване на контролни списъци, които могат да се използват за конкретен клиент, за да се сведе до минимум рискът от пропуски.

Функционалности на изброените по-горе програми:

- *Празни сегменти* (Empty segments) е функционалност, която следи да няма празни сегменти в текстовете за превод.
- *Целевият текст съвпада с изходния текст* (Target text matches the source text) следи за съответствие между целевия и изходни език.
- *Несъответствие* в таговете (Tag mismatch) следи дали тагове (етикети, прикрепени към конкретна дума с цел идентификация или за предоставяне на друга информация) от изходния сегмент присъстват в целевия сегмент.

¹⁶⁵ <http://www.qa-distiller.com/en>

¹⁶⁶ <https://www.xbench.net/>

¹⁶⁷ <http://autoupdate.lionbridge.com/LTB3/>

- *Несъответствие в цифрите* (Number mismatch) е функционалност, която следи дали цифрите в изходния текст присъстват в целевия текст.
- *Граматична грешка* (Grammar) следи за допуснати граматични грешки.
- *Несъответствие в линковете* (URL mismatch) е функционалност, която следи дали линковете в изходния текст присъстват в целевия текст.
- *Правописни грешки* (Spelling).
- *Буквено-цифрово несъответствие* (Alphanumeric mismatch) е функционалност, която следи дали комбинациите от цифри и букви в изходния текст присъстват в целевия текст.
- *Липса на двоен символ* (Unpaired symbols) следи дали при двойните символи (например скоби или кавички) липсва отварящ или затварящ символ.
- *Частичен превод* (Partial translation) е настройка, която следи за минимален брой непреведени последователни думи. При този вид превод някои части от текста на изходния език остават непреведени и се прехвърлят директно в целевия език.
- *Повторен празен символ* (Double blanks) следи за последователности от няколко празни символа.
- *Повтарящи се думи* (Repeated words) следи за повторения на дума.
- *Последователност в изходния текст* (Source consistency) е функционалност, която следи за сегменти с еднакъв превод, но с различни изходни сегменти.
- *Последователност в целевия текст* (Target consistency) е функционалност, която следи за еднакви сегменти в изходния език, но с различни преводи на целевия език.
- *Доклад за извършени промени* (Change report) е функционалност, която позволява да се проследи кой е направил дадени промени в текста.
- *Обработка на множество файлове едновременно* (Multiple files) е настройка, която позволява проверката на повече от един файл.
- *Камелкейс* (CamelCase) е функционалност, която позволява да се проследи дали съкращения (комбинация от букви и цифри и от малки и главни букви) от изходния език присъстват и целевия.
- *Терминология* (Terminology) проследява за последователната употреба на термините в целия текст.
- *Списък за проверка* (Checklists) е предварително уговорен списък между преводача и клиента, в който са описани стъпките за гарантиране на качеството.

- *Силно търсене* (PowerSearch) е режим на търсене, който използва регулярни изрази.
- *Профили* (Profiles) са персонализирани настройки за проверка на качеството, правописа и граматиката, които са избрани за конкретен потребител.
- *Доклад* (Report) визуализира откритите от програмата грешки в текста.
- *Команден панел* (Command line) дава възможност за работа без графичния потребителски интерфейс на програмата.
- *Списък от думи, които не трябва да се превеждат* (DNT List).

В сравнение с възможностите за проверка на качеството посредством системите за компютърноподпомогнат превод, специално създадените програми за проверка на качеството разполагат със значително по-голям брой функционалности. Те са разработени от различни компании. Обикновено преводачите работят с повечето от тях, за да могат да отговорят на нуждите на всеки проект за превод. При по-задълбочен анализ се откриват разлики в техните възможности. По тази причина за целите на дисертацията бе направена съпоставка на техническите им възможности. Бе прегледана техническата документация на всяка една от описаните по-горе програми, като всяка налична функционалност беше включена в списък. След това списъкът бе сравнен за всяка програма, за да се види какво е съотношението между тях (Таблица 4).

	Ексбенч	Верифика	Ерърснай	Елтиби	Кюей дистилър
Празни сегменти	X	X	X	X	X
Целевият текст съвпада с изходния текст	X	X	X	X	X
Несъответствие в тагове	X	X	X	X	X
Несъответствие в цифрите	X	X	X	X	X

Граматична грешка		X		X	
Несъответствие в линковете	X	X	X	X	X
Правописни грешки	X	X	X	X	X
Буквено-цифрово несъответствие	X	X	X	X	X
Липса на двоен символ	X	X	X	X	X
Частичен превод	X	X		X	X
Повторен празен символ	X	X		X	X
Повтарящи се думи	X	X	X	X	
Последователност в изходния текст	X	X	X	X	X
Последователност в целевия текст	X	X	X	X	X
Доклад за извършени промени				X	
Обработка на множество файлове едновременно	X	X		X	X
Камелкейс	X	X	X	X	X
Терминология	X	X	X	X	X

Списък за проверка	X	X			X
Силно търсене	X	X	X		X
Профили		X		X	X
Отчет	X	X	X	X	X
Команден панел					X
Списък от думи, които не трябва да се превеждат	X	X	X	X	X

Таблица 4: Съпоставка на техническите характеристики на самостоятелни програми за проверка на качеството на превода

Както се вижда от таблицата, всички програми разполагат с функционалности като проверка на правописа, несъответствия в цифрите, пунктуацията и таговете. Всички програми засичат също наличието на празни сегменти в целевия език, липсата на конкретен символ в целевия език и последователното използване на терминология. Други функционалности обаче са присъщи само на конкретна програма, например възможността за извличане на доклад за извършени промени (само Елтиби разполага с такава). Всички различия трябва да се имат предвид при употребата на програмите за проверка на качеството на превода, за да може да се подбере най-подходящата за целите на конкретен превод.

По време на работата с външните програми бе забелязано, че всяка програма степенува откритите грешки по различен начин. Например правописна грешка може да е много по-сериозен пропуск в сравнение с двоен празен символ. В Традос идентифицираните проблеми при превода се делят на: грешка (Error), предупреждение (Warning) и забележка (Note), като към грешка принадлежат липсващи или сгрешени цифрови изрази, към предупреждение – различна пунктуация в края на изречението, а към забележки – двоен празен символ. В Ексбенч разделението е между основни грешки (Basic) – непреведен сегмент или несъответствие в текста, грешки в текста (Content) – пунктуационни несъответствия, двойни празни символи, пропуснати или сгрешени цифри и други, списък с грешки (Checklists) и проверка на правописа (Spell-checker).

III. Класификация на грешките при превод

Както бе посочено по-горе, програмите за проверка на качеството на превода могат да засичат значителен брой грешки. Значението на грешките обаче варира според влиянието, което оказват върху превода. Някои от програмите предлагат персонализиране на сериозността на даден пропуск в зависимост от изискванията на потребителя. Кюей дистилър определя сериозността на грешката според цифрова скала. Например: двоен празен символ би получил Тежест 1 (Severity 1), пунктуационна грешка би получила Тежест 2 (Severity 2), докато пропуск или несъответствие в превода би получил Тежест 5 (Severity 5). Лингвистик тулбокс, която няма да бъде използвана за настоящия експеримент, ги разделя на сегментация (Segment Checker) – несъответствия на ниво граници на сегмент, проверка на правописа (Speller), термини и пунктуация (Term & Punctuation), обща информация (General Information) – името на файла или използваната конфигурация, като между отделните групи няма йерархия. Програмата мемоКю, която също като Традос е програма за компютърнопомогнат превод, не степенува грешките по важност, а по вид, но без да им дава конкретна стойност.

Както се вижда, всяка програма оценява грешките, които открива, с различна важност. Само по себе си това не би могло да се смята за проблем, ако се работи само е една програма. В случаите, когато се използва повече от една програма обаче, има риск за объркване, най-вече относно степенуване за сериозността на конкретната грешка. Поради тази причина се предлага следната класификация в опит да се анализират грешките и техните реални последици за качеството на превода. Класификацията отчита възможните негативни ефекти, които дадената грешка може да окаже на качеството на превода:

Скала	Описание	Вид на грешката
-------	----------	-----------------

<p>1</p>	<p>Оказва слабо или никакво влияние върху качеството на превода; рядко може да доведе до объркване</p>	<p><u>Несъответствие в пунктуацията:</u></p> <ul style="list-style-type: none"> - ограждащи символи (скоби, кавички) – липса на единия знак, липса на двата знака; липса на единия или и на двата знака и наличие на друг знак; наличие на различни знаци около една дума; - различен, липсващ или излишен пунктуационен знак в края на изречението; - излишни или липсващи празни символи (спейсове, табулации); - излишен (липса на) нов ред.
<p>2</p>	<p>Може да доведе до объркване, в зависимост от контекста</p>	<p><u>Графични несъответствия</u></p> <ul style="list-style-type: none"> - несъответствие в цифровите изрази: липсващи, объркани или разместени цифрови изрази (несъответствия в числа, дати, часове, дроби, номерация и др.). При определени ситуации обаче, програмите могат да сметнат за грешка нещо, което е вярно (превръщане на градуси от Целзий към Фаренхайт, локализиране в различни метрични системи и др.); - несъответствие в интернет линкове, имейли и др.; - несъответствие в комбинациите от цифри и букви.

3	Риск за сериозна подмяна на съдържанието на текста	<p><u>Несъответствия в пълнотата на превода</u></p> <ul style="list-style-type: none"> - празни сегменти; - сегментите в целевия език и в езика източник съвпадат. В различните инструменти се разглежда и като текст, който не е преведен; - несъответствие в главните букви в началото на сегмента или изречението; - изпуснати, добавени, разменени букви; - наличие или липса на специфични символи; - наличие на знаци от друга азбука в думата.
4	Неприемлива грешка	<p><u>Правопис:</u></p> <ul style="list-style-type: none"> - повторени думи; - изпуснати думи (могат да бъдат засечени само когато се отнасят до съкращения или думи, намиращи се между кавички, скоби или в комбинация с числа); - несъответствие в съкращения (за да бъдат засечени подобни несъответствия, е необходимо да бъдат въведени предварително в настройките); - липса на преведени части от изречението (подобна липса се отразява най-често, като бъде сравнена дължината на изреченията в %); - непоследователна употреба на конкретни думи и термини в текста. <p><u>Граматика</u></p> <p>Само някои инструменти са снабдени с подобна функция, и то за най-разпространените езици. При езици с малък брой носители като шведски, датски,</p>

		<p>фински, български и др., програмите за оценка на превода не поддържат проверка на граматиката. Избраните за експеримента програми не притежават подобна функция и тя няма да бъде тествана.</p>
--	--	--

Таблица 5: Класификация на грешки при превод и тяхното влияние за правилното пренасяне на съдържанието

Грешките варират по важност и последствия при правилното пренасяне и разбиране на съдържанието при превод от един език на друг.

IV. Сравнение на резултатите от автоматичните средства за измерване на качеството на превода

Самостоятелните програми за оценка на качеството на превода имат много общи характеристики, но между тях се наблюдават и разлики. Общото между тях е следното: проверяват терминологията, несъответствия при цифри, линкове, интернет адреси и възможността да се генерира отчет с грешки, най-вече във формат excel), които след това могат да бъдат проверени от специалист по контрол на качеството (QA specialist) или изпратени до преводача, който е работил по проекта. Последната стъпка зависи от това

какви практики са възприети от участниците в дадения проект (по-долу са изброени използваните практики).

Въпреки че тези програми гарантират добри резултати, когато се използват за проверка на формалните характеристики на превода, в някои случаи могат да открият фалшиви грешки (false positives). Такива са: разлика в правилата за разстояние в изходния и в целевия език (например във френски език между дума и двоеточие има разстояние), инструкциите относно начина на записване на цифрите и др. В същото време конкретна програма за проверка на качеството може да е по-добра в откриването на определен тип грешки в сравнение с останалите. В списъка с грешки на Елтиби това ще бъде отразено като грешка. Друг такъв пример е разликата в дължината на текста в езика източник и в целевия език. Когато тестът в целевия език е с 20% по-дълъг, Верифика го посочва като възможна грешка, въпреки че езиците имат ясно изразени лексикални и граматични структури, които се различават.

Инструментите за компютърноподпомогнат превод също са снабдени с функционалности за проверка на качеството. Най-новите инструменти за компютърноподпомогнат превод имат много общи функции с програмите, специално предназначени за проверка на качеството на превода.

Беше направено сравнение на начина на работа на различни програми и на тяхната успеваемост при откриването на грешки при превода от английски на български. За целта бяха използвани специализираните програми Кюей дистилър и Ексбенч и функционалностите за проверка на качеството на Традос 2019 и мемоКю 9.2. Тоест, сравняват се две програми, специално създадени за проверка на качеството на превода, и два инструмента за компютърноподпомогнат превод, в които проверката на качеството на превода е вградена функционалност. За тази цел бяха използвани текстовете от първия експеримент.

По-долу са представени обобщено резултатите от проведеното сравнение.

Грешки със степен на важност 1 се откриват винаги и от всички програми. Понеже става въпрос за липсващ ограждащ символ или пропуснати празни символи, подобни грешки са с минимални последствия за качеството на превода. Например:

Изходен език: *(1) of the Treaty on the Functioning of the European Union (TFEU) provides...*

Целеви език: *(1 ... от Договора за функционирането на Европейския съюз ДФЕС) Съюзът подкрепя...*

Грешки със степен на важност 2 също се откриват винаги (несъответствие между цифри, линкове и комбинации от цифри). Тук е необходимо да се подчертае, че от контекста може да зависи дали са реални, или фиктивни грешки. Към тази група принадлежат и значителна част от намерените „фалшиви грешки“:

Изходен език: *It will be launch on March 17, 2020 at 2pm Bulgarian time, as the local authorities...*

Целеви език: *Той ще бъде отворен на 17 март 2020 г. в 14.00 ч. българско време като преди това местните власти...*

Изходен език: *https://data.europa.eu/euodp/en/legal_notice*

Целеви език: *https://data.europa.eu/euodp/bg/legal_notice*

Пример за „фалшива грешка“ със стойност 3. е засичане на липсата на скоби в целевия език и сигнализиране на пропуската:

Изходен език: *...territory of the Member States shall not...*

Целеви език: *...територията на държава/и-членка/и не*

При откриването на грешки, характеризиращи се със стойност 4, се забелязват различия в поведението на отделните програми. Ексбенч и Кюей дистилър не отчитат проблеми с правописа, докато Традос се справя значително по-добре. Въпреки това се отразяват като грешки правилни думи: *правоприлагане, дихлоробензен, ароматизатори, договорка, киберсигурност* и др.

Безспорното и изключително ценно при превода е предимството на тези програми, когато става въпрос за грешки от тип 4 от скалата – възможността да се открива несъответствие при употребата на конкретни думи, което може да доведе до загуба или объркване при интерпретацията на значението. Става въпрос за последователната употреба на едни и същи думи/термини в целия текст на целевия език.

Когато става дума за несъответствие в употребата на думи или термини, системите работят и по посока на езика източник. Това е допълнителна настройка, която трябва да бъде активирана ръчно.

Изходен език: *Such family members shall retain their right of residence exclusively on personal basis.*

Изходен език: *Such family members shall retain their right of residence exclusively on a personal basis.*

Поведението на програмите при работа с български език заслужава отделно внимание. Най-общо, не се отчитат някои специфични пунктуационни правила и знаци (по-точно, програмите не се справят с правилното отчитане на българските кавички), които не присъстват в други езици:

Целеви език: ...към 1 август 2019 г., при условие че са спазени...

За програмите българските кавички са неправилни поради различието им с тези в изходния език:

Изходен език: *“Almost 90 % of the inspected products do not contain substances...has an obligation to provide it when requested,”*

Целеви език: *„Почти 90 % от проверените продукти не съдържат вещества...има задължението да я предоставя при поискване“*

Както беше посочено, има случаи, в които програмите се държат по различен начин при един и същи текст. Кюей дистилър открива немските букви като „повредени знаци“ (corrupt characters), защото са в български текст:

Изходен език: *...Südfrüchte Großhandlung Lidl & Co. at Heilbronn as a shareholder in 1930, which was then renamed...*

Целеви език: *...Südfrüchte Großhandlung Lidl & Co. в Хайлброн като акционер през 1930 г., който бива преименуван...*

Ексбенч е особено чувствителна програма по отношение на съкращенията и комбинациите от цифри и букви. В примера по-долу програмата е засекала, че при подчертаните съкращения в целевия език липсва s и го тълкува като грешка:

Изходен език: *... fluorinated greenhouse gases – such as hydrofluorocarbons (HFCs), perfluorocarbons (PFCs), sulphur hexafluoride (SF6)*

Целеви език: ... *флуорирани парникови газове – като хидрофлуоровъглероди (HFC), перфлуоровъглероди (PFC), серен хексафлуорид (SF6)*

Безспорно е предимството на програмите за проверка на качеството. За оптимални резултати обаче винаги са необходими персонализирани настройки и човешка преценка. Независимо от факта, че тези програми намират „фалшиви“ грешки, основното им предимство е възможността да открият несъответствия в превода (например, дали в целия текст се използват едни и същи термини). Представените по-долу грешки вероятно биха били игнорирани в реална работна среда, но дават добра представа за начина на работа на тези инструменти. Всички програми са открили, че предлогът *на* е на различни места в два сегмента от целевия език. Понеже програмите са предвидени да откриват всякакъв вид несъответствия, това разместване се приема като грешка, макар качеството на превода да не е компрометирано:

Исходен език: *a document attesting to the existence of a family relationship or of a registered partnership;*

Целеви език: *документ, удостоверяващ съществуването на семейни взаимоотношения или на регистрирано съжителство;*

Целеви език: *документ, удостоверяващ съществуването на семейни взаимоотношения или регистрирано съжителство;*

Още един много ясен пример за „фалшива“ грешка е липсата на отварящата скоба при изреждането на буквите:

Исходен език: *in cases falling under points (c) and (d) of Article 2(2), documentary evidence that the conditions laid down therein are met;*

Целеви език: *в случаите, попадащи в приложното поле на член 2, параграф 2, букви в) и г), документи, доказващи, че са изпълнени условията, посочени в тези разпоредби;*

Целеви език: *в случаите, попадащи в приложното поле на член 2, параграф 2, букви в) и г), документално доказателство, че са изпълнени условията, постановени в тези разпоредби;*

За да се постигнат най-добрите възможни резултати с тези програми, е задължително да се зададат конкретни настройки за всеки превод. Предлагаме и

допълнителна схема за оценяване на инструментите за проверка на качеството (Таблица б).

Първо ниво	интерфейс и основни характеристики	<ul style="list-style-type: none"> - лесно ли се използва инструментът - самостоятелен ли е, или е част от програма за компютърноподпомогнат превод - какви файлове може да обработва - кои езици поддържа - позволява ли персонализация
Второ ниво	често използвани функции	<ul style="list-style-type: none"> - как се справя с общите проблеми (пропуснати знаци, цифри и т.н.) - настройки по подразбиране - непоследователност при употребата на терминология - количество на фалшивите грешки
Трето ниво	персонализирани настройки	<ul style="list-style-type: none"> - как се справя с локализацията - как се справя с настройките за конкретен език в сравнение с друг - персонализирани настройки за конкретен език/клиент – формат на дати, елемент, който не е за превод и др. - разчитане на специални символи - количество на „фалшивите“ грешки при персонализирани настройки
Четвърто ниво	допълнителни предимства	<ul style="list-style-type: none"> - аудиокоманди - редакция на преводния файл

		- други предимства (например директни връзки между различни програми)
--	--	---

Таблица 6: Допълнителна схема за оценяване на програмите за проверка на качеството на превода

Програмите за автоматична проверка на качеството (било то вградени в системите за компютърнопомогнат превод или самостоятелни) са полезни, защото, до известна степен улесняват работата на преводачите. Въпреки това, не трябва да се забравя, че те не могат да заместят хората. Последното се доказва от присъствието на „фалшивите“ грешки.

V. Международни стандарти за качеството на превода

Международните стандарти за качество се отнасят до средствата за измерване на качеството, последователността, в която трябва да се преглеждат и обработват файловете за превод, уменията и квалификацията на преводачите и други, свързани с превода компоненти.

С цел унифициране на работните процеси, терминологията, определяне на уменията на преводачите и гарантиране на качеството на превода се приемат международни стандарти.

Към днешна дата актуалните международни стандарти при превод на документи са:

- ISO 9001:2015¹⁶⁸ – управление на качеството на системите;
- ISO 17100:2015¹⁶⁹ – преводачески услуги;
- ISO 27001:2013¹⁷⁰ – системи за информационна сигурност;
- ISO 18587:2017¹⁷¹ – редакция на машинен превод.

Въпреки че стандартите за превод съществуват от известно време, едва през 2006 година организациите започват официално да работят заедно за стандартизиране на практиките и създаване на официални документи за стандарти, както и на схеми за сертифициране за областта на превода. Крайната им цел е защитата и информираността на потребителите чрез гарантиране на определено ниво на качество при превода. Има

¹⁶⁸ <https://www.iso.org/iso-9001-quality-management.html>

¹⁶⁹ <https://www.iso.org/standard/59149.html>

¹⁷⁰ <https://www.iso.org/standard/54534.html>

¹⁷¹ <https://www.iso.org/standard/62970.html>

няколко елемента, общи за всички стандарти за превод: необходимото или договорено ниво на постигане на качество, необходимостта от ясни договорени спецификации на проекта, изискване за споразумение между клиент и доставчик на преводаческа услуга, процес на управление на проекти, технически капацитет, както и някои други условия, свързани с предоставянето на превод (Бендана и Мелби 2012: 81).

Стандартите, описани по-горе, включват стандартите от серията ISO 9000¹⁷², но с нарастването на интереса към управлението на качеството, както и с все по-големите обеми на преводите, са разработени специфични стандарти за качество на преводаческите услуги. Те включват италианския UNI 10574¹⁷³, немския DIN 2345¹⁷⁴, австрийските Önorm D 1200¹⁷⁵ и Önorm D 1201¹⁷⁶ и канадския CAN CGSB 131.10¹⁷⁷. През 2006 г. бива разработен EN 15038, „...европейски стандарт, определящ изискванията към доставчиците на преводачески услуги по отношение на човешките и техническите ресурси, управлението на качеството и на проектите, договорната рамка и процедурите“¹⁷⁸. Посредством стандарта се цели да се унифицира терминологията, използвана в областта на превода, да се определят основните изисквания за доставчиците на преводачески услуги и да се създаде рамка за взаимодействие между клиентите и доставчиците на услуги по отношение на техните права и задължения. Акцентираща се върху процесите на административно, документално и редакционно ниво, както и върху функциите на различни специалисти, които участват в процеса на превода. Минимално изискване според този стандарт е преводите да включват най-малко две отделни лица: едното да извърши превода, а другото да го редактира. Въпреки че е заменен с от ISO 17100 през 2015 г., много компании все още продължават да спазват зададените от този стандарт изисквания.

Към днешна дата международните стандарти, които регламентират и определят качеството на превода, са ISO 17100, ISO 9000, ASTM F2575-14 и ISO 18587. Стандартът ISO 17100 определя изискванията към процесите, ресурсите и всички други аспекти, необходими за предоставянето на качествена преводаческа услуга. Също така описва

¹⁷² <https://www.iso.org/iso-9001-quality-management.html>

¹⁷³ http://store.uni.com/catalogo/uni-10574-2007?josso_back_to=http://store.uni.com/josso-security-check.php&josso_cmd=login_optional&josso_partnerapp_host=store.uni.com

¹⁷⁴ <https://www.beuth.de/en/standard/din-2345/3364650>

¹⁷⁵ https://shop.austrian-standards.at/action/de/public/details/72071/OENORM_D_1200_2000_02_01

¹⁷⁶ https://shop.austrian-standards.at/action/de/public/details/320045/OENORM_D_1201_2009_01_01;jsessionid=98BDF04F451AB2AAA5AB3FBAF2235921

¹⁷⁷ <https://www.scc.ca/en/standardsdb/standards/28935>

¹⁷⁸ https://www.bds-bg.org/standard/?national_standard_id=51617

средствата, чрез които доставчикът на услуги на превод може да демонстрира способността си да предостави преводаческа услуга, която да отговаря на изискванията на клиента (както и на самия доставчик на преводаческа услуга). ISO 9000 адресира различни аспекти на управлението на качеството и съдържа някои от най-известните стандарти на ISO, които предоставят насоки и инструменти за организации, които искат да гарантират, че техните продукти и услуги последователно отговарят на изискванията на клиента. ASTM F2575-14 предлага рамка за клиентите и доставчиците на преводачески услуги, желаещи да се споразумеят за специфичните изисквания на конкретен преводачески проект. Този стандарт не предоставя критерии за превод или качество на проекта поради факта, че изискванията на клиентите могат да се различават и да бъдат строго индивидуални, но посочва параметри, които трябва да бъдат взети под внимание преди започването на проект за превод. Най-нов е стандартът ISO 18587¹⁷⁹, който описва процесите на редактиране на машинен превод и нужната компетентност на такъв тип редактори.

Съществуват предложения, тъй като машинният превод и системите за компютърноподпомогнат превод по своята същност са компютърни програми, тяхното оценяване да се смята за етап от софтуерното инженерство, към което са приложими съществуващите стандарти за оценка на софтуер. Съвместно разработен от Международната организация по стандартизация и Международната електротехническа комисия, стандартът ISO/IEC 9126 (1991; 2001) предоставя модел за качество, идентифицира няколко вида метрики за оценка на софтуера и определя качеството на софтуера като „свкупността от характеристики на даден продукт или услуга, които се отнасят до способността му да задоволява заявени или подразбиращи се нужди“. Идентифицирани са шест характеристики, а именно функционалност, надеждност, използваемост, ефективност, поддръжка и преносимост, всяка от които е допълнително разделена на редица подхарактеристики (Чуниу и Уонг Так-минг 2015: 218).

Най-популярните модели за проверка на качеството (Linguistic Quality Assurance или LQA), основаващи се на човешка преценка, са ЛИСА Кюей метрик (LISA QA Metric), ТиКюАй (TQI) и ЕмКюЕм (MQM-DQF). ЛИСА е съкратено от Асоциация за индустриални стандарти за локализация (Localization Industry Standards Association¹⁸⁰) (1990 – 2011) на компютърен софтуер (и свързаните с него материали) на множество

¹⁷⁹ <https://www.iso.org/standard/62970.html>

¹⁸⁰ <https://web.archive.org/web/20110101184308/http://www.lisa.org/> чрез Internet Archive

езици. LISA КюЕй метрик се основава на обратната връзка от страна на редактора или коректора на превода, която се структурира според нивото на сериозност на намерените грешки (те се категоризират като незначителни (Minor), сериозни (Major) или критични (Critical)). Ако в даден сегмент има две грешки с различна степен на сериозност, всяка една се записва поотделно в съответното поле.

ТиКюАй¹⁸¹ (Translation Quality Index или TQI) е количествен метод за оценка на качеството на превода, който измерва броя и вида на грешките, открити в даден текст, и изчислява резултатите. Индексът за качество на превода приписва определена стойност на преведения текст, като 100 означава „безгрешен“ превод. Всяка от грешките може да бъде маркирана в зависимост от последствията от нея. За да се постигне определен резултат, оценителят трябва да спазва определени правила при маркиране на грешки. В противен случай употребата на ТиКюАй може да доведе до приспособяване на данните от оценителя за това колко добър или лош е даден превод. Добрият оценител трябва да бъде обективен и да е в състояние да прави разлика между фактически грешки и стилистични предпочитания. Следователно е необходимо да се установят ясни правила, които определят какво е грешка и какво не е грешка. За съжаление, методологията на ТиКюАй е предназначена само за измерване на фактически грешки и не е ефективна при висока степен на творчество от страна на преводача.

Поради невъзможността да се създаде универсален инструмент за проверяването на качеството на превода, друг опит в тази посока е разработването на „7 стъпки на Игълс“ (EAGLES¹⁸²) – персонализирана схема за проверката на качеството на превода, която предлага седем основни стъпки, необходими за извършване на успешна оценка на системите или компонентите на езиковите технологии:

1. Защо се извършва оценката?
2. Създаване на модел.
3. Определяне на характеристиките, показващи високо ниво на качество.
4. Създаване на подробни изисквания за оценяващата система (на базата на 2. и 3.).
5. Създаване на показатели, които да се прилагат към изискванията (специфицирани в точка 4.).
6. Тестване на подготвената методология.

¹⁸¹ <https://pdfs.semanticscholar.org/adda/d58a3789946e181f7939577ead3b7e238dc1.pdf>

¹⁸² <https://www.issco.unige.ch/en/research/projects/eagles/ewg99/7steps.html>

7. Оценяване.

Следвайки сходни принципи при проверката на качеството, Д. Гаудек (2007: 74) препоръчва по време на работата си преводачът да извърши пет основни проверки:

а) Проверка на качеството на материала, т.е. да провери дали всичко, което е трябвало да бъде преведено, действително е преведено и че преводът отговаря на всички приложими спецификации за предаване на имена, формули, заглавия, оформление и т.н.

б) Проверка на качеството на езика, стила и регистъра: проверка дали правописът, граматика, синтаксисът, терминологията, фразеологията, стилът и регистърът са (1) правилни, (2) хомогенни и (3) в съответствие с всички приложими спецификации.

в) Техническа, фактическа и семантична проверка на качеството: проверка дали цялата фактическа информация, данни, логически или хронологични последователности са адекватни и съответстват на приложимите спецификации.

г) Контрол на качеството на трансфера/превода: проверка дали всички важни и значими елементи в източника присъстват в превода (с отчитане на необходимите адаптации) и дали преводът отговаря на професионалните стандарти, спецификациите на доставчика на работа и (всички специфични ограничения, свързани с нуждите и изискванията на крайния потребител.

д) Проверка за хомогенност и непротиворечивост: проверка дали стилът, терминологията, фразеологията и регистърът са напълно хомогенни. Това е особено важно при работа с пакетни/групови преводи (материали, преведени от различни преводачи). Тези проверки са необходими, защото преводачът извършва корекции и промени в текста (Гаудек 2007: 74).

Дж. Макушина описва нивото на осведоменост за съществуващите инструменти за автоматизация на проверката на качеството, отделните подходи за осигуряване на качеството, видовете инструменти, които се използват, готовността за автоматизиране на проверката за качество и причините да не се направи проверка (Макушина 2007: 7). Според нейното проучване 86,5% от потребителите на тези инструменти са компании, предоставящи услуги за превод или локализация. Една трета от анкетиранияте съобщават, че прилагат процедури за осигуряване на качество в края на всеки превод. Малки компании прилагат проверка на качеството преди доставка; 30% от анкетиранияте ги използват както за изходните файлове, така и тези след редакция. Над 5% от анкетиранияте компании, най-вече големи, не прилагат вътрешни процедури за осигуряване на качеството, защото предпочитат да ги възлагат на външни изпълнители. Други методи за проверка на качеството (избрани от 4,62% от анкетиранияте) включват

проверка на файловете след редакция и проверка на терминологията, докато най-популярният отговор в тази категория е „зависи от проекта“. Най-малко използваната проверка е последователност на ниво дума, която често е една от най-важните проверки, но от друга страна, е много трудна и отнема много време. Най-популярните инструменти за автоматизация са вградените в Традос. Почти 17% от големите компании посочват, че използват собствени инструменти. Други отговори включват проверка на правописа на Майкрософт и Традос. Освен това са споменати SAE J2450 и моделът ЛИСА Кюей, които всъщност не са инструменти за автоматизация на проверката на качеството, а само показатели за състоянието на анализирания документ.

През 2013 г. КюТиЛаунчпад (QTLaunchPad¹⁸³) анализира кои модели се използват за оценка на качеството на превода. Близко 500 респонденти посочват, че използват повече от един модел за проверка на качеството. Причината за това е, че в определени случаи моделите зависят от областта на приложение. Това води до използването на вътрешни или модифицирани модели в допълнение към горното. Вътрешните модели са доминиращи – 45%. Опциите за проверка за качеството, включени в инструментите за компютърноподпомогнат превод, също са популярни – 32%. Най-широко използваният външен стандарт е EN 15038 (30%), следван от ISO 9000 (27%). Други нямат официален модел (17%), а 16% използват този на Асоциацията за индустриални стандарти за локализация. На въпроса кои инструменти използват, повечето респонденти отговарят, че използват вградената функционалност на инструментите за компютърноподпомогнат превод (48%) или собствените си вътрешни инструменти за оценка на качеството (39%), като в някои случаи се прилагат повече от един инструмент. Особено популярни са Ексбенч (30%) и Кюей дистилър (12%); а 22% заявяват, че въобще не използват инструменти за проверка на качеството. Ситуацията не се е променила много, както се вижда, от анкета на Традос през 2016 г. Тя се основава на отговорите на изследвания за технология (Research Technology Insights Research). Един от ключовите изводи на изследването е преобладаващото значение на качеството на превода (посочено е като 2.5 пъти по-важно от скоростта и 6 пъти по-важно от цената). 64% от анкетираните се налага системно да преработват своите проекти, а 48% трябва да разрешават несъответствия в използването на терминологията. Оценката на качеството до голяма степен продължава да бъде субективна. 59% от анкетираните

183

<http://www.qt21.eu/launchpad/content/new-goal-quality-translation.html>;
<https://aclanthology.org/2014.eamt-1.30.pdf>

изобщо не измерват качеството на превода или използват лошо определени критерии. Само 4% разчитат изцяло на официални, стандартизирани показатели за оценка на качеството. Те търсят и обратна връзка, но два пъти повече се получава субективна, отколкото обективна обратна връзка. Според същата анкета, за да се подобри качеството на превода, е необходимо да се даде приоритет на управлението на терминологията (тъй като терминологичните несъответствия са основната причина за преработка), участниците трябва да се запознаят със съществуващите международни стандарти и да приемат официален обективен подход за измерване на качеството.

В анкетата, описана в Първа глава (Фигура 23), българските преводачи посочват, че проверяват качеството на своите преводи чрез повторен преглед на вече преведения от тях текст (почти 62% от отговорилите). Останалата една трета от отговорилите използват автоматизирани методи като вградените функционалности и външните инструменти за измерване на качеството (някои от тях в комбинация със собствена редакция). Вероятно причините за тези резултати се дължат на относително негативното отношение на преводачите към програмите за проверка на качеството на превода, което пък, от своя страна, може да се дължи на недоброто качество на работа на самите програми.

Как проверявате качеството на превода?

73 отговора



Фигура 23: Методите за проверка на качеството на превода, използвани от българските преводачи

Обобщение

Качеството на превода е трудно за дефиниране. Инструментите за измерване на качеството на превода са стъпка към автоматичното идентифициране на грешките при превод. Както вградените в системите за компютърнопомогнат превод

функционалности, така и самостоятелните програми за проверка на качеството на превода могат да сравняват сегментите от изходния с тези на целевия език. Това е особено важно, когато става въпрос за измерване на точността на превода с терминологични бази. Спазването на зададената терминология е от ключово значение за успешното предаване на съдържанието на даден текст. От направените експерименти става ясно, че функционалностите за проверка на качеството на превода се справят добре с разпознаването на липсващи или непълни термини, но не отчитат промени като форми за множествено число.

Функционалностите за проверка на качеството на превода служат, за да се проверяват автоматично използваните термини, форматирането и дължината на сегментите, съгласуваността на сегментите с преводната памет, дали промените в източника (ако има) са приложени към целевия текст, дали цифрите в изходния език присъстват в целевия сегмент, дали препинателните знаци, кавичките, интервалите, главните букви, специалните символи и вградените тагове се използват в целевия сегмент правилно.

Значително по-голямо разнообразие и възможности предлагат самостоятелните програми за проверка на качеството. При сравнението между най-разпространените сред тях стана ясно, че всички засичат различни видове несъответствия, но само отделни програми разполагат с функции като персонализирани профили или възможността за търсене с регулярни изрази.

Разликата в настройките на отделните програми при определянето на сериозността на дадена грешка е повод за създаването на класификация на грешките при превод, която отчита степента на сериозност на грешките във връзка с последиците за качеството на превода.

Както се вижда от сравнението на резултатите на автоматичните средства за измерване на качеството, тези програми помагат частично за откриване на грешките. Срещат се много на брой „фалшиви“ грешки и това налага човешка намеса, за да може преводачът да прецени дали всеки отделен случай има нужда от поправка.

Заклучение

Основната цел на настоящата дисертация бе да анализира технологиите и средствата, използвани за създаването на терминологични ресурси в контекста на системите за компютърноподпомогнат превод и да предложи (без претенции за изчерпателност) унифицирана методика за тяхното създаване.

За тази цел бяха анализирани системите за компютърноподпомогнат превод, техническите им характеристики, техните компоненти, както и измененията, които тези инструменти налагат на работните процеси, образованието и обучението на преводачите. Последователно бяха разгледани характеристиките и начинът на работа на преводната памет, машинния превод и терминологичните бази: в обобщен вариант бяха представени основните им прилики, разлики, предимствата и недостатъци.

Изводът е, че преводната памет, поне към момента, е най-ценният компонент в системите за компютърноподпомогнат превод заради съхранението на данни от предишни преводи. Терминологичната база също има важно значение за преводачите, защото спестява нуждата от търсене на подходящите термини и подпомага избора на термин. Технологията, която изцяло преобръща представите за преводаческата дейност, е машинният превод и неговата употреба ще се увеличава все повече. Многоезиковите терминологични бази в системите за компютърноподпомогнат превод също са важен компонент от преводаческата работа. Те следват принципите на стандартните двуезикови терминологични бази: имат същата структура, функционират по същия начин и се подчиняват на аналогични принципи.

Направено е описание на новите функции на системите за компютърноподпомогнат превод. Разработени са допълнителни възможности за работа, например регулярни изрази, автоматичен редактор (към момента достъпен за английски и френски), управление на съдържанието за изграждане на уебсайтове, мобилни приложения и формуляри, възможност за свързване със софтуер за проследяване на извършената работа по даден проект и други.

Описано е използването на системите за компютърноподпомогнат превод в зависимост от лингвистичните характеристики на различните типове текстове. Програмите за компютърноподпомогнат превод са предназначени основно за превод на текстове със стандартизирано съдържание. Всяко изречение от текста за превод е

разделено на отделни сегменти, които след като бъдат преведени, се запазват в преводната памет. Всеки нов текст за превод се сканира автоматично за възможни съвпадения със сегментите, налични в паметта. Това е полезно, тъй като вече преведени сегменти се използват отново, а не се налага един същи израз (или дори изречение) да се превежда втори или трети път.

Разгледани са необходимите умения за работа със системите за компютърноподпомогнат превод. Направено е заключението, че професионалните преводачи са висококвалифицирани експерти, които имат не само езикови познания и умения да ги прилагат за два или повече езика, но и технически познания, както заради съдържанието, което превеждат, така и заради различните и все по-сложни инструменти и софтуер, които трябва да използват: различни системи за компютърноподпомогнат превод, управление на преводна памет, създаване и редактиране на терминологични бази, последващото редактиране на машинен превод, програмиране, работа със субтитри, творчески превод, локализация, обработка на изображения, разпознаване на глас, търсене и намиране на информация, обработка на файлове във формати като HTML, XML, XHTML и т.н.

Представени са редица проучвания, които изследват уменията и предпочитанията на преводачите в чужбина. Създадена е нова анкета, насочена специално към професионалните преводачи в България. Анализът на резултатите показва, че (с малки изключения) българските преводачи са добре запознати с инструментите за компютърноподпомогнат превод и се стремят да ги използват, макар че не всички ги намират за полезни и достъпни (подобно е и отношението към преводната памет и терминологичните бази). Преводачите очакват функционалностите на системите за компютърноподпомогнат превод да бъдат все по-интуитивни и по-опростени. Невронният машинен превод създава известни притеснения, че ако продължава да се развива и усъвършенства, може да представлява конкуренция за работата на преводачите.

Предлага се описание на основните понятия, свързани с термините, терминологичните речници и терминологичните бази. За най-подходяща от гледна точка на разработването на терминологични бази се приема дефиниция, според която „терминът е израз, принадлежащ към техническия език“ (Кис 2005: 106).

В дисертацията се разглежда структурата на терминологичните речници и техните видове. Прави се съпоставка с терминологичните бази, като са показват приликите, разликите и преимуществата при използването на терминологичните бази в

системите за компютърноподпомогнат превод. Посочени са примери за световноизвестни терминологични бази, които се използват както във, така и извън системите за компютърноподпомогнат превод.

Представя се процесът по създаване и управление на терминологичните бази в системите за компютърноподпомогнат превод. Основните начини за създаване и управление на терминологични бази са вградени в системите за компютърноподпомогнат превод. Освен вградените възможности съществуват редица разработени притурки, които могат да се използват както заедно със системите за компютърноподпомогнат превод, както и самостоятелно. Използват се и отделни програми за работа с терминология, които работят независимо от системите за компютърноподпомогнат превод.

Накратко са посочени уменията, които се изискват от преводачите за работа с терминология и терминологични бази. Те се състоят предимно в това преводачите да могат свободно да боравят с вградените функционалности в системите за компютърноподпомогнат превод, да прилагат преобразуване между различни файлови формати и уменията за търсене и извличане на информация. С първостепенна важност остава отличното познание на езика или езиците, към които принадлежат избраните термини, както и познанието за областта, към която спадат термините. С други думи, съдържателната страна на терминологичните бази: правилният подбор на термините и значението, с което се използват, и правилният им превод, има ключова роля за коректността на превода. В някои случаи се налага да се прилагат контролирани езици, за да се поддържа последователността и непротиворечивостта при употребата на термините.

Разглежда се и използването на терминологичните бази в зависимост от различните типове текст за превод. Основно внимание се обръща на изискванията за последователност и непротиворечивост при употребата на термини. Това е много важно, защото едни и същи термини имат различни значения в различни тематични области. Термините и терминологични бази като структура и предназначение се разглеждат в контекста на терминологията като начин за представяне на специфично човешко познание. Терминологичните бази не трябва да се противопоставят на традиционните терминологични речници, които няма да изгубят своята важност.

Предлага се кратък анализ на съществуващите езикови ресурси за български език, които могат да подпомогнат работата на преводачите и създаването на двуезикови и многоезикови терминологични бази (паралелни корпуси, терминологични речници), като се достига до извода, че те са недостатъчни. Наблюдава се и липса на свободностъпни системи за автоматично извличане на термини за един или повече езици. Необходимостта преводачите самостоятелно да създават терминологични бази при превода на документи от различни тематични области мотивира предложението на нова методика за създаване на терминологични ресурси.

Разгледани са редица програми за автоматично извличане на термини и са сравнени техните характеристики и функционалности: програмите варират според достъпността си (дали могат да се използват чрез абонамент, или безплатно); според това колко и кои езици поддържат (много малка част от разгледаните инструменти предлагат извличане на термини от български език, особено що се отнася до безплатните версии на програмите); в зависимост от начините на извличане на термините и настройките, които се предлагат в тази връзка; както и по отношение на възможността за поддържане на различни формати на терминологичните бази.

В дисертацията се предлага методика за полуавтоматично създаване на многоезикова терминологична база от данни в дадена област и за дадени езици. В конкретния случай терминологичната база е двуезикова, езиците са български и английски, като посоката на превода е от английски към български, а областта е компютърната терминология. Първата стъпка при създаването на нова терминологична база е подборът на достоверен източник: паралелен корпус, който представя дадена тематична област достатъчно пълно. Следващите стъпки са: избор и работа с програма за извличане на термини от изходния език (в случая английски), ръчно извличане на преводните еквиваленти на целевия език от подравнените изречения на корпуса, проверка и техническо оформяне на базата.

При анализа на създадената терминологична база в областта на компютърната терминология се показва, че изходният корпус оказва съществено влияние върху качеството и пълнотата на извлечените термини, което, впоследствие може да повлияе на превода. От една страна, доколкото методиката се основава на паралелни езикови корпуси и автоматично извличане на термини, може да се твърди, че се предлага унифициран начин на работа при създаването на терминологични бази. От друга страна, ръчният подбор на преводни еквиваленти осигурява проверка на качеството на

подбраните термини, като трябва да се посочи, че това в някои случаи може да отнеме много време.

Предложената методика за разработване на терминологични бази не може да се разглежда като универсално решение за подобряване и улесняване на работата на преводачите, и има известни недостатъци (основно включването на ръчния подбор като компонент), но е първи опит към разработването на специализиран софтуер за извличане на термини и създаването на двуезикови терминологични бази, в които единият език е български.

Разгледани са възможностите за измерване на качеството на превода в системите за компютърнопомогнат превод, които имат вградени функционалности за това. Те служат за автоматична проверка на използваните термини; форматирането и дължината на сегментите; съгласуваността на сегментите с преводната памет; дали промените в изходния език (ако има такива) са приложени към целевия текст; дали числовите изрази, специалните символи, препинателните знаци, главните букви и под. се използват правилно. Повече възможности предлагат самостоятелните програми за проверка на качеството на превода. Сравнението между най-разпространените програми за проверка на качеството показва, че имат съотносими възможности, но разликата в настройките им може да доведе до различни резултати. Това е поводът да се създаде класификация на грешките при превод, която отчита степента на сериозност на грешките във връзка с последиците за качеството на превода.

Бяха направени два експеримента, за да се провери поведението на програмите за проверка на качеството при превод от английски на български език. При сравнение на резултатите на автоматичните средства за измерване на качеството на превода се вижда, че тези програми помагат частично за откриване на грешките при превод. Отчитат се „фалшиви“ грешки и това налага преценка от преводача дали всеки регистриран случай има нужда от поправка.

Като цяло в изследването се предлага описание, анализ и сравнение на съвременните средства, които се използват 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. Всичко това е насочено към

постигането на основната цел – да се разработи методика за (полу)автоматично създаване на терминологични ресурси за дадена специализирана област.

Приноси на дисертацията

Приносите, които настоящото изследване предлага, могат да се разделят на теоретични и на такива с практическа насоченост. Теоретичните приноси са както следва:

- Представянето, описването и анализа на възможностите на системи за компютърноподпомогнат превод и промените, които те носят със себе си;
- Преглед на актуалните проучвания в областта на компютърноподпомогнатия превод;
- Проучване на наличните начини за създаване на терминологични бази за системите за компютърноподпомогнат превод;
- Описание и анализа на системите за проверка на качеството.

Приносите с практическа насоченост са както следва:

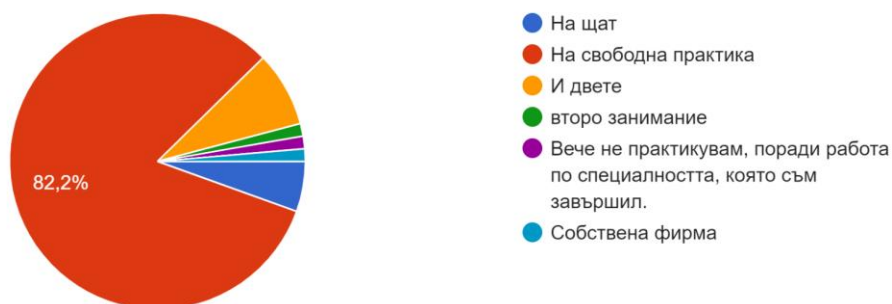
- Разработването на анкета, конкретно насочена към преводачите в България;
- Разработването на методология за полуавтоматично създаване на терминологични ресурси за български език за системите за компютърноподпомогнат превод, като се вземат под внимание някои ограничения (безплатен софтуер, поддръжка на български и в областта на компютърната терминология);
- Разработването на унифицираща класификация на грешките при превод.

Приложение 1

По време на работата по дисертацията бяха разгледани различни анкети, насочени към преводачите и изследващи техните предпочитания, професионални навици и виждания относно инструментите за компютърнопомогнат превод. Тези изследвания обхващат различни аспекти от преводаческата индустрия и мненията на преводачи. Настоящата анкета сред преводачите в България бе създадена, за да се провери дали заключенията от международните анкети са валидни и за България. Отговорите са копирани от формуляра за анкети на Гугъл.

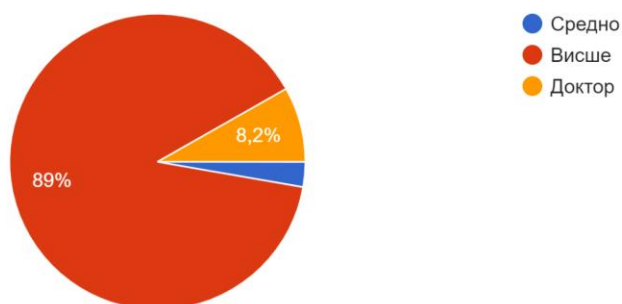
Каква е Вашата професионална заетост?

73 отговора



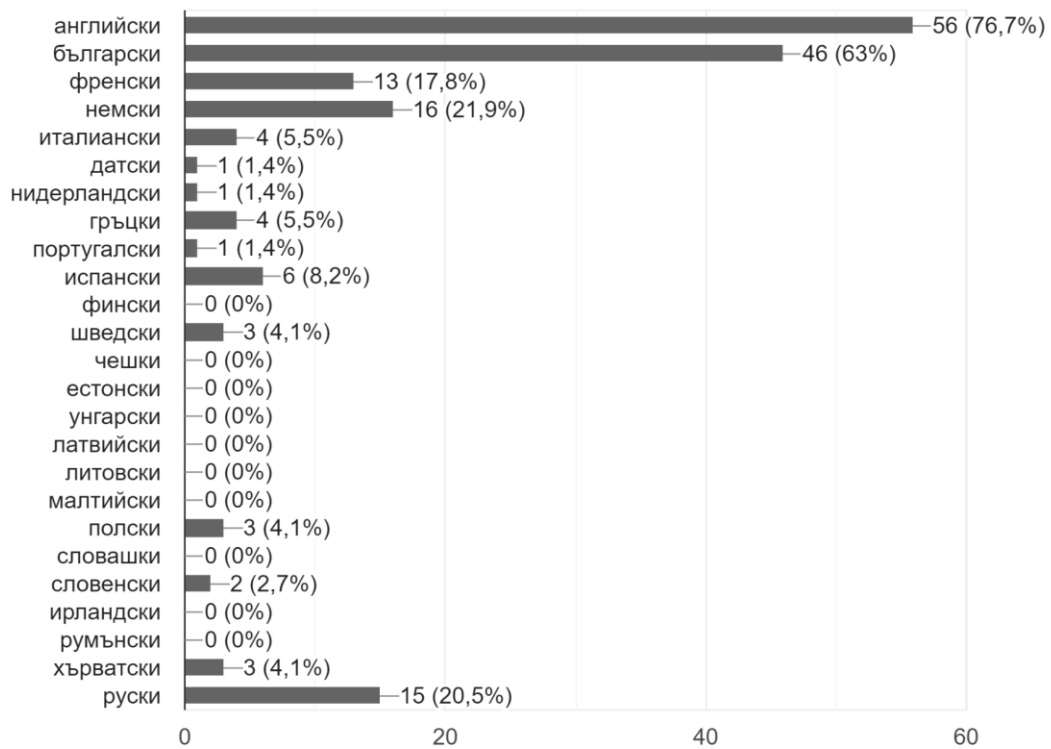
Какво е Вашето образование?

73 отговора



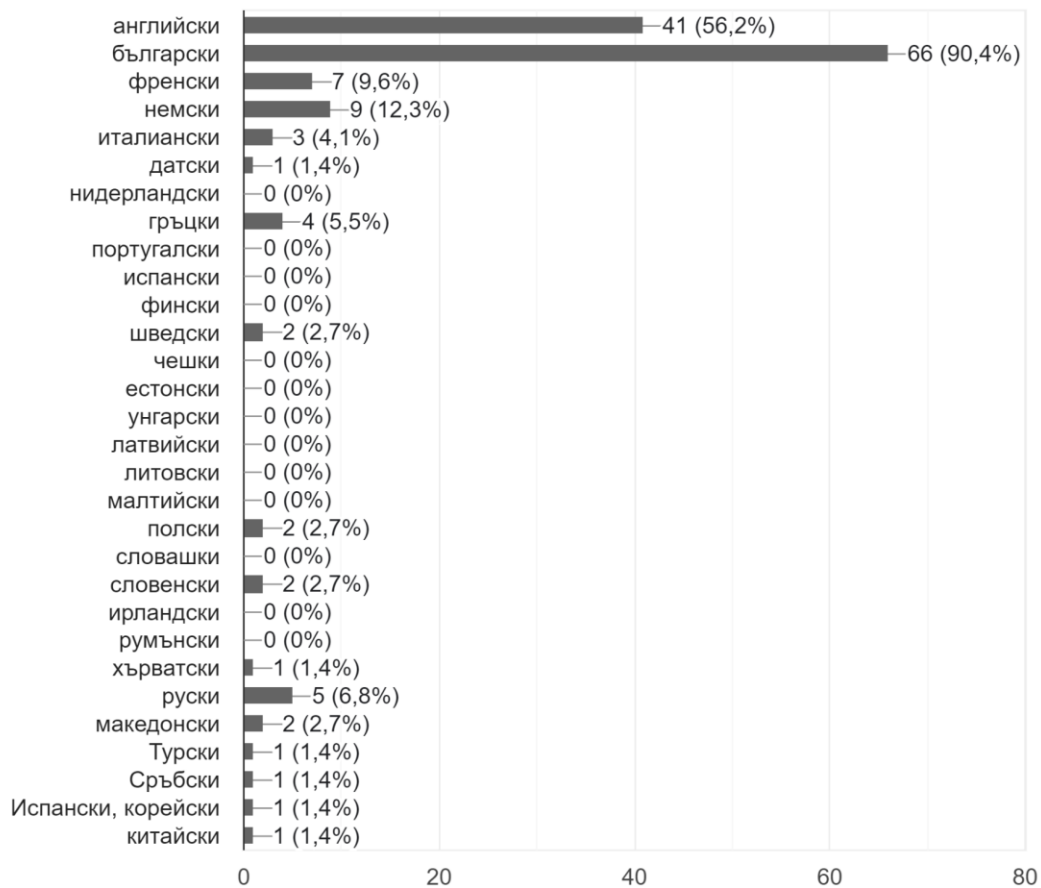
От кои езици превеждате?

73 отговора



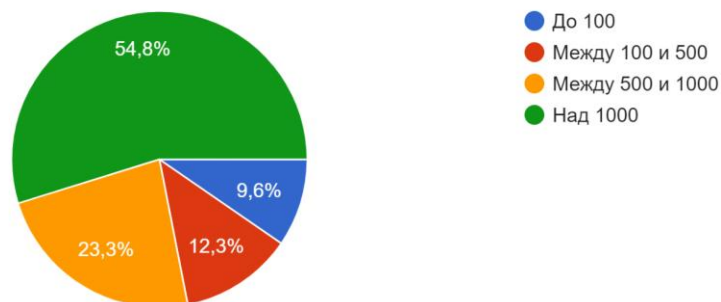
На кои езици превеждате?

73 отговора



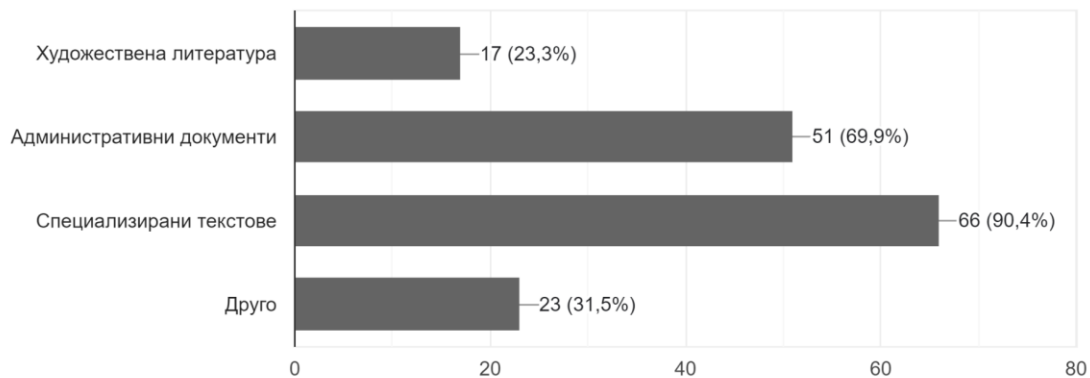
Колко стандартни страници превеждате приблизително на година?

73 отговора



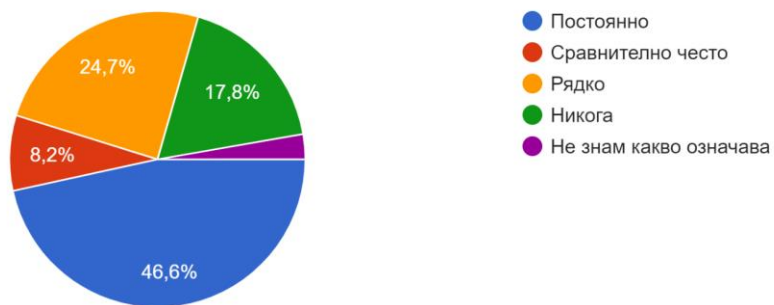
Какви видове текстове превеждате?

73 отговора



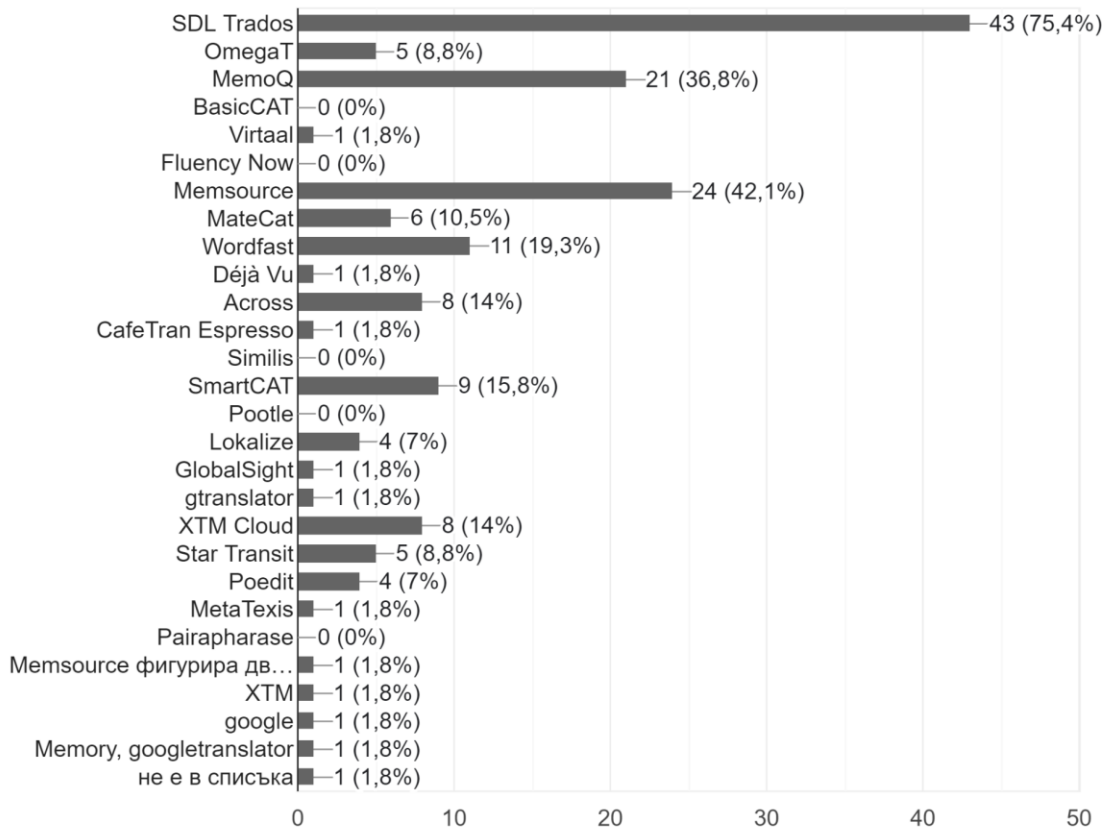
Използвате ли инструменти за компютърнопомогнат превод?

73 отговора



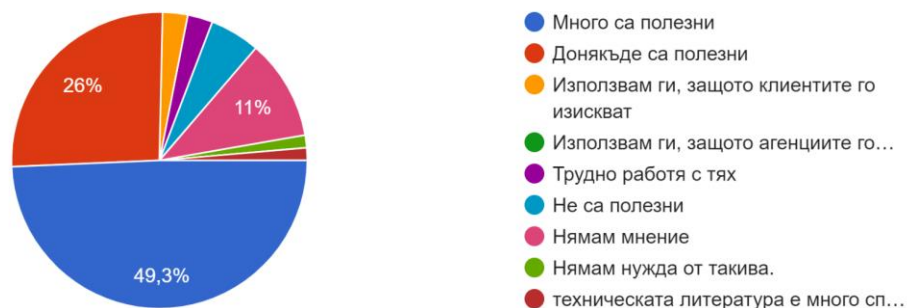
Ако да, кои инструменти за компютърнопомогнат превод използвате?

57 отговора



Как оценявате употребата на инструментите за компютърнопомогнат превод?

73 отговора



Как според Вас трябва да се развият и подобрят инструментите за компютърнопомогнат превод? 55

отговора

нямам мнение

Да имат адекватни инструкции и на български език. Не всички преводачи ползват английски
По-добра база данни.

Да са по-точни при повторенията.

Подобрен OCR, гъвкавост при работа с различни видове файлове, подобрен контрол върху оформлението на целевите файлове и по-добра файлова съвместимост. Връзка с онлайн ресурси.

Не използвам в практиката си и нямам мнение по този въпрос.

Трябва да са по-евтини!!!

Аз съм доволна от сегашното ниво.

Обогатяване на речника и включване на идиоматични изрази, да се работи за по - добра граматика и лексикална прецизност

да се обогати базата данни, за да бъдат повече двойки езици пълноценни помощници при превод

В посока форматиране и по-лесен превод на fuzzy matches при тагове, както и в посока quality assurance.

Да се опростят, защото станаха твърде сложни, с все повече и повече функционалности, повечето от които безполезни за масовия потребител. Повече интегриране на различни добавъчни функции.

X

Да са достъпни.

Улесняване на достъпността

На първо място, да могат да се превеждат pdf файлове или да има вграден конвертор. Да има по-големи възможности за визуализация, редактиране и форматиране на текста на целевия език и за сравнението му с изходния текст, без да се налага използване на MS Word. Подобряване на възможностите за обмен и добавяне на преводачески памети. Усъвършенстване на автоматизирания превод. Осигуряване на по-големи облачни пространства за съхранение, гарантиращи конфиденциалност на информацията.

Необходима е по-добра работа с преводачески памети и терминологични бази; възможност за директно търсене на термини напр. в IATE

Те непрекъснато се развиват и подобряват с оглед удобство и леснота.

Имат нужда от по-добър интерфейс, а онлайн версиите трябва да бъдат цялостно подобрени както за по-лесна употреба, така и за по-ефективна работа.

не смятам, че мога да дам съвет

Да се подобри преводът по подразбиране.

Необходимо е подобряване на съвместимостта между проекти, създавани с помощта на различни инструменти. По-лесно създаване на преводаческа памет от документи, превеждани без помощта на инструмент.

По-унифицирани функции и интерфейс, по-голяма достъпност в ценово отношение

Да са още по-полезни :D

достатъчно развити са

Не трябва.

Както и да се развият няма да могат да заместят живия преводач.

Не знам, добре са си, обаче QA дава много false positives

Да станат по-усъвършенствани спрямо българския

При някои има проблеми с функциите Cut - Copy - Paste и трябва да се използват клавишни комбинации, което за мен лично не е особено удобно

Следва да се комбинират с гласово разпознаване, за да може преводачът да не набира, а да диктува текста на превода.

Много сме далеч от достигането на адекватни инструменти за специализирани преводи.

Трябва да предоставят подобрени възможности за QA

Нямам забележки

Не знам, не работя с такъв софтуер.

Нямам мнение

Нямам мнение.

нямам мнение

Да станат по-интуитивни и самообучаващи се.

Не знам

Не съм достатъчно запозната с този вид превод, за да потвърдя

С промени в интерфейса.

По-прецизна проверка на правописа, добър вариант би бил, ако е възможно да се интегрира диктува е на текст

Ползвам безплатна версия, а тя е доста ограничена. За платените нямам поглед.

Смятам, че някои, в това число Trados, Smart at, memoq, работят почти перфектно

AI

В някои инструменти липсват елементарни функции за редакция на текст.

трябва да бъдат по-бързи в търсенето и предлагането на алтернативи, и да са още значително "по-умни" разбира се, напр. да поддържат гласови команди

Не мога да преценя.

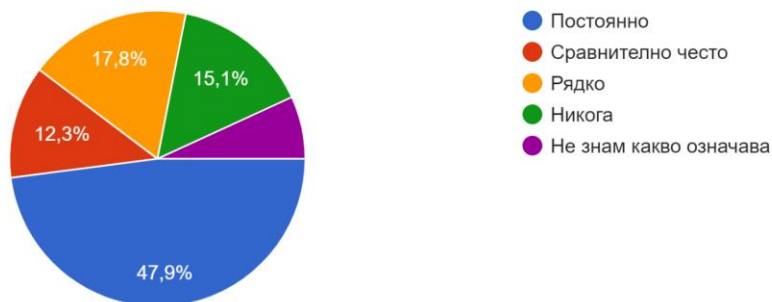
Да дават по-малко грешки при експорт на файл.

Да бъдат по-модуларни, т.е., по леки и съдържащи само основни функции и останалите им функции да се добавят поотделно като plugin, както в Trados, която въпреки това е прекалено тежка програма.

САТ инструментите трябва да взимат пред вид особеностите на българския език. За хубаво или лошо, те са ориентирани предимно от/към английски език. Инак са добър помощник...

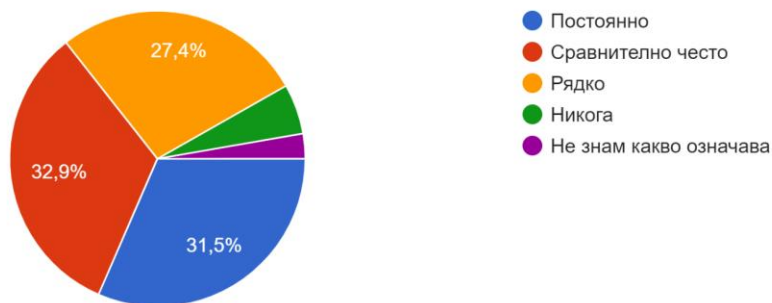
Използвате ли преводна памет?

73 отговора



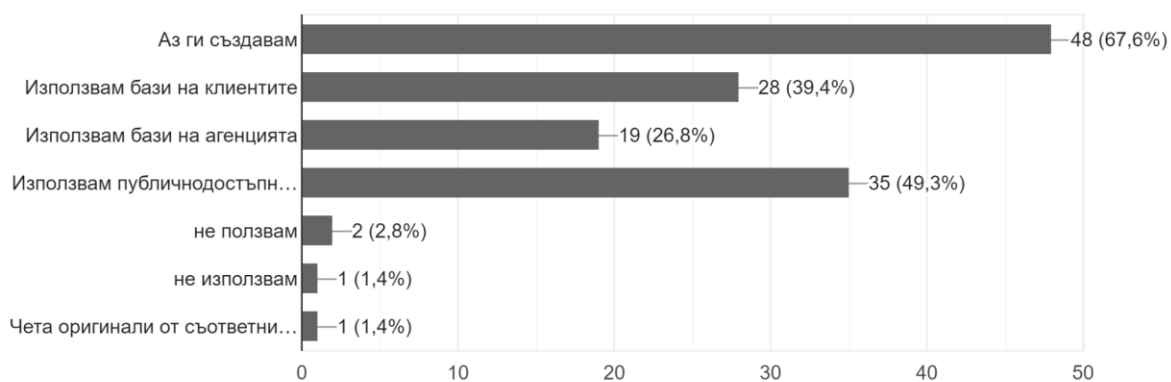
Използвате ли терминологични бази?

73 отговора



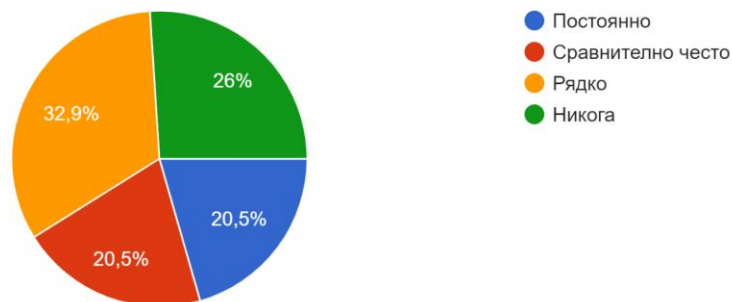
Ако използвате терминологични бази, какъв е техният произход?

71 отговора



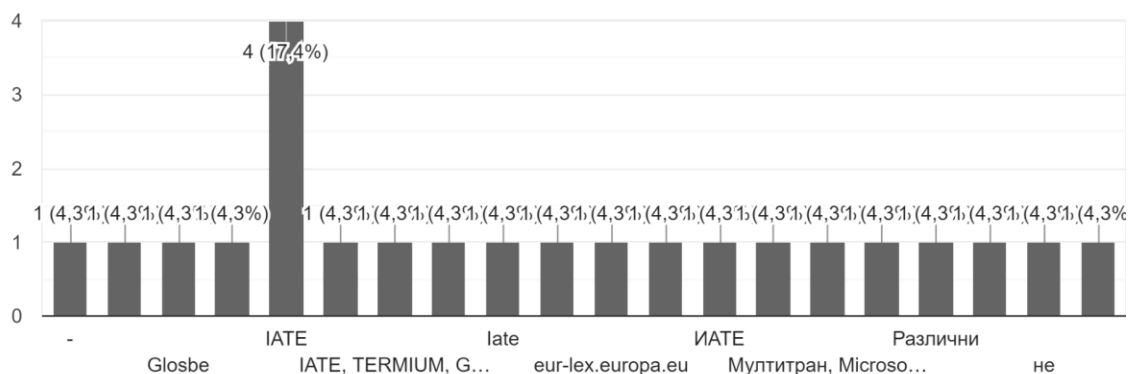
Използвате ли публичнодостъпни терминологични бази?

73 отговора



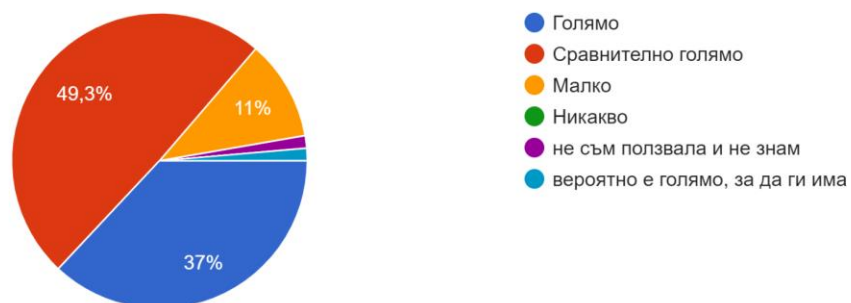
Ако използвате публичнодостъпни терминологични бази, кои са те?

23 отговора



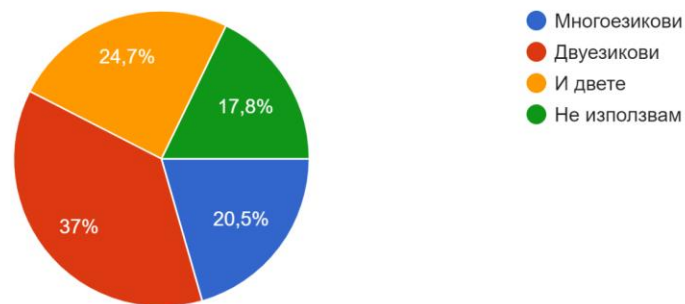
Според Вас какво е значението на терминологичните бази за качеството на превода?

73 отговора



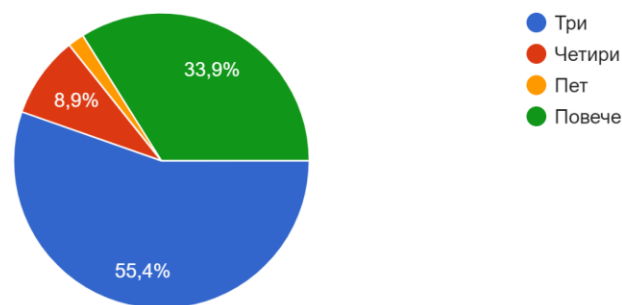
Използвате ли двуезикови или многоезикови терминологични бази?

73 отговора



Ако използвате многоезикови терминологични бази, от колко езика са създадени?

56 отговора



Ако сами създавате терминологични бази, какви методи използвате? 34 отговора

Не създавам

Създавам файлове

Не използвам

Запазвам намерените точни преводи в мултитерм.

Нямам мнение

Метод

Използвам ресурсите на съответната програма за компютърнопомогнат превод.

Инструментите за компютърнопомогнат превод и друг специализиран софтуер

Следвам спецификите на тематиката (кулинария, свибодно време и др.)

записване във файл (word, excel) и MultiTerm на SDL Trados

Създавам свои двуезикови терминологични бази в области, в които работя често - най-често това е техническа лексика - автомобилостроене, машиностроене, пътно строителство. Многоезиковите терминологични бази, които използвам, най-често са предоставени от клиента или са публичнодостъпни.

notepad

Не създавам.

Речников подход

Чрез SDLMultiTerm

Aligne

Записвам си преведените термини, в случай, че отново се сблъскам с тях.

не създавам

Не

Генерирани от превод в програмата

Excel, MultiTerm, ...

не

Използвам SDL MultiTerm или просто файл в excel

Ръчно

Excel

Зависи, Word, Excel, които трябва да конвертирам, както и инструментите, предлагани от Trados.

MultiTerm

manual, term-mining

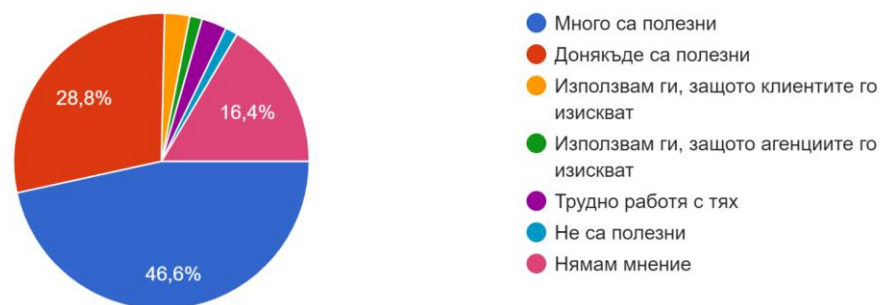
Сравнение на тестове на двата езика

Записвам ги в собствени файлове, тетрадки.

Съответната функция за добавяне на термини в програмата за компютърноподпомогнат превод

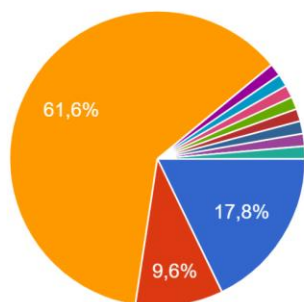
Как оценявате работата с терминологични бази?

73 отговора



Как проверявате качеството на превода?

73 отговора

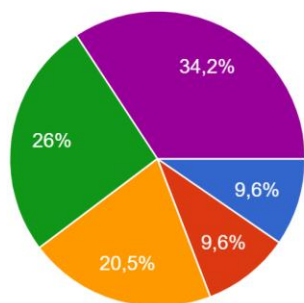


- Чрез функционалност на инструмен...
- Чрез програма, външна за инструме...
- Чрез повторно преглеждане
- Не проверявам
- В зависимост от указанията на заяв...
- консултация със специалисти в обл...
- С помощта на вградените функции...
- QA и собствена редакция

▲ 1/2 ▼

Използвате ли инструменти за локализация?

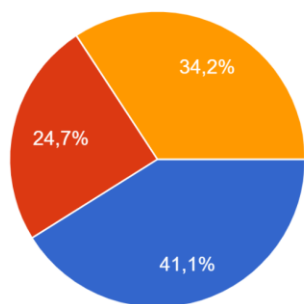
73 отговора



- Постоянно
- Сравнително често
- Рядко
- Никога
- Не познавам такива

Запознати ли сте с ISO стандартите за качество на превода?

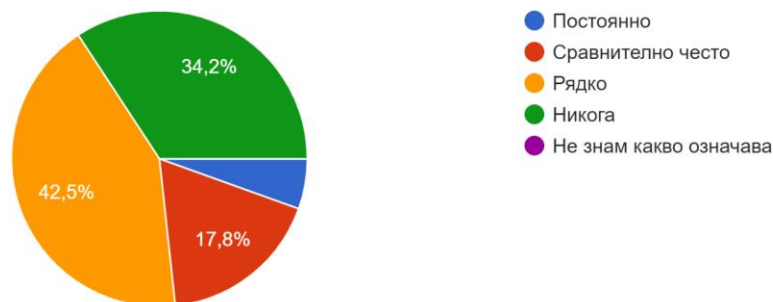
73 отговора



- Да
- Не
- Отчасти

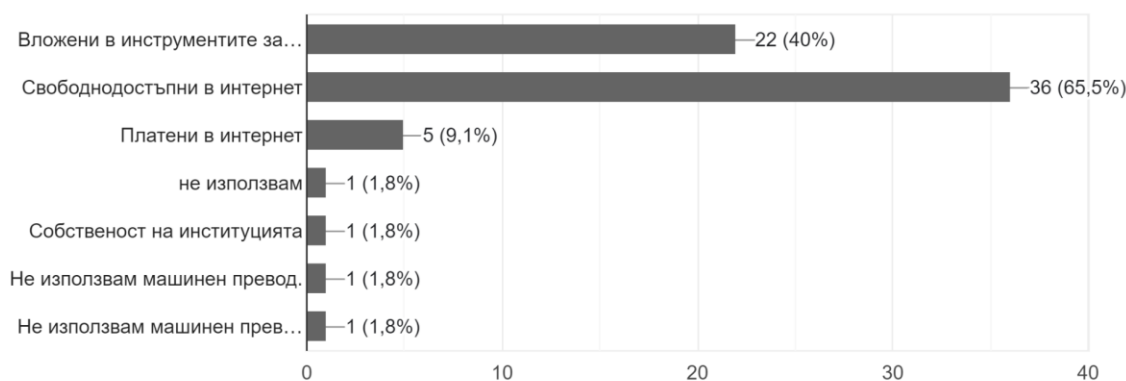
Използвате ли машинен превод?

73 отговора



Ако използвате машинен превод, какви приложения ползвате?

55 отговора



Как според Вас трябва да се развие и подобри машинният превод? 56 отговора

нямам мнение

-

Да се увеличи обемът на въведените термини

По-развит софтуер.

Като се даде възможност на специалистите да го допълват.

Вероятно чрез подобряване на алгоритмите за постигане на по-високо ниво на точност

Не използвам машинен превод

Да може да взаимодейства с паметта и терминологичната база на програмите за компютърноподпомогнат превод и да взема термини и фрази оттам. Да се диверсифицира по области и да разграничава специфични термини по области.

По - добро програмиране, работа в екип на филолози и програмисти

.

Не одобрявам използването му и не вярвам да може да се развие до степен да е личи, че е превод.

Той се развива постоянно и става все по-добър. Грешките стават все по-малко очевидни и трудни за откриване.

Х

Да дава възможност за подобряване на предложения превод.

Подобряване на търсачките и обогатяване на базите данни.

Да стане по-адекватен и по-точен.

Необходимо е допълнително обучение на ИИ, тъй като при необучен ИИ машинният превод се придържа към синтаксиса на изходния език и използва неподходящи колокации и термини. Това не се наблюдава при добре обучени програми.

Не смятам, че на този етап е възможно кой знае какво развитие на машинния превод, въпреки че непрекъснато се работи в тази посока.

Не трябва. Ще останем без работа.

нямам компетентност в областта

Да се приспособява според тематиката.

Машинният превод в доста области става все по-добър с натрупването на все по-големи бази данни и така или иначе изпреварва очакванията ми, затова не бих могла да отговоря.

Нямам мнение.

да стане по-човешки" :D

Не трябва повече. Преводачите ще станат излишни.

Да помага на преводача, не да се опитва да го замества и да се плаща за редакция на нещо, което е почти изцяло нов превод.

Ако можеше да превежда вместо мен би било безценно, но това няма как да стане.

Терминологията да е вярна

Тотално

да се подобри качеството на превода на български език

Да се работи по усъвършенстване на алгоритмите на невромашинния превод.

Не мисля, че може да се подобри дотолкова, че да се използва в професионалните преводи.

Бъ... сериозно ли?

Добавяне на повече корпуси с реални преводи и др.

С повече примери ще се подобри, няма как да стане идеален.

Нямам интерес да се подобрява

Тъй като се занимавам с художествен превод, го смятам за неуместен в случая.

с повече техническа терминология

Да стане по-интуитивен и самообучаващ се.

Не е приложимо

Не съм достатъчно запозната

С въвеждане на невронни мрежи и качествено им обучение

Не работя и нямам намерение да работя с машинен превод.

преглед на предложени от случайни лица "преводи", които служат само за забавление, а реано не са верни изобщо.

Често срещаните проблеми са свързани с непостоянство в термините, проблеми с рода и съгласуването, както и незачитане на контекста

AI

Поверителността на информацията

Като се събират непрекъснато словосъчетания и обогатяват, съответните термини на различни езици би трябвало да се уеднаквят.

Не знам

Към разпознаване на различните родове като начало

С увеличаване на обхвата.

Трябва да се проверява от преводачи.

Създаване на огромни бази данни, съобразени със конкретен език

Приложение 2

Представят се извлечените едноезикови термини на английски с програмата Термостат. Форматът по подразбиране е .txt. Термините са подредени в табличен вид, като към всеки термин е приложена честотата му в текста, специфичността му и вариантите, с които се е срещнал в текста.

Candidat de regroupement	Fréquence	Spécificité	Variantes orthographiques	Matrice
wireless network	42	80.44	wireless network___wireless networks	Nom Nom
hard drive	37	75.37	hard drive___hard drives	Adjectif Nom
google doc	23	58.93	google doc___google docs	Nom Nom
facebook account	15	47.02	facebook account	Nom Nom
facebook page	10	37.73	facebook page___facebook pages	Nom Nom
window xp	10	37.73	window xp___windows xp	Nom Nom
internet access	9	35.58	internet access	Nom Nom
parental control	9	33.73	parental control___parental controls	Adjectif Nom
google drive	8	33.3	google drive	Nom Nom
web page 8	33.3	web page___web pages	Nom Nom	
email account	8	33.3	email account___email accounts	Nom Nom
wireless router	7	30.85	wireless router	Nom Nom
product key	7	30.85	product key	Nom Nom
page break	7	30.85	page break___page breaks	Nom Nom
graphics card	7	30.85	graphics card___graphics cards	Nom Nom
mobile phone	7	30.85	mobile phone	Adjectif Nom
internet connection	7	30.85	internet connection___internet connections	Nom Nom
email address	6	28.19	email address___email addresses	Nom Nom
text box 6	28.19	text box	Nom Nom	
wireless internet	6	28.19	wireless internet	Nom Nom
sd card 6	28.19	sd card___sd cards	Nom Nom	
android phone	6	28.19	android phone	Adjectif Nom
sd card 6	28.19	sd card	Adjectif Nom	
flash drive	5	25.26	flash drive	Adjectif Nom
facebook chat	5	25.26	facebook chat	Nom Nom
external hard drive	5	25.26	external hard drive	Adjectif Adjectif Nom
internet browser	5	25.26	internet browser___internet browsers	Nom Nom
facebook message	5	25.26	facebook message___facebook messages	Nom Nom
facebook friend	5	25.26	facebook friend___facebook friends	Nom Nom
profile picture	5	25.26	profile picture___profile pictures	Nom Nom
antivirus software	5	25.26	antivirus software___antivirus softwares	Nom Nom
google play	5	25.26	google play	Nom Nom
network name	5	25.26	network name___networks name	Nom Nom
cell content	5	25.26	cell content	Nom Nom

much ram	4	21.96	much ram	Adjectif Nom	
background image	4	21.96	background image	Nom Nom	
facebook notification		4	21.96	facebook notifications	Nom Nom
font size	4	21.96	font size	Nom Nom	
android tablet	4	21.96	android tablet	Adjectif Nom	
facebook messenger		4	21.96	facebook messenger	Nom Nom
factory setting	4	21.96	factory settings	Nom Nom	
video call	4	21.96	video call___video calls	Adjectif Nom	
android device	4	21.96	android device	Adjectif Nom	
recycle bin	4	21.96	recycle bin	Nom Nom	
facebook password	4	21.96	facebook password	Nom Nom	
youtube video	4	21.96	youtube video___youtube videos		Nom Nom
hard disk	7	20.23	hard disk	Adjectif Nom	
contact list	4	19.61	contact list	Nom Nom	
much space	6	19	much space	Adjectif Nom	
text document	3	18.11	text document	Nom Nom	
iphone weather application		3	18.11	iphone weather application	Nom Nom Nom
wireless network password		3	18.11	wireless network password	Nom Nom Nom
speed test	3	18.11	speed test___speed tests	Nom Nom	
thermal paste	3	18.11	thermal paste___thermal pastes		Adjectif Nom
pc monitor	3	18.11	pc monitor	Nom Nom	
external drive	3	18.11	external drive	Adjectif Nom	
network cable	3	18.11	network cable	Nom Nom	
spell checker	3	18.11	spell checker	Nom Nom	
audio file	3	18.11	audio file	Adjectif Nom	
old file	3	18.11	old files	Adjectif Nom	
google chrome	3	18.11	google chrome	Nom Nom	
user account	3	18.11	user account___users accounts		Nom Nom
video card	3	18.11	video card___video cards		Adjectif Nom
side bar	3	18.11	side bar	Nom Nom	
gb ram	3	18.11	gb ram	Nom Nom	
wifi password	3	18.11	wifi password	Nom Nom	
page margin	3	18.11	page margin___page margins		Nom Nom
command prompt	3	18.11	command prompt	Nom Nom	
cell phone	3	18.11	cell phone	Nom Nom	
youtube account	3	18.11	youtube account___youtube accounts		Nom Nom
page size	3	18.11	page size	Nom Nom	
internet speed	3	18.11	internet speed___internet speeds		Nom Nom
ipad mini	3	18.11	ipad mini	Nom Nom	
blue screen	3	18.11	blue screen	Adjectif Nom	
dual link	3	18.11	dual link	Adjectif Nom	
flash drive	3	18.11	flash drive	Nom Nom	
macbook pro	3	18.11	macbook pro	Nom Nom	
network key	3	18.11	network key	Nom Nom	
ethernet cable	3	18.11	ethernet cable	Nom Nom	

hp laptop	3	18.11	hp laptop	Nom Nom	
column width	3	18.11	column width	Nom Nom	
window calculator	3	18.11	windows calculator	Nom Nom	
network connection	3	18.11	network connection	Nom Nom	
dual core	3	18.11	dual core	Adjectif Nom	
page orientation	3	18.11	page orientation	Nom Nom	
warning message	3	18.11	warning message	Nom Nom	
full screen	3	18.11	full screen	Adjectif Nom	
samsung galaxy	3	18.11	samsung galaxy	Adjectif Nom	
dell laptop	3	18.11	dell laptop	Nom Nom	
voice message	3	18.11	voice message	voice messages	Nom Nom
itunes library	3	18.11	itunes library	itunes libraries	Nom Nom
safe mode	3	18.11	safe mode	Adjectif Nom	
density memory	3	18.11	density memory	Nom Nom	
galaxy note	3	18.11	galaxy note	Nom Nom	
twitter account	3	18.11	twitter account	Nom Nom	
background color	3	18.11	background color	Nom Nom	
hidden file	3	18.11	hidden files	Adjectif Nom	
usb port	3	18.11	usb port	Nom Nom	
old phone	3	18.11	old phone	Adjectif Nom	
pop up	3	18.11	pop ups	Adjectif Nom	
free software	3	18.11	free software	Adjectif Nom	
text alignment	3	18.11	text alignment	Nom Nom	
much memory	3	15.64	much memory	Adjectif Nom	
programming language	3	13.96	programming language	programming languages	Nom Nom
unblock someone	2	13.3	unblock someone	Nom Nom	
upload speed	2	13.3	upload speed	Nom Nom	
gaming mouse	2	13.3	gaming mouse	Nom Nom	
multiple computer	2	13.3	multiple computers	Adjectif Nom	
mouse pad	2	13.3	mouse pad	Nom Nom	
security answer	2	13.3	security answer	Nom Nom	
factory reset	2	13.3	factory reset	Nom Nom	
micro sd card	2	13.3	micro sd card	Nom Nom Nom	
share photo	2	13.3	share photos	Nom Nom	
ipad mini screen	2	13.3	ipad mini screen	Nom Adjectif Nom	
picture file	2	13.3	picture file	picture files	Nom Nom
mobile version	2	13.3	mobile version	Adjectif Nom	
low density memory	2	13.3	low density memory	Adjectif Nom Nom	
wireless signal	2	13.3	wireless signal	Nom Nom	
boot device	2	13.3	boot device	Nom Nom	
account name	2	13.3	account name	Nom Nom	
audio powerpoint	2	13.3	audio powerpoint	Adjectif Nom	
video file	2	13.3	video file	video files	Nom Nom
dropbox file	2	13.3	dropbox file	Nom Nom	
automatic reply	2	13.3	automatic reply	automatic replies	Adjectif Nom

say user	2	13.3	said user	Adjectif	Nom
internet datum	2	13.3	internet data	Nom	Nom
bit computer	2	13.3	bit computer	Nom	Nom
gaming keyboard	2	13.3	gaming keyboard	Nom	Nom
video call	2	13.3	video call	Nom	Nom
iphone contact	2	13.3	iphone contacts	Nom	Nom
computer monitor	2	13.3	computer monitor	Nom	Nom
gaming desktop	2	13.3	gaming desktop	Nom	Nom
wifi range	2	13.3	wifi range	Nom	Nom
battery life	2	13.3	battery life	Nom	Nom
hdmi cable	2	13.3	hdmi cable	Nom	Nom
confirmation code	2	13.3	confirmation code	Nom	Nom
pdf file	2	13.3	pdf file	Nom	Nom
total security	2	13.3	total security	Adjectif	Nom
quad core	2	13.3	quad core	Nom	Nom
laptop battery	2	13.3	laptop battery	Adjectif	Nom
email signature	2	13.3	email signature	Nom	Nom
spell check	2	13.3	spell check	Nom	Nom
facebook timeline	2	13.3	facebook timeline	Nom	Nom
google map	2	13.3	google maps	Nom	Nom
email message	2	13.3	email message	email messages	Nom
printer cartridge	2	13.3	printer cartridge	Nom	Nom
movie maker	2	13.3	movie maker	Nom	Nom
shape fill color	2	13.3	shape fill color	Nom	Nom
usb cable	2	13.3	usb cable	Nom	Nom
pc tuneup	2	13.3	pc tuneup	Nom	Nom
wifi router	2	13.3	wifi router	Nom	Nom
dell inspiron	2	13.3	dell inspiron	Nom	Nom
airplane mode	2	13.3	airplane mode	Nom	Nom
pc restarts	2	13.3	pc restarts	Nom	Nom
boolean query	2	13.3	boolean query	boolean queries	Adjectif
dvd drive	2	13.3	dvd drive	dvd drives	Nom
web site	2	13.3	web site	web sites	Nom
web browser	2	13.3	web browser	web browsers	Nom
screen resolution	2	13.3	screen resolution	Nom	Nom
facebook friend list	2	13.3	facebook friends list	Nom	Nom
digit pin	2	13.3	digit pin	Nom	Nom

Приложение 3

Представя се двуезиковата терминологична база с компютърни термини на английски и български. Форматът е ексел, защото е един от най-разпространените формати за работа с терминология. Добавени са три допълнителни полета (за синоними, бележки и части на речта).

EN	BG	Синоним	Част на речта	Бележки
facebook	фейсбук			
laptop	лаптоп			
internet	интернет			
iphone	iphone			
password	парола			
wireless	безжичен			
router	рутер			
ipad	ipad			
email	електронна поща			
file	файл			
computer	компютър			
pc	настолен компютър			
google	google			
folder	папка			

itunes	itunes
wireless network	безжична мрежа
hard drive	харддиск
powerpoint	powerpoint
website	уебсайт
wifi	wifi
antivirus	антивирусна
usb	USB
mac	мак
window	прозорец
video	видео
desktop	работен плот
browser	браузър
youtube	youtube
doc	док
google doc	google doc
gaming	игри
download	Изтегли
email	имейл
program	програма

server	сървър
color	цвет
motherboard	дънна платка
ram	рам
dropbox	dropbox
network	мрежа
monitor	монитор
skype	скайп
icon	икона
printer	принтер
app	приложение
facebook account	Фейсбук профил
screen	екран
chrome	chrome
drive	драйвер
notification	уведомление
gmail	gmail
hdd	hdd
macbook	macbook
disk	диск

modem	модем
website	уебсайт
gb	gb
ios	ios
ssd	ssd
xp	xp
bookmark	отметка
instagram	инстаграм
username	потребителско име
id	идентификатор
page	страница
software	софтуер
web	уеб
update	актуализация
notepad	notepad
login	логин
app	приложение
dvd	DVD
shortcut	пък път
facebook page	фейсбук страница

pdf	pdf
window xp	window xp
document	документ
keyboard	клавиатура
chat	чат
scan	сканиране
account	акаунт
adapter	адаптер
photoshop	фотошоп
internet access	интернет достъп
sd	sd
photo	снимка
message	съобщение
parental control	родителски контрол
upload	качване
skydrive	skydrive
cpu	процесор
web page	уеб страница
android	Android
email account	имейл акаунт

xbox	xbox
google drive	google диск
phone	телефон
os	os
webmail	уеб поща
webpage	уеб страница
page break	разделител на страница
gpu	гру
wireless router	безжичен рутер
utorrent	utorrent
graphics card	графична карта
product key	продуктов ключ
internet connection	интернет връзка
mobile phone	мобилен телефон
font	шрифт
cable	кабел
spam	спам
admin	администратор
text box	текстово поле
icloud	icloud

wireless internet	безжичен интернет
android phone	телефон с Android
email address	имейл адрес
homepage	начална страница
sd card	sd карта
hdmi	hdmi
bios	биос
url	URL адрес
twitter	туитър
firefox	firefox
java	java
sd card	sd карта
audio	аудио
text	текст
publisher	publisher
word	word
picture	снимка
add-on	add-on
backup	резервно копие
hyperlink	хиперлинк

facebook message	фейсбук съобщение
facebook chat	фейсбук чат
offline	офлайн
bluetooth	Bluetooth
facebook friend	фейсбук приятел
profile picture	снимка на профила
internet browser	интернет браузър
antivirus software	антивирусен софтуер
google play	Google Play
network name	име на мрежата
downloading	изтегляне
linux	linux
ip	ip
flash drive	флаш памет
imac	imac
toolbar	лента с инструменти
external hard drive	външен твърд диск
card	карта
tab	раздел
error	грешка

access	достъп
address	адрес
key	ключ
tablet	таблет
mouse	мишка
android device	устройство с Android
facebook notification	фейсбук известие
prompt	prompt
microsoft	Майкрософт
recycle bin	кошче
youtube video	youtube видео
pic	снимка
libreoffice	libreoffice
font size	размер на шрифта
ghz	GHz
video call	видео разговор
http	http
android tablet	таблет с android
factory setting	фабрична настройка
ethernet	ethernet

facebook messenger	фейсбук месинджър
unblock	деблокиране
facebook password	фейсбук парола
chromebook	chromebook
timeline	timeline
webcam	уебкамера
outlook	outlook
virus	вирус
repeater	повторител
playlist	плейлист
favorite	любими
cellphone	мобилен телефон
ping	пинг
version	версия
com	com
battery	батерия
processor	процесор
hard disk	твърд диск
slide	слайд
mode	режим

hotspot	hotspot
contact list	списък с контакти
cd	cd
image	изображение
headphones	слушалки
brightness	яркост
default	по подразбиране
geforce	geforce
twitter account	акаунт в Twitter
gb ram	gb ram
pc monitor	компютърен монитор
nvidia	nvidia
ipod	ipod
side bar	странична лента
smartphone	смартфон
wireless network password	парола за безжична мрежа
fps	fps
inspiron	inspiron
gifs	gifs
thermal paste	термична паста

google chrome	Google Chrome
youtube account	акаунт в youtube
wifi password	wifi парола
onedrive	onedrive
dual core	Двухядрен процесор
exe	екз
voice message	гласово съобщение
imessage	imessage
network connection	мрежова връзка
internet speed	скорост на интернет
video card	видео карта
norton	нортън
network key	мрежов ключ
speed test	тест за скорост
text document	текстов документ
network cable	мрежов кабел
snapchat	снапчат
sapo	sapo
page orientation	ориентация на страницата
ebay	ebay

jpeg	jpeg
hp laptop	лаптоп hp
ethernet cable	ethernet кабел
samsung galaxy	samsung galaxy
hidden file	скрит файл
user account	потребителски акаунт
dual link	двойна връзка
window calculator	калкулатор за прозорец
density memory	памет за плътност
audio file	аудио файл
text alignment	подравняване на текста
usb port	USB порт
itunes library	itunes library
taskbar	лента на задачите
dll	dll
firewall	защитна стена
free software	безплатен софтуер
external drive	външен диск
macbook pro	MacBook Pro
warning message	предупредително съобщение

blue screen	син екран
ipad mini	айпад мини
downloads	изтегляния
command prompt	командна конзола
slideshow	слайдшоу
iphoto	iphoto
background color	Цвят на фона
page size	размер на страницата
ubuntu	ubuntu
spell checker	проверка на правописа
page margin	поле на страницата
column width	ширина на колоната
blog	блог
gif	gif
connection	Връзка
device	устройство
dell	dell
adobe	adobe
tv	телевизия
presentation	представяне

galaxy	galaxy
menu	меню
adaptor	адаптер
beep	звуков сигнал
mic	микрофон
contact	контакт
setting	настройка
profile	профил
chart	графика
button	бутон
sound	звук
torrent	торент
format	формат
reset	нулиране
programming language	програмен език
startup	стартиране
friend	приятел
datum	данни
application	приложение
remote access	отдалечен достъп

email password	парола на електронната поща
case insensitive filesystem	нечувствителна към главни и малки букви файлова система
sound card	звукова карта
web site	уеб сайт
https	https
unix	unix
google play store	Google Play Store
inbox	входяща кутия
install window	прозорец за инсталиране
internal page	вътрешна страница
pt	pt
laptop keyboard	клавиатура на лаптоп
online picture	онлайн снимка
nvidia geforce	nvidia geforce
wireless connection	безжична връзка
gaming mouse	мишка за игри
avast	avast
dvd drive	dvd устройство
chart type	тип диаграма
friend someone	добавям някого за приятел

picasa	Picasa
printer cartridge	касета за принтер
ipad mini screen	екран на iPad мини
playstation	конзола за игри
browser plugins	добавки за браузър
unblock someone	деблокирам някого
avi	avi
sim card	СИМ-карта
page change	промяна на страницата
jpg	jpg
automatic reply	автоматичен отговор
video call	видео повикване
celeron	селерон
mac book	mac book
email signature	имейл подпис
dell inspiron	dell inspiron
laptop battery	батерия за лаптоп
dpi	dpi
external graphics card	външна графична карта
second monitor	втори монитор

asus	asus
font color	цвет на шрифта
free antivirus	безплатна антивирусна програма
adress	адрес
wireless printer	безжичен принтер
boolean query	булева заявка
picture file	картинен файл
spell check	проверка на правописа
hdmi cable	hdmi кабел
external monitor	външен монитор
pc tuneup	настройка на компютър
html	html
internet page	интернет страница
png	png
account name	име на акаунта
quad core	четириядрен
debian	debian
share photo	сподели снимка
javascript	javascript
upload speed	скорост на качване

internet datum	интернет данни
gimp	gimp
video file	видео файл
acer	acer
usb multiport	usb мултипорт
avg	ср
pdf file	pdf файл
signal repeater	повторител на сигнала
facebook friend list	списък с приятели във facebook
mobile version	мобилна версия
email message	имейл съобщение
zip file	zip файл
user name	потребителско име
pc restarts	компютърът се рестартира
dropbox file	dropbox файл
iphones	iphones
audio powerpoint	аудио powerpoint
mouse pad	Подложка за мишка
blu-ray	Блу-рей
digit pin	цифров щифт

overheat	прегриване
shape fill color	цвет на запълване на формата
airplane mode	самолетен режим
battery life	живот на батерията
start button	бутон за стартиране
internet explorer	Internet Explorer
internet access	достъп до интернет
boot device	устройство за зареждане
apple logo	лого на apple
imacs	imacs
imovie	imovie
gmail account	gmail акаунт
autorecover	автоматично възстановяване
screen brightness	яркост на екрана
gaming keyboard	клавиатура за игри
facebook timeline	времева линия във фейсбук
vram	vram
web browser	уеб браузър
confirmation code	код за потвърждение
cable internet	кабелен интернет

wireless signal	безжичен сигнал
screen resolution	Резолюция на екрана
facebook setting	фейсбук настройка
uninstall	деинсталиране
facetime	facetime
version of microsoft	версия на microsoft
computer monitor	Монитор на компютър
wifi router	wifi рутер
pdf	pdf файл
iwork	работя
firmware	фърмуер
vpn	vpn
motherbord	дънна платка
teamviewer	teamviewer
directx	directx
intel	intel
screen saver	скрийнсейвър
desktop background	фон на работния плот
fb	fb
iphone contact	iphone контакт

number of word	брой на думите
lose access	губят достъп
google map	Гугъл карта
home wireless network	домашна безжична мрежа
usb cable	USB кабел
flashplayer	флаш плейър
wifi range	wifi обхват
admin password	администраторска парола
podcast	подкаст
gaming desktop	десктоп за игри
disk error	грешка на диска
battery status	състояние на батерията
gaming performance	игрово представяне
anti-virus	антивирусен
wireless internet access	безжичен интернет достъп
original file	оригинален файл
micro sd card	micro sd карта
factory reset	фабрични настройки
dvi	dvi
game mode	режим на игра

smart tv	умен телевизор
total security	пълна сигурност
movie maker	movie maker
spacing	разстояние
fill	запълвам
messenger	messenger
user	потребител
hp	hp
bandwidth	трафик
yahoo	yahoo
disk space	дисково пространство
memory card	карта памет
text file	текстов файл
command line	команден ред
attachment	прикачен файл
logo	лого

Списък на някои от използвани (по-рядко употребявани)

термини:

- *Бекенд автоматизиране* (Backend automation) е автоматизация на работното натоварване на Апи (API) за конкретно дефинирани транзакции с голям капацитет.
- *Джава* (Java) е обектно ориентиран език за програмиране, разработен и пуснат в употреба през 1995 година.
- *Джейес Прокси локализация* (JS Proху Localization) е процесът, чрез който цялата комуникация между крайния потребител на браузър и оригиналния сървър на уебсайт преминават през сървър за локализация (т.е. прокси).
- *Локализация* (Localization) е процесът на адаптиране на продукт или съдържание към конкретна държава, област или пазар.
- *Плъгин* (Plugin) е софтуерна добавка, която се инсталира в програма, подобрявайки нейните възможности, като предоставя допълнителни функционалности.
- *Сегмент* (Segment) е единицата, на която се разделят изходният и целевият език в преводна памет.
- *Таб* (tab) е раздел или страница в документ или браузър.
- *Творчески превод* (Transcreation) е процесът, чрез който се адаптира съдържанието от един език на друг, като същевременно се запазва съществуващият стил, регистър и цялостно внушение на текста.
- *Дотнет* (.NET) е патентована софтуерна рамка, разработена от Майкрософт, която работи предимно с Уиндоус.
- *Уебавтоматизиране* (Web automation) е процесът на автоматизиране на действия в интернет като попълване на формуляр, щракване върху бутон или изтегляне на файл.
- *Файлово съдържание* (File content) е информация, съдържаща се във файлове, която е четима от хора.

БИБЛИОГРАФИЯ

1. Али 2017: Aly, El-H. Workplace-Oriented Definition of Translation Competence. – In: Tariq Khan (Eds.) *Translation Today*, Volume 11, Issue 2, pp. 49–75. National Translation Mission, Central Institute for Indian Language, [https://www.ntm.org.in/download/ttvol/volume11-2/E-TT%20Vol_11-2_2017.pdf] (посетен на 20.07.2022).

2. Алшехаб 2017: Alshehab, M. The Effect of Using Mobiles in Students' Ability in Translation from English into Arabic at Jadara University in Jordan, In: – *International Journal of Applied Linguistics and Translation*, Vol. 3, No. 3, pp. 32–39. doi: 10.11648/j.ijalt.20170303.11

[https://www.researchgate.net/publication/321653014_The_Effect_of_Using_Mobiles_in_Students%27_Ability_in_Translation_from_English_into_Arabic_at_Jadara_University_in_Jordan] (посетен на 20.07.2022).

3. Ан 2021: Anh, T. The Difference Effects of Paper Dictionaries vs. Online Dictionaries, In: – *AsiaCALL Online Journal*, Vol. 12, No. 3, pp. 28–38. ISSN 1936-9859 <https://asiacall.info/acoj> (посетен на 20.07.2022).

4. Астраханцев 2016: Astrakhantsev, N. ATR4S: Toolkit with State-of-the-art Automatic Terms Recognition Methods in Scala. [<https://arxiv.org/abs/1611.07804>] (посетен на 20.07.2022).

5. Аренас 2020: Arenas, A. G. Pre-editing and post-editing. – In: E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury Companion to Language Industry Studies* (1 ed., pp. 333–360). (Bloomsbury Companions). Bloomsbury Academic [<https://research.rug.nl/en/publications/pre-editing-and-post-editing>] (посетен на 20.07.2022).

6. Бактиорович 2021: Bakhtiyorovich, S. D. Term: definition of the concept and its essential features. – In: *Central Asian Journal of Theoretical & Applied Sciences*, 2(6), 12–18 [<https://cajotas.centralasianstudies.org/index.php/CAJOTAS/article/view/181>] (посетен на 20.07.2022).

7. Биану-Гил и Пим 2006: Biau-Gil, J. R. and A. Pym. Technology and translation (a pedagogical overview). – In: Anthony Pym, Alexander Perestrenko & Bram Starink (Eds.). *Translation Technology and its Teaching*. Intercultural Studies Group, Universitat Rovira i

Virgili; Tarragona (Spain), pp 5–19. [<https://www.scribd.com/document/251311642/Pym-e-Biau-Gil-Translation-Technology-2006>] (посетен на 20.07.2022).

8. Бейкър 1992: Baker, M. *In other words. A coursebook on translation*, London: Routledge 11 New Fetter Lane. EC4P 4EE.

9. Бейкър 2017: Baker, W. *Controlled vocabularies in the digital age: are they still relevant?*, Dissertation at University of North Texas, US [https://digital.library.unt.edu/ark:/67531/metadc1011802/m2/1/high_res_d/BAKER-DISSERTATION-2017.pdf] (посетен на 20.07.2022).

10. Бендана и Мелби 2012: Bendana, L. and A. Melby. Everything you ever wanted to know about Translation, – In: Multi-Languages Corporation, Toronto, Canada [https://www.academia.edu/34926861/Everything_you_ever_wanted_to_know_about_translation_melby_bendana_1_] (посетен на 20.07.2022).

11. Благоева и др. 2012: Blagoev, D., S. Koeva and V. Murdarov. The Bulgarian Language in the Digital Age, In – Rehm G. and H. Uszkoreit). *White Paper Series*. Springer [<http://www.meta-net.eu/whitepapers/e-book/bulgarian.pdf>] (посетен на 20.07.2022).

12. Боукър 2005: Bowker, L. *Productivity vs Quality? A pilot study on the impact of translation memory systems*. Localisation Reader, pp. 133–140.

13. Боукър 2015: Bowker, L. *Computer-aided translation: translator training*. In S.-W. Chan (Ed.), *The Routledge encyclopedia of translation technology* (pp. 88–104), Routledge.

14. Боукър и Сиро 2019: Bowker L. and J.B. Ciro. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*, 37–54.

15. Боукър 2020: Bowker, L. Terminology. – In: Baker M. and Saldanha G. (Eds) *Routledge Encyclopedia of Translation Studies*, pp. 579–583, London and New York, by Routledge [<https://www.routledge.com/Routledge-Encyclopedia-of-Translation-Studies-3rd-Edition/Baker-Saldanha/p/book/9781138933330>] (посетен на 20.07.2022).

16. Бояджиев 2011: Бояджиев, Т. *Българска лексикология*. Издател УИ „Св. Климент Охридски”.

17. Бояджиев 2012: Бояджиев, Т. Езиковата ситуация у нас в исторически и съвременен план и европейската езикова политика. – В: Жобов, В. и Г. Колев (ред.) *Книжовен език и диалекти. Избрани трудове*. София: Университетско издателство „Св. Климент Охридски”, стр. 41–83.

18. Бриньолфсон и др. 2019: Brynjolfsson, E, X. Hui, M. Liu. Does Machine Translation Affect International Trade? Evidence from a Large Digital Platform. – In:

Management Science, Vol. 65, No. 12. [<https://doi.org/10.1287/mnsc.2019.3388>] (посетен на 20.07.2022).

19. Булте и Тецкан 2019: Bulté, B. and A. Tezcan. Neural Fuzzy Repair: Integrating Fuzzy Matches into Neural Machine Translation. 10.18653/v1/P19-1175, [https://www.researchgate.net/publication/334745562_Neural_Fuzzy_Repair_Integrating_Fuzzy_Matches_into_Neural_Machine_Translation] (посетен на 20.07.2022).

20. Вакуленко 2014: Vakulenko, M. Term and terminology: basic approaches, definitions, and investigation methods (Eastern-European perspective). – In: *Terminology Science & Research*. 24. pp. 13–28, [https://www.researchgate.net/publication/336605560_Term_and_terminology_basic_approaches_definitions_and_investigation_methods_Eastern-European_perspective_In_Terminology_Science_Research] (посетен на 20.07.2022).

21. Вапра 2013: Varga, C. Virtualization of Research in Terminology. Cloud-based Terminology Management Tools. – In: *Transactions on Modern Languages*, Vol. 12, No. 1-2, Scientific Bulletin of the Politehnica University of Timișoara. [https://www.academia.edu/5631484/Virtualization_of_Research_in_Terminology_Cloud_based_Terminology_Management_Tools?email_work_card=title] (посетен на 20.07.2022).

22. Видал и др. 2020: Sergi, A.-V., O. Antoni y T. Badia. Post-editing for Professional Translators: Cheer or Fear?. – In: *Tradumàtica. Tecnologies de la Traducció*, 18, pp. 49–69. [<https://doi.org/10.5565/rev/tradumatica.275>] (посетен на 20.07.2022).

23. Виеира 2017: Vieira, L. Automation anxiety and translators. – In: *Translation Studies*, 13:1, pp. 1–21, Routledge, [<https://www.tandfonline.com/doi/full/10.1080/14781700.2018.1543613>] (посетен на 20.07.2022).

24. Вокалович 2021: Vukalović, N. An Analysis of Computer - Assisted Translation (CAT) Tools. – In: *Sveučilište u Rijeci*, Hrvatska, Filozofski fakultet, Odsjek za anglistiku [<https://repository.ffri.uniri.hr/islandora/object/ffri:2967>] (посетен на 20.07.2022).

25. Гамбие 2019: Gambier, Y. Impact of technology on Translation and Translation Studies, – In: *Russian Journal of Linguistics*, Серия: Лингвистика, Т. 23. No 2. С. 344–361, [https://www.academia.edu/39818451/Impact_of_technology_on_Translation_and_Translation_Studies_1?email_work_card=title] (посетен на 20.07.2022).

26. Гарсия 2015: Garcia, I. Computer-aided translation: systems. – In: In S.-W. Chan (Ed.), *Routledge Encyclopedia of Translation Technology*, University of Western Sydney,

Australia, pp. 68–87. Retrieved from <http://UWSAU.ebib.com.au/patron/FullRecord.aspx?p=1843560> (посетен на 20.07.2022).

27. Гаудек 2007: Gouadec, D. Translation as a profession. University of Rennes, Benjamins translation library, 0929-7316 ; v. 73, 2007 [<https://benjamins.com/catalog/btl.73>] (посетен на 20.07.2022).

28. Гилбърт 2022: Gilbert, D. Solving Terminology Management Pain Points: One Termbase Management System to Rule Them All. – In: *9th Asian-Pacific Forum on Translation and Intercultural Studies*, [https://www.researchgate.net/publication/359586445_Solving_Terminology_Management_Pain_Points_One_Termbase_Management_System_to_Rule_Them_All/citations] (посетен на 20.07.2022).

29. Дара и др. 2013: Dara, A. A., S. Dandapat, D. Groves and J. Van Genabith. TMTprime: a recommender system for MT and TM integration. – In: Proceedings of the 2013 NAACL HLT Demonstration Session (pp. 10–13). [<https://aclanthology.org/N13-3003.pdf>] (посетен на 20.07.2022).

30. Дер Мер 2020: van der Meer, J. Translation technology—past, present and future. – In: E. Angelone, M. Ehrensberger-Dow, & G. Massey (Eds.), *The Bloomsbury Companion to Language Industry Studies* (1 ed., pp. 285–310), Bloomsbury Academic, [<https://www.bloomsburycollections.com/book/the-bloomsbury-companion-to-language-industry-studies/ch13-translation-technology-past-present-and-future>] (посетен на 20.07.2022).

31. Де Калуве и Ван Сантен 2003: De Caluve, J. and A. Van Santen. Phonological, morphological and syntactic specifications in monolingual dictionaries. – In: P. van Sterkenburg (Ed.), *A Practical Guide to Lexicography*, Institute for Dutch Lexicology, Leiden, pp. 71–82. [<https://benjamins.com/catalog/tlrp.6.10dec>] (посетен на 20.07.2022).

32. Делавиние 2017: Delavigne, V. Term usage and socioterminological variation. The impact of social and local issues on the movement of terms. – In: Drouin P., A. Francœur, J. Humbley and A. Picton (Eds.), *Multiple Perspectives on Terminological Variation* Université Sorbonne Nouvelle – Paris 3, Chapter 2, pp. 31–55, John Benjamins Publishing Company. [<https://benjamins.com/catalog/tlrp.18.02del>] (посетен на 20.07.2022).

33. Дохърти 2016: Doherty, S. The Impact of Translation Technologies on the Process and Product of Translation. – In: *International Journal of Communication*, 10, pp. 947–969 [<https://ijoc.org/index.php/ijoc/article/view/3499>] (посетен на 20.07.2022).

34. Друан 2003: Drouin, P. Term extraction using non-technical corpora as a point of leverage. – In: *Terminology*. 9. 10.1075/term.9.1.06dro, [https://www.researchgate.net/publication/228683045_Term_extraction_using_non-technical_corpora_as_a_point_of_leverage] (посетен на 20.07.2022).
35. Дьо Шривер 2003: De Schryver, G.-M. Lexicographers' dreams in the electronic dictionary age. – In: *International Journal of Lexicography*, 16(2): 143–199. [https://academic.oup.com/ijl/article-abstract/16/2/143/925134] (посетен на 20.07.2022).
36. Еригуки и др. 2019: Eriguchi, A., S. Rarrick, H. Matsushita. Combining Translation Memory with Neural Machine Translation. – In: *Proceedings of the 6th Workshop on Asian Translation, Association for Computational Linguistics*, pp. 123–130 Hong Kong, China. [https://www.aclweb.org/anthology/D19-5214] (посетен на 20.07.2022).
37. Занетин 2015: Zanettin, F. Concordancing. – In: S.-W. Chan (Ed.), *Routledge Encyclopedia of Translation Technology*. University of Perugia, Italy, pp. 475–487. [https://www.taylorfrancis.com/chapters/edit/10.4324/9781315749129-40/concordancing-federico-zanettin] (посетен на 20.07.2022).
38. Зарецкая и др. 2015: Zaretskaya, A., G. Corpas Pastor and M. Seghiri. Translators' requirements for translation technologies: a user survey. – In: *New Horizons in Translation and Interpreting Studies*, pp. 247–254, Geneva, Switzerland. Tradulex. [https://www.researchgate.net/publication/282658144_Translators'_Requirements_for_Translation_Technologies_a_User_Survey] (посетен на 20.07.2022).
39. Зарецкая и др. 2015: Zaretskaya, A., G. Corpas Pastor, G and M. Seghiri. Integration of Machine Translation in CAT Tools: State of the Art, Evaluation and User Attitudes. – In: *SKASE Journal of Translation and Interpretation*, vol. 8, no. 1. [https://www.researchgate.net/publication/283667119_Integration_of_Machine_Translation_in_CAT_Tools_State_of_the_Art_Evaluation_and_User_Attitudes] (посетен на 20.07.2022).
40. Зарецкая и др. 2016: Zaretskaya, A., G. Corpas Pastor, G and M. Seghiri and C. Hernani. Nine Terminology Extraction Tools: Are they useful for translators?. – In: *Multilingual*, 27(3), April/May, pp. 14–20. [https://wlv.openrepository.com/handle/2436/622550] (посетен на 20.07.2022).
41. Зарей и Фумани 2017: Zarei, S. and M. Fumani. Impact of Online Versus Hardcopy Dictionaries' Application on Translation Quality of Iranian M. A. Translation Students. [https://www.researchgate.net/publication/324835559_Impact_of_Online_Versus_Hardcopy_

Dictionaries'_Application_on_Translation_Quality_of_Iranian_M_A_Translation_Students] (посетен на 20.07.2022).

42. Зафра 2006: Zafra, J. *The adoption of computer-aided translation tools by freelance translators in the UK*. Loughborough University. Thesis. [<https://hdl.handle.net/2134/7812>] (посетен на 20.07.2022).

43. Зехналова и др. 2013: Zehnalová, J., O. Molnár and M. Kubánek. A Comprehensive Survey of Multilingual Neural Machine Translation. – In: *Tradition and Trends in Trans-Language Communication*, Olomouc, Univerzita Palackého. [<https://arxiv.org/pdf/2001.01115.pdf>] (посетен на 20.07.2022).

44. Зидарова 2018: Зидарова, В. *Процеси на терминологизация и детерминологизация в българската лексикална система*. – В: *Славянски диалози*, КН. 21, Пловдивски университет „Паисий Хилендарски“, [<https://dialozi.uni-plovdiv.bg/wp-content/uploads/2020/10/Slavjanski-dialozi-kn.21-2018-56-66.pdf>] (посетен на 20.07.2022).

45. Имре 2013: Imre, A. Term Bases Reloaded. – In: *Philologia 14*, Studia Universitatis „Petru Maior”. pp. 204–210. [https://www.researchgate.net/publication/287911571_Term_Bases_Reloaded] (посетен на 20.07.2022).

46. Имре 2015: Imre, A., Quality Translations with the Help of Term Bases. – In: *Sapientia Discourse as a Form of Multiculturalism in Literature and Communication*, Section: Language and Discourse Arhipelag XXI Press, University of Tîrgu Mureş, [https://www.researchgate.net/publication/292047910_QUALITY_TRANSLATIONS_WITH_THE_HELP_OF_TERM_BASES] (посетен на 20.07.2022).

47. Имре 2020: Imre, A. An Updated View on Translation Studies. – In: *Journal of Romanian Literary Studies*, Issue No. 8/2016, pp. 149–157. [https://www.researchgate.net/publication/339289785_AN_UPDATED_VIEW_ON_TRANSLATION_STUDIES] (посетен на 20.07.2022).

48. Имре 2021: Imre, A. An Introduction to Translator Studies. Publisher: Editura Universității Transilvania din Braşov, [https://www.researchgate.net/publication/349395585_An_Introduction_to_Translator_Studies] (посетен на 20.07.2022).

49. Имре и Барабас 2020: Imre, A. and B. Barabás. The Efficiency of a Multilingual Term Base. [https://www.researchgate.net/publication/339289691_THE_EFFICIENCY_OF_A_MULTILINGUAL_TERM_BASE] (посетен на 20.07.2022).

50. ИЮН 2018: Uyun, R. M. *Khaled Hosseini's style in the kite runner and the mountains echoed: A study on corpus stylistics*. Thesis, Universitas Airlangga. [<https://repository.unair.ac.id/79103/>] (посетен на 20.07.2022).

51. Кареура и УМИНО 1996: Kageura, K. and B. Umino. Methods of automatic term recognition. – In: *Terminology* vol 3(2), pp. 259–289, John Benjamins Publishing Co. [http://www.iro.umontreal.ca/~felipe/IFT6010-Automne2011/resources/tp3/gabriel_bc/Kageura_Umino_1996.pdf] (посетен на 20.07.2022).

52. Канавос и Картсаклис 2010: Kanavos, P. and D. Kartsaklis. Integrating Machine Translation With Translation Memory: A Practical Approach. – In: Conference: Proceedings of 2nd Joint EM+/CNGL Workshop “*Bringing MT to the User: Research on Integrating MT in the Translation Industry*” (JEC’10) At: Denver, USA [https://www.researchgate.net/publication/269519356_Integrating_Machine_Translation_With_Translation_Memory_A_Practical_Approach] (посетен на 20.07.2022).

53. Кабре 1998: Cabré, M. T. *Terminology: Theory, methods, and applications* (Vol. 1). John Benjamins Publishing.

54. Кабре 2002: Cabré, M. T. Terminología y normalización lingüística. – In: Jornadas (EHU: LEIOA) *Terminología y lenguajes de especialidad*. Euskara Institutua EHU-LEIOAKO CAMPUSA País Basc.

55. Карейра 2020: Carreira, O. Defining Transcreation from the Practitioners Perspective: an Interview-based Study. – In: C. Botella Tejera, J. Franco Aixelá, & C. Iliescu Gheorghiu (Eds.), *Translatum Nostrum. La traducción y la interpretación en el ámbito humanístico*. Comares. [https://www.researchgate.net/profile/Oliver-Carreira/publication/346494704_Defining_Transcreation_from_the_Practitioners_Perspective_An_Interview-Based_Study/links/602076a092851c4ed5578d7a/Defining-Transcreation-from-the-Practitioners-Perspective-An-Interview-Based-Study.pdf] (посетен на 20.07.2022).

56. Кели и Мартин 2020: Kelly, D. and A. Martin. Training and education, curriculum. – In: Baker M. and G. Saldanha (Eds.). *Routledge Encyclopedia of Translation Studies*, pp. 591–596, Third edition, London and New York, by Routledge [<https://www.routledge.com/Routledge-Encyclopedia-of-Translation-Studies-3rd-Edition/Baker-Saldanha/p/book/9781138933330>] (посетен на 20.07.2022).

57. Кени 2012: Kenny, D. The Ethics of Machine Translation. – In: S. Ferner (Ed.) *Proceedings of the NZSTI Annual Conference*, pp. 121–131. Auckland: NZSTI.

58. Килгариф и др. 2014: Kilgarriff, A., M. Jakubíček, V. Kovář, P. Rychlý, V. Suchomel. Finding terms in corpora for many languages with the Sketch Engine. – In: *EACL 2014*, 53. [https://www.researchgate.net/publication/271849661_Finding_Terms_in_Corpora_for_Many_Languages_with_the_Sketch_Engine] (посетен на 20.07.2022).
59. Кирали 2000: Kiraly, D. *A Social Constructivist Approach to Translator Education: Empowerment from theory to practice*. London, Routledge.
60. Кирова 2002: Кирова, Л. Еволюция на българската компютърна терминология и компютърен жаргон. – В: *LiterNet*, № 2 (27), [https://litenet.bg/publish3/lkirova/evolution.htm] (посетен на 20.07.2022).
61. Кирова 2004: Кирова, Л. Компютърната лексика – актуални процеси и тенденции. – В: *LiterNet* № 5 (54), [https://litenet.bg/publish3/lkirova/lex-procesi.htm] (посетен на 20.07.2022).
62. Кис 2005: Kis, Á. Terminusalkotás: a terminológiai helyzet és a terminológiai szerep. Mindent fordítunk, és mindenki fordít” Értékek teremtése és közvetítése a nyelvészetben, 105–112.
63. Кнежевич и др. 2021: Millennials as Dictionary Users: A Study of Dictionary Use Habits of Serbian EFL Students. – In: *SAGE Journals*, Volume: 11 issue: 2. [https://journals.sagepub.com/doi/full/10.1177/21582440211008422] (посетен на 20.07.2022).
64. Кодура 2016: Kodura, M. Dictionary-using skills of translation students. –In: *Spoleczeństwo Edukacja. Język*, 4(3), pp. 235–242. [http://cejsh.icm.edu.pl/cejsh/element/bwmeta1.element.desklight-83b0667e-7fa9-4ce5-8da0-ec9f4f91eea3] (посетен на 20.07.2022).
65. Кобан 2015: Coban, F. Analysis and training of the required abilities and skills in translation in the light of translation models and general theories of translation studies. – In: *7th World Conference on Educational Sciences*, (WCES-2015). University of Sakarya, Department of Translation Studies, Sakarya, 54187, 707–714.
66. Коева 2007: Коева, S. Multi-word term extraction for Bulgarian. – In: *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing Information Extraction and Enabling Technologies - ACL '07*. [https://doi.org/10.3115/1567545.1567556] (посетен на 20.07.2022).
67. Коева 2021: Коева, С. The Bulgarian WordNet: Structure and specific features. – In: *Papers of BAS, Humanities and Social Sciences*, Vol. 8, 2021, No. 1,

[https://www.papersofbas.eu/images/Papers_2021-1/Koeva_optimized.pdf] (посетен на 20.07.2022).

68. Коева, Благоева и Колковска 2010: Коева, S., D. Blagoeva and S. Kolkovska. Bulgarian National Corpus Project. – In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

[https://www.academia.edu/26480996/Bulgarian_National_Corpus_Project] (посетен на 20.07.2022).

69. Коева и Стефанова 2022: Коева, S., V. Stefanova. Report on the Bulgarian Language. – In: *European Language Equality (ELE)* [https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D1_5__Language_Report_Bulgarian_.pdf] (посетен на 20.07.2022).

70. Колина 2020: Colina, S. Quality, translation. – In: Baker M. and G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies*, Third edition by Routledge [<https://www.routledge.com/Routledge-Encyclopedia-of-Translation-Studies-3rd-Edition/Baker-Saldanha/p/book/9781138933330>] (посетен на 20.07.2022).

71. Колковска 2010: Колковска, С. Лексикални неологизми в българския език, възникнали от чужди инициални абривиатури. – В: *Български език*, кн. 4, с. 21–30. [<http://ibl.bas.bg/neolex/wp-content/uploads/2013/Issues/Publications/20%20Sia%20Leksikalni%20neologizmi%20ot%20initsialni%20abreviaturi.pdf>] (посетен на 20.07.2022).

72. Континен и др. 2020: Konttinen K., L. Salmi and M. Koponen. Revision and post-editing competences in translator education. – In: M. Koponen, B. Mossop, S. Isabelle, G. Scocchera (Eds.). *Translation Revision and Post-editing Industry Practices and Cognitive Processes*. London and New York, Published by Routledge [<https://www.routledge.com/Translation-Revision-and-Post-editing-Industry-Practices-and-Cognitive/Koponen-Mossop-Robert-Scocchera/p/book/9781138549715>] (посетен на 20.07.2022).

73. Коста и др. 2016: Hernani, C., A. Zaretskaya, G. Corpas Pastor and M. Seghiri. Nine Terminology Extraction Tools: Are they useful for translators? – In: *Multilingual*, 27(3), April/May, pp. 14–20. [<https://wlv.openrepository.com/handle/2436/622550>] (посетен на 20.07.2022).

74. Кристал 2004: Crystal, D. *The Language Revolution*, by Polity Press Ltd.

75. Кристенсен и др. 2017: Christensen, T., M. Flanagan and A. Schjoldager. Mapping Translation Technology Research in Translation Studies. An Introduction to the Thematic Section. *Hermes*. – In: *Journal of Language and Communication in Business* no 56-2017. [https://tidsskrift.dk/her/article/download/97199/146005] (посетен на 20.07.2022).

76. Лий 2016: Lee, C. Multilingual Resources and practices in Digital Communication. – In: A. Georgakopoulou and T. Spilioti (Eds.) *The Routledge Handbook of Language and Digital Communication*, London and New York Routledge, pp. 118–132.

77. Лефевеер 2009: Lefever, E., L. Macken and V. Hoste. Language-independent bilingual terminology extraction from a multilingual parallel corpus. – In: *Proceedings of the 12th Conference of the European Chapter of the ACL*, pp. 496–504, Athens, Greece [https://aclanthology.org/E09-1057.pdf] (посетен на 20.07.2022).

78. Лукашик 2012: Łukasik, M. Terminological dictionary as a comprehensive cognitive and linguistic tool. – In: *Language in Different Contexts: Research papers = Kalba ir kontekstai*, Volume 5 (1), pp. s.98-108. [https://www.researchgate.net/publication/294836193_Terminological_dictionary_as_a_comprehensive_cognitive_and_linguistic_tool] (посетен на 20.07.2022).

79. Лю 2013: Lew, R. *From paper to electronic dictionaries: Evolving dictionary skills*. Pp. 79–84, Airlangga University Press. [https://www.researchgate.net/publication/265252133_From_paper_to_electronic_dictionaries_Evolving_dictionary_skills] (посетен на 20.07.2022).

80. Лю и Дьо Шривер 2014: Lew, R. and G.-M. de Schryver, Dictionary Users in the Digital Revolution, – In: *International Journal of Lexicography*, Volume 27, Issue 4, pp. 341–359, [https://doi.org/10.1093/ijl/ecu011] (посетен на 20.07.2022).

81. Макушина 2007: Makoushina, J. Translation Quality Assurance Tools: Current State and Future Approaches. – In: *Translating and the Computer*. Tomsk, Russia, [https://aclanthology.org/2007.tc-1.8.pdf] (посетен на 20.07.2022).

82. Матео 2016: Mateo, M. Aligning qualitative and quantitative approaches in professional translation quality assessment. – In: *Revista de investigación e innovación en la clase de idiomas*, 25, pp. 36–44. Publicaciones de la Universidad de Alcalá [https://ebuah.uah.es/dspace/handle/10017/30326?locale-attribute=es] (посетен на 20.07.2022).

83. Мичъл-Шуитевörder 2020: Mitchell-Schuitevoerder, R. *A Project-Based Approach to Translation Technology*, London and New York, Publisher: Routledge.

84. Милър и др. 1990: Miller, G., R. Beckwith, C. Fellbaum, C. D. Gross, and K. Miller. *Introduction to WordNet: An On-line Lexical Database*. [https://www.researchgate.net/publication/31441225_Introduction_to_WordNet_An_On-line_Lexical_Database] (посетен на 20.07.2022).

85. Николова 2018: Николова, Д. Лингвокултурен аспект на съвременната компютърна терминология (върху материал от руски и български език), – В: *Автореферат на дисертационен труд*, Шуменски университет „Епископ Константин Преславски“, [https://www.shu.bg/wp-content/uploads/file-manager-advanced/users/razvitie/as/2018/phd-D.Nikolova/avtoreferat_D_Nikolova.pdf] (посетен на 20.07.2022).

86. Олохам 2020: Olohan, M. „Self-translation. – In: Baker M. and G. Saldanha (Eds.), *Routledge Encyclopedia of Translation Studies*, Third edition by Routledge [https://www.routledge.com/Routledge-Encyclopedia-of-Translation-Studies-3rd-Edition/Baker-Saldanha/p/book/9781138933330] (посетен на 20.07.2022).

87. О’Хаган 2012: O’Hagan, M. Translation as the new game in the digital era. – In: *Translation Spaces*, Volume 1, Issue 1, pp. 123–141. [https://doi.org/10.1075/ts.1.06oha] (посетен на 20.07.2022).

88. П. Нюмарк 1988: Newman, P. *A textbook of translation. Shanghai foreign language education press*. First published by Prentice Hall International.

89. Пим 2003: Pym, A. Redefining Translation Competence in an Electronic Age. In defence of a minimalist approach... – In: *Meta* Volume 48, Number 4, December 2003, pp. 481–497. [https://www.erudit.org/en/journals/meta/1900-v1-n1-meta725/008533ar/abstract/] (посетен на 20.07.2022).

90. Пим 2009: Pym, A. *The moving text: Localization, translation, and distribution*. Amsterdam-Philadelphia, John Benjamins.

91. Пим и Торес-Симон 2021: Pym, A. and E. Torres-Simón. Efectos de la automatización en las competencias básicas del traductor: la traducción automática neuronal. – In: *Ocupaciones y Lenguaje. Indicadores y Análisis De Competencias lingüísticas En El Ámbito Laboral*. Tarragona, Urv. Universidad Rovira I Virgili. [https://www.academia.edu/45109338/Efectos_de_la_automatizaci%C3%B3n_en_las_competencias_b%C3%A1sicas_del_traductor_la_traducci%C3%B3n_autom%C3%A1tica_neuronal] (посетен на 20.07.2022).

92. Попова 2016: Попова, М. За някои видове терминографски параметри. Институт за български език – БАН, София

[<https://ispan.waw.pl/ireteslaw/bitstream/handle/20.500.12528/705/Lexikografia-2016-70-78.pdf?sequence=1&isAllowed=y>] (посетен на 20.07.2022).

93. Попова 2019: Попова, М. Някои въпроси на терминологията във връзка с обучението. Journal: Български език и литература, Issue Year: 61/2019, Issue No: 2, pp. 149–155, [<https://www.cceeol.com/search/article-detail?id=758654>] (посетен на 20.07.2022).

94. Попова 2020: Попова, М. Българският език и неговата терминология в ерата на глобализацията. – В: *Доклади от международната научна конференция на Института за български език*, I том. София, Институт за български език – БАН.

95. Райт 2001: Wright, S. E. Data Categories for Terminology Management. – In: Wright S. E. and G. Budin (Eds), *Handbook of Terminology Management: Volume 2: Application-Oriented Terminology Management*, pp. 552–571, John Benjamins Publishing Company.

96. Райт 2020: Wright, S. E. Standards for the language, translation, and localization industry. – In: (O'Hagan M.) *The Routledge Handbook of Translation and Technology*, pp. 21-44, London and New York, Routledge.

97. Рей и Кели 2010: Ray, R. and Kelly N. *Reaching New Markets through Transcreation. Lowell: Common Sense Advisory*. [<https://www.novilinguists.com/sites/default/files/Common%20Sense%20Advisory%20-%20Reaching%20New%20Markets%20through%20Transcreation.pdf>] (посетен на 20.07.2022).

98. Руоконен и Коскинен 2017: Ruokonen, M. and K. Koskinen. Dancing with technology: translators' narratives on the dance of human and machinic agency in translation work. – In: *The Translator*, Volume 23, Issue 3, pp. 310-323. [<https://www.tandfonline.com/doi/full/10.1080/13556509.2017.1301846>] (посетен на 20.07.2022).

99. Рутьер 2020: Roturier, J. XML for translation technology. – In: O'Hagan M. *The Routledge Handbook of Translation and Technology*, pp. 45 - 60, London and New York, Routledge.

100. Сайер 2001: Sager J. Terminology theory, In: – Baker, M. (Ed.), *Routledge Encyclopedia of Translation Studies*. London and New York, Routledge. pp. 250–261 [https://books.google.bg/books?id=ewBfSBo8rRsC&printsec=frontcover&hl=bg&source=gb_s_ge_summary_r&cad=0#v=onepage&q&f=false] (посетен на 20.07.2022).

101. Сдобников 2018: Sdobnikov, V.V. New trends in translation. – In: *Kazan bulletin of young scientists*, 4, pp. 72-79.

102. Сикора 2018: Sikora, I. Developing terminological competence in a specialized translation classroom-problems, resources, and tools. – In: Maliszewski J. (red.) *Częstochowa: Wydawnictwo Politechniki Częstochowskiej*. [https://www.academia.edu/37422238/Developing_terminological_competence_in_a_specialized_translation_classroom_problems_resources_and_tools] (посетен на 20.07.2022).
103. Сингх, Триулзи и Маги 2021: Singh A., G. Triulzi, L. C. L. Magee. Technological improvement rate predictions for all technologies: Use of patent data and an extended domain description. – In: *Research Policy*. Volume 50, Issue 9, [https://doi.org/10.1016/j.respol.2021.104294] (посетен на 20.07.2022).
104. Син-уи 2015: Sin-wai, C. *Routledge Encyclopedia of Translation Technology*. First published 2015 by Routledge. [https://www.routledge.com/Routledge-Encyclopedia-of-Translation-Technology/Chan/p/book/9780367570439] (посетен на 20.07.2022).
105. Син-уи 2017: Sin-wai, C. *The Future of Translation Technology. Towards a World without Babel*. London and New York, by Routledge [https://www.routledge.com/The-Future-of-Translation-Technology-Towards-a-World-without-Babel/Sin-wai/p/book/9780367409050] (посетен на 20.07.2022).
106. Стайнър 2014: Стайнър, Дж. *След Вавилон. Аспекти на езика и превода*, издателство Изток-Запад.
107. Съогард и др. 2015: Søgaard et al. Inverted indexing for cross-lingual NLP. – In: *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pp. 1713–1722, Beijing, China, July 26-31, Association for Computational Linguistics [https://www.aclweb.org/anthology/P15-1165.pdf] (посетен на 20.07.2022).
108. Тайбех 2008: Tayebeh, M. Translation Memories and the Translator. – In: *International Journal of Translation*. 20, pp. 97-106. [https://www.researchgate.net/publication/271602313_Translation_Memories_and_the_Translator] (посетен на 20.07.2022).
109. Терин и др. 2020: Terryn, A. R., V. Hoste, P. Drouin and E. Lefever. TermEval 2020: Shared Task on Automatic Term Extraction Using the Annotated Corpora for Term Extraction Research (ACTER) Dataset. – In: *Proceedings of the 6th International Workshop on Computational Terminology (COMPUTERM 2020)*, pp. 85–94 Language Resources and Evaluation Conference (LREC 2020) [https://biblio.ugent.be/publication/8661811/file/8661815.pdf] (посетен на 20.07.2022).

110. Телен 2015: Thelen, M. The Interaction between Terminology and Translation. – In: Seite 377, *trans-kom* 8 [2]: pp. 347-381, [http://www.trans-kom.eu/bd08nr02/trans-kom_08_02_03_Thelen_Terminology.20151211.pdf] (посетен на 20.07.2022).

111. Тимко 2021: Timko, N. Localization, Transcreation, Transadaptation, Transculturation: New Types of Translation or Trendy Names?, Department of Moscow State Institute of International Relations. [<https://ssrn.com/abstract=3794808>] (посетен на 20.07.2022).

112. Тодоранова и Тодоранова 2017: Тодоранова, Л, А. Тодоранова. Некоректно предадена компютърна терминология и стратегии за преодоляване на нееквивалентността. – В: *Известия на съюза на учените – Варна* 2, pp. 209–2016, [<https://www.su-varna.org/izdaniy/2017/ikonomika-017-2/209-216.pdf>] (посетен на 20.07.2022).

113. Торес-Хостенч и др. 2017: Testing interaction with a mobile MT post-editing app. – In: *The International Journal of Translation and Interpreting Research*, 9(2), pp. 138–150. [<https://search.informit.org/doi/10.3316/informit.986278878398978>] (посетен на 20.07.2022).

114. Тофтедал 2020: Toftedahl, M. Localization and regional aspects of game production – a research overview. – In: *Sweden International Conferences Interfaces and Human Computer Interaction 2020 and Game and Entertainment Technologies*. pp. 135–143, [https://www.ihci-conf.org/wp-content/uploads/2020/07/01_202010L017_F034.pdf] (посетен на 20.07.2022).

115. Тръмбъл и Стивънсън 2002: Trumble, W. R., & Stevenson, A. (Eds.), *Shorter Oxford English Dictionary*. Fifth Edition, North Carolina, publisher: Oxford University Press.

116. Уанг и Пинг 2020: Wang, B. and Ping Y. Perceptions of Machine Translation and Computer-Aided Translation by Professionals and the General Public: A Survey Study Based on Articles in Professional Journals and in the Media. – In: *International Journal of Translation, Interpretation, and Applied Linguistics*, Volume 2, Issue 2, pp. 1–14, [https://www.researchgate.net/publication/342605061_Perceptions_of_Machine_Translation_and_Computer-Aided_Translation_by_Professionals_and_the_General_Public_A_Survey_Study_Based_on_Articles_in_Professional_Journals_and_in_the_Media/citations] (посетен на 20.07.2022).

117. Уорбъртън 2015: Warburton, K. Terminology management. In Sin-wai, C. (Ed), *The Routledge Encyclopedia of Translation Technology*, pp. 644–661, New York, Routledge,

[<https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781315749129&type=googlepdf>] (посетен на 20.07.2022).

118. Уорбъртън 2021: Warburton, K. *The Corporate Terminologist*. University of Illinois at Urbana-Champaign, John Benjamins.

[<https://www.jbe-platform.com/content/books/9789027260093>] (посетен на 20.07.2022).

119. Фабер 2011: Faber, P. The dynamics of specialized knowledge representation: Simulational reconstruction or the perception–action interface. – In: *Terminology* 17:1, pp. 9–29, John Benjamins Publishing Company,

[https://www.researchgate.net/publication/263063290_The_dynamics_of_specialized_knowledge_representation_Simulational_reconstruction_or_the_perception-action_interface] (посетен на 20.07.2022).

120. Фабер и Арауз 2021: Faber, P. and Araúz P. Designing Terminology Resources for Environmental Translation. – In: Meng, Ji. and S. Laviosa (Eds.), *The Oxford Handbook of Translation and Social Practices*, pp. 587–615, Oxford Academic
[https://www.researchgate.net/publication/350213272_Designing_Terminology_Resources_for_Environmental_Translation] (посетен на 20.07.2022).

121. Хансен 2010: Hansen, G. Translation ‘errors’. - In: Y. Gamber and Van Doorslaer L. (Eds.) *Handbook of Translation Studies. Volume I*. Amsterdam: John Benjamins, pp. 383–388.

[https://books.google.bg/books?hl=bg&lr=&id=BTwzAAAAQBAJ&oi=fnd&pg=PA385&dq=translation+error+classification&ots=bhrc4Gmjv2&sig=3l5Dpv8GDPsvv1SXo9Z8pp3ioxs&redir_esc=y#v=onepage&q=translation%20error%20classification&f=false] (посетен на 20.07.2022).

122. Хаус 2015: House, J. *Translation Quality Assessment. Past and Present*. London, Routledge.

123. Хейлън и Хертъг 2015: Heylen, K. and D. Hertog. Automatic Term Extraction. – In: H. J. Kockaert and F. Steurs (Eds.). *Handbook of Terminology: Volume 1*, pp. 203–221, [<https://doi.org/10.1075/hot.1.aut1>] (посетен на 20.07.2022).

124. Хойзъл и др. 2015: Hoisl, K., T. Stelzer, S. Biala. Forecasting technological discontinuities in the ICT industry. – In: *Res. Policy*, 44, pp. 522–532, [<https://www.sciencedirect.com/science/article/abs/pii/S0048733314001796>] (посетен на 20.07.2022).

125. Хосари и др. 2019: Hosarri, M., Dev S., Kelleher J. D. TEST: A Terminology Extraction System for Technology Related Terms. – In: *ICCAE* pp. 23-25, Perth, Australia [https://arxiv.org/pdf/1812.09541.pdf] (посетен на 20.07.2022).

126. Хьониг 1997: Hönig, H.G. Positions, Power and Practice: Functionalist approaches and translation quality assessment. – In: *Current Issues in Language and Society* 4(1): pp. 6–34. [https://www.tandfonline.com/doi/abs/10.1080/13520529709615477] (посетен на 20.07.2022).

127. Чарчунеско и др. 2004: Craciunescu, O., C. Gerding-Salas, S. Stringer-O'Keeffe. Machine Translation and Computer-Assisted Translation: a New Way of Translating?. – In: *Translation Jurnal Translators and Computers*, Vol 8, No 3, [https://aclanthology.org/www.mt-archive.info/TranslationJ-2004-Craciunescu.pdf] (посетен на 20.07.2022).

128. Четинер 2021: Çetiner, C. Evaluating Quality Assurance (QA) Checks in CAT tools in Relation to Professional Quality Standards. – In: *12th Eurasian Conferences on Language & Social Sciences*, pp 338–349, Ekaterinburg, Russia. [https://www.researchgate.net/publication/353719751_Evaluating_Quality_Assurance_QA_Checks_in_CAT_tools_in_Relation_to_Professional_Quality_Standards] (посетен на 20.07.2022).

129. Чуню и Уонг Так-минг 2015: Chunyu, K. and Wong Tak-ming B. Evaluation in Machine Translation and Computer-Aided Translation. – In: Chan Sin-wai (Ed.) *The Routledge Encyclopedia of Translation Technology*, pp. 213–236, New York, Routledge. [https://www.routledgehandbooks.com/doi/10.4324/9781315749129.ch12] (посетен на 20.07.2022).

130. Шаятович и др. 2019: Šajatović, A., M. Buljan, J. Šnajder and B. Bašić. Evaluating Automatic Term Extraction Methods on Individual Documents. – In: *Proceedings of the Joint Workshop on Multiword Expressions and WordNet*, pp. 149–154, Florence, Italy. Association for Computational Linguistics. [https://aclanthology.org/W19-5118/] (посетен на 20.07.2022).

131. Шилдс 1999: Shields, M. Slaves to the Computer. – In: *Institute of Translation and Interpreting Bulletin*, October 1999, pp. 4-5.

132. Шрев 2020: Shreve, G.M. Professional translator development from an expertise perspective. – In: E. Angelone, M. Ehrensberger-Dow and G. Massey (Eds.) *The Bloomsbury Companion to Language Industry Studies*, (1 ed., pp. 153–177). (Bloomsbury Companions). Bloomsbury Academic.

[https://books.google.bg/books?hl=bg&lr=&id=jGawDwAAQBAJ&oi=fnd&pg=PA153&dq=Professional+translator+development+from+an+expertise+perspective&ots=efhXW-w9HQ&sig=04reUDLEzCVvFEgVtlMy06JTU4&redir_esc=y#v=onepage&q=Professional%20translator%20development%20from%20an%20expertise%20perspective&f=false] (посетен на 20.07.2022).

133. Швитер 2015: Schwitter, R. Controlled Language. – In: Chan Sin-wai (Ed.) *The Routledge Encyclopedia of Translation Technology*, pp. 450–464, Macquarie University, Australia.

134. ЯНГ 2021: Yang, Z. Effective Computer-assisted Terminology Management Through SDL MultiTerm. – In: *Journal of Physics: Conference Series 1861*, Xi'an Fanyi University, Xi'an, Shaanxi, China, [<https://iopscience.iop.org/article/10.1088/1742-6596/1861/1/012106/pdf>] (посетен на 20.07.2022).

Източници от интернет:

Байер 2012: Bier, K. MT journey: MT use at SYBASE, a SAP company. – In: *TAUS Open Source Machine Translation Showcase*. Sybase, 4 June 2012 <<https://www.slideshare.net/TAUS/4-june-2012-taus-moses-open-source-mt-showcase-paris-kerstin-bier-sybase>> (посетен на 20.07.2022).

What is Machine Translation Literacy?

<<https://sites.google.com/view/machinetranslationliteracy/home/resources>> (посетен на 20.07.2022).

The Shallowness of Google Translate

<<https://www.theatlantic.com/technology/archive/2018/01/the-shallowness-of-google-translate/551570/>> (посетен на 20.07.2022).

Атлас на лингвистичните технологии Nimdzi 2019 <<https://www.nimdzi.com/language-technology-atlas-2019/>> (посетен на 20.07.2022).

The 2021 Nimdzi 100: The Ranking of Top 100 Largest Language Service Providers <https://www.nimdzi.com/nimdzi-100-top-lsp/?utm_source=ZohoCampaigns&utm_campaign=FULL+REPORT%3A+State+of+the+language+industry%2C+data+is+now+live%21+&utm_medium=email>

(посетен на 20.07.2022).

The Nimdzi Language Technology Atlas 2020 - [<https://www.nimdzi.com/nimdzi-language-technology-atlas-2020/>] (посетен на 20.07.2022)

Automated TM apps for Trados <https://appstore.sdl.com/language/list/automated-translation/last-updated/?utm_medium=organic-social&utm_source=linkedin&utm_campaign=sdl-trados-studio&utm_tactic=tp-2021-02-19-image&utm_content=ap-last-updated> (посетен на 20.07.2022).

LISA QA Metric
<http://producthelp.sdl.com/SDL_TMS_2011/en/Creating_and_Maintaining_Organizations/Managing_QA_Models/LISA_QA_Model.htm> (посетен на 20.07.2022).

The 8 most used standards and metrics for Translation Quality Evaluation
<<https://blog.taus.net/the-8-most-used-standards-and-metrics-for-translation-quality-evaluatio>> (посетен на 20.07.2022).

Seven Machine Translation Trends in 2020 - TAUS
<https://blog.taus.net/seven-machine-translation-trends-in-2020?utm_content=128026586&utm_medium=social&utm_source=linkedin&hss_channel=1cp-894145> (посетен на 20.07.2022).

Report on UK Translator Survey 2016-17 <<https://www.ciol.org.uk/report-uk-translator-survey-2016-17>> (посетен на 20.07.2022).

UK Translator Survey 2017 <<http://www.ciol.org.uk/sites/default/files/UKTS2016-Final-Report-Web.pdf>> (посетен на 20.07.2022).

What protects minority languages from extinction?, ScienceDaily. ScienceDaily, 22 April 2020. <www.sciencedaily.com/releases/2020/04/200422112301.htm> (посетен на 20.07.2022).

Localization Industry Standards Association
<https://en.wikipedia.org/wiki/Localization_Industry_Standards_Association> (посетен на 20.07.2022).

Quality Translation 21, D3.1: Harmonised Metric <<http://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf>> (посетен на 20.07.2022)

What is Translation Quality Assurance? <<https://www.memoq.com/tools/translation-quality-assurance>> (посетен на 20.07.2022)

Machine Translation Report Q2 2021 <<https://go.memsource.com/machine-translation-report>> (посетен на 20.07.2022).

Quality Translation 21 Harmonised Metric. part of the Research and Innovation Action. This project has received funding from the European Union's Horizon 2020 program for ICT under grant agreement no. 645452 <<http://www.qt21.eu/wp-content/uploads/2015/11/QT21-D3-1.pdf>> (посетен на 20.07.2022).