

БЪЛГАРСКА АКАДЕМИЯ НА НАУКИТЕ  
ИЗВЕСТИЯ НА ИНСТИТУТА ЗА БЪЛГАРСКИ ЕЗИК „ПРОФ. ЛЮБОМИР АНДРЕЙЧИН“  
Книга XXXVI

## ПОЛУАВТОМАТИЧНО РАЗРАБОТВАНЕ НА МНОГОЕЗИКОВИ ТЕРМИНОЛОГИЧНИ РЕСУРСИ\*

**Виктория Петрова-Любенова**  
Институт за български език

## DEVELOPMENT OF SEMI-AUTOMATIC MULTILINGUAL TERMINOLOGICAL RESOURCES

**Viktoriya Petrova-Lyubanova**  
Institute for Bulgarian Language

**Резюме:** Текстът предлага методология за полуавтоматично създаване на терминологични ресурси за български език за системите за компютърноподпомогнат превод. За целта са описани и анализирани техническите характеристики на системите за компютърноподпомогнат превод и техните компоненти (преводна памет, терминологична база, машинен превод). Описани са резултатите от проучване, насочено към преводачите в България и тяхното мнение относно този вид технологии. Предложената методология няма претенции за изчерпателност, но взема под внимание процесите, към които преводачите се придържат в своята работа. Чрез предложената методология е създадена още и английско-българска терминологична база в областта на компютърната терминология.

**Abstract:** The text suggests a methodology for semi-automatic creation of terminological resources for the Bulgarian language for computer-aided translation systems. For this purpose, the technical characteristics of computer-assisted translation systems and their components (translation memory, terminological base, machine translation) are described and analyzed. The results of a survey aimed at translators in Bulgaria and their opinion regarding this type of technology are described. The proposed methodology does not claim to be exhaustive, but takes into account the processes that translators follow in their work. Through the proposed methodology, an English-Bulgarian terminological base in the field of computer terminology was also created.

**Keywords:** computer-aided translation, translation memory, termbase, automatic term extraction, translator

---

\* Copyright of each paper stays with the respective authors. The works in the Papers of the Institute for Bulgarian Language are licensed under a Creative Commons Attribution 4.0 International Licence (CC BY 4.0): <http://creativecommons.org/licenses/by/4.0>.

## 1. Системи за компютърноподпомогнат превод и техните 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. Променят се също и сроковете за изпълнение на възложените преводи и методите на работа. Така технологиите продължават да се развиват и през 2015 г. са изброени шестдесет и седем налични системи на пазара (Гарсия/ García 2015: 22). През 2020 г. бройката се покачва на 164<sup>1</sup>, като вече не се вземат предвид системите за компютърноподпомогнат превод в 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 компютърноподпомогнат превод дават следните дефиниции за този вид програми: „Инструментите за компютърноподпомогнат превод са софтуерни приложения, които помагат за превода на текстове от един език на друг. По-конкретно, инструментите за компютърноподпомогнат превод се използват за възлагане, редактиране, управление и съхранение на преводи“ според Мемсорс<sup>2</sup>; „Инструментите за компютърноподпомогнат 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 мемоКю<sup>3</sup>; „Това е софтуер, използван от преводачи и лингвисти. Има различно предназначение, но преди всичко подпомага процеса на превода. Позволява редакция, управление и съхраняване на превода“ според Традос<sup>4</sup>. Изображенията по-долу илюстрират интерфейса на няколко такива програми. Най-голямо пространство на екрана заема текстовият редактор.

---

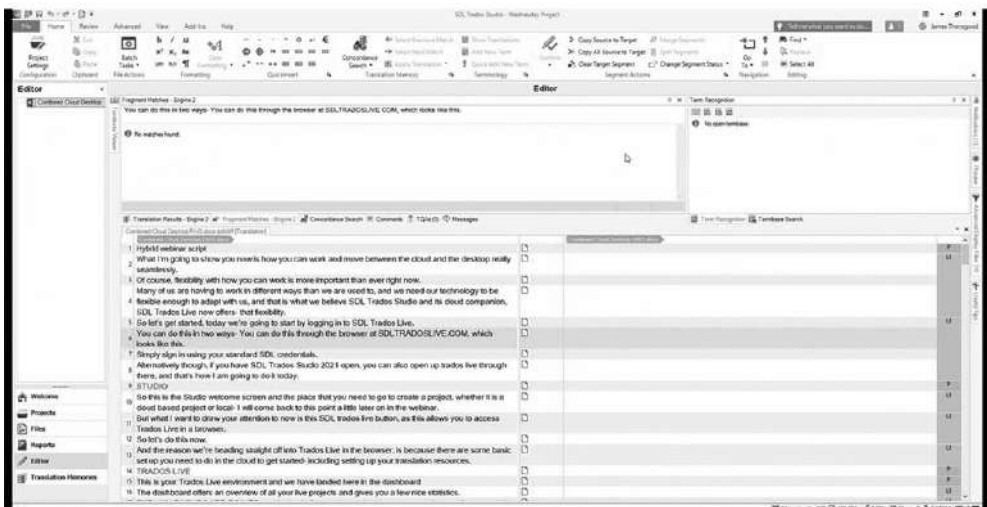
<sup>1</sup> Според годишния доклад на Нимзи за 2020 г. <https://www.nimdzi.com/nimdzi-language-technology-atlas-2020/?hilite=trados>

<sup>2</sup> <https://www.memsource.com/what-are-cat-tools/>

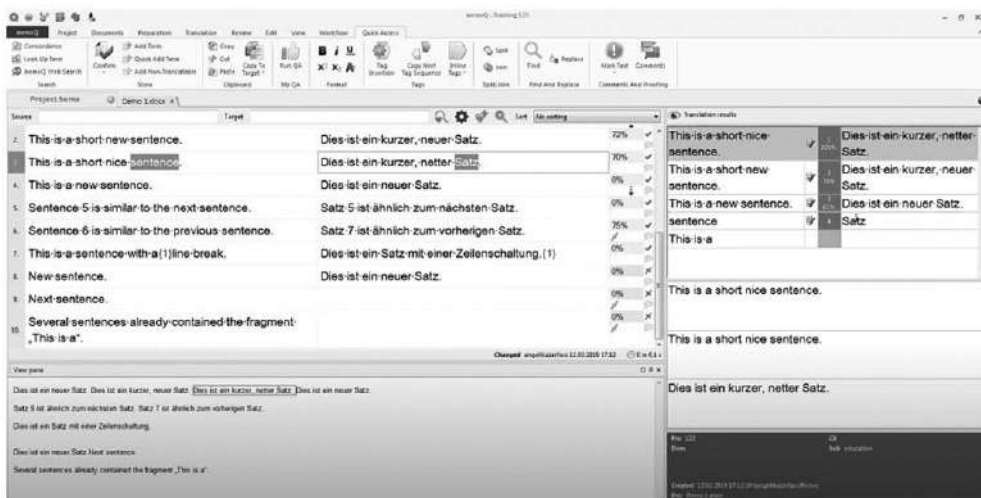
<sup>3</sup> <https://www.memoq.com/tools/what-is-a-cat-tool>

<sup>4</sup> <https://www.trados.com/solutions/cat-tools/translation-101-what-is-a-cat-tool.html>

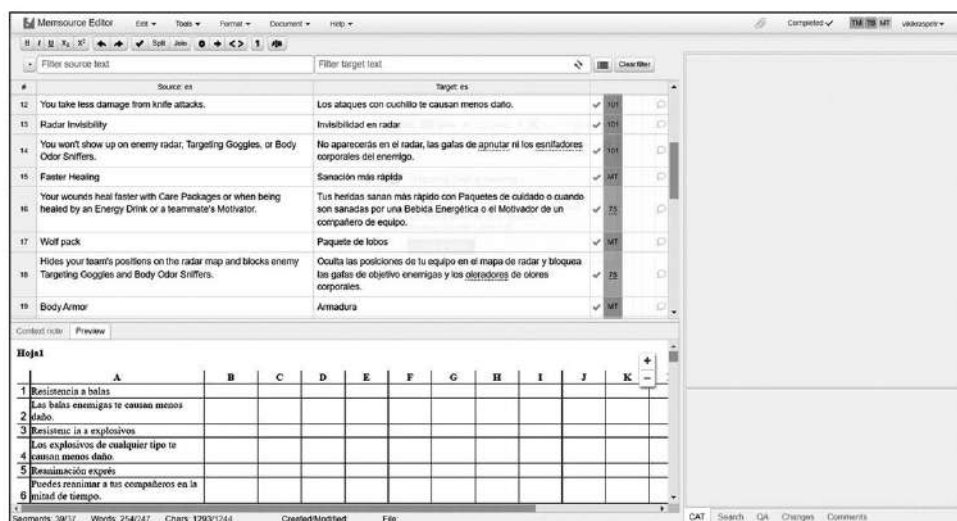
Той е разделен на две части: от лявата страна е изходният текст, разделен на отделни сегменти, а от дясната – целевият текст. Към всеки целеви сегмент са посочени процентите на съвпадение с наличната информация в преводната памет. Както се вижда от Фигура 2, сегментът върху който се работи, е ограден в рамка. При кликане 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.



Фиг. 1. Интерфейс на Традос 2021



Фиг. 2. Интерфейс на мемоКю 2021



Фиг. 3. Интерфейс на Мемсорс

Въпреки различията във функционалностите, доколкото ни е известно, всички системи за компютърнопомогнат 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.

### 1.1. 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 компютърнопомогнат 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/ Bulté and Tezcan 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, който точно да 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 превеждания 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 Левенщайн<sup>5</sup>, 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 Градос 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 съвпадение<sup>6</sup>. 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/Garcia 2015: 71-72).

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 Юролекс<sup>7</sup> и Юропарл<sup>8</sup>

---

<sup>5</sup> 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.

<sup>6</sup> 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).

<sup>7</sup> 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>

<sup>8</sup> Европар (Europar) – портал с актуална информация относно работата на Парламента на ЕС: <https://www.europarl.europa.eu/portal/bg>

или Аки Комюотер<sup>9</sup> (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/Zanettin 2015: 445). Преводната памет се нуждае от три различни вида софтуер, за да функционира правилно:

- система за подравняване на сегменти (*aligner*), посредством която се създават файловете за преводна памет. Може да бъде както самостоятелна програма, така и вградена функция в системите за компютърноподпомогнат 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 компютърноподпомогнат 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 различни преводни паметни, изтриване, експортиране на части от съдържанието, както и историята на всеки сегмент, намиращ се в паметта (Гуадек/Gouadec 2007: 271 – 272).

## 1.2. Машинен превод

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 (Уд<sup>10</sup>). Според статия на Д. Адауси от август 2021 г., „Гугъл транслейт се използва за превод на над сто милиарда думи на ден“<sup>11</sup>. 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 Торрес-Симон/Pym and Torres-Simón 2021: 479).

Константин Дранч<sup>12</sup> 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

---

<sup>9</sup> Аки Комюотер (*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>

<sup>10</sup> Twenty years of building for everyone. Company Announcements: <https://www.blog.google/inside-google/company-announcements/twenty-years-building-everyone/>

<sup>11</sup> 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/#>

<sup>12</sup> Журналист, преводач и съосновател на Custom.MT, специализирана в създаването на персонализирани услуги за невронен машинен превод. <https://custom.mt/>

своите продажби. Константин Дранч уточнява, че делът на машинния превод „е поне два пъти по-голям от системите за преводни памет и расте по-бързо“.

Необходимо уточнение е техническата разлика 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 (Боукър и Сиро/Bowker and Ciro 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.

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/Kanavos and Kartsaklis 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. Сайбейс Айти<sup>13</sup> отчита увеличаване на производителността след такова комбиниране, при условие обаче, че машинният превод е обучен със специфични данни (Байер/Bier 2012).

През 2013 г. група учени предлагат системата Тиемти Прайм<sup>14</sup>, която има за цел да улесни интегрирането на машинния превод и преводната памет (Дара

---

<sup>13</sup> Днешното име на компанията е Sap – <https://www.sap.com/index.html>

<sup>14</sup> <http://www.meta-net.eu/events/german-lt-day-2013/posters/cngl.pdf>

и др./Dara et al. 2013: 10). Опити в тази посока са правени и в комбинация със статистически машинен превод и след това с невронен (Еригуки и др./Eriguchi et al. 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 компаниите и са достъпни само като част от разпространението на продукта) и ограничения във възможностите за промяна (Дер Мер/der Meer 2020: 293).

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. 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“<sup>15</sup> (Аренас/Arenas 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

---

<sup>15</sup> Важно е да се подчертае, че критериите за качество се определят заедно с възложителя на проекта преди започването на работата.



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.

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 версии<sup>16</sup> на мемоКю и Традос позволяват интеграция с машинен превод чрез така наречения приложно-програмен интерфейс (Application Programming Interface или API). Модерн МТ<sup>17</sup>, Амазон Транслейт<sup>18</sup> и ДийпЕл<sup>19</sup> предлагат подобна интеграция.

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/Wang and Ping 2020: 3).

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/Shields 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/Zafra 2006)<sup>20</sup>. С времето всъщност не мал-

---

<sup>16</sup> Тази информацията е валидна към 1 октомври 2021 г. [https://marcinbas.github.io/api\\_mt/](https://marcinbas.github.io/api_mt/)

<sup>17</sup> Модерн МТ <https://www.modernmt.com/>

<sup>18</sup> Амазон Транслейт (AWS) <https://aws.amazon.com/translate/>

<sup>19</sup> Дийпел <https://www.deepl.com/translator>

<sup>20</sup> 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).

ко 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. Ч. Син-уи констатира, че: „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/Sin-wai 2017: 258).

Сравнително добрите резултати на невронния машинен превод отново пораждаат 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)<sup>21</sup> и Мултилингвал (MultiLingual)<sup>22</sup> 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/Wang and Ping 2020: 11). Въпреки това спорът дали машинният превод и инструментите за компютърнопомогнат 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 полезрението на много по-големия контекст на изкуствения интелект, най-вече заради невронния машинен превод“ (доклад на ЕсДиЕл<sup>23</sup> 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. През 2020 г. учени от Университета Помпеу Фабра разглеждат

---

<sup>21</sup> Инститют фор транслейшън енд интерпретинг (Institute for Translation and Interpreting) <https://www.iti.org.uk/>

<sup>22</sup> <https://multilingual.com/>

<sup>23</sup> <https://www.sdl.com/software-and-services/translation-software/research/>

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./Vidal et al. 2020: 65).

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 (Олохан/Olohan 2020: 578).

### **1.3. 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 компютърнопомогнат 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“. 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.

### **1.4 Приложение на многоезикови терминологични бази в системите за компютърнопомогнат превод**

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 въз-

можни 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/Imre and Barabás 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.

### 1.5. 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 Assistant<sup>24</sup> – функция с готови регулярни изрази, която спестява на преводачите нуждата сами да програмират), Антидот (Antidote<sup>25</sup> – интеграция на автоматичния редактор за английски и френски Антидот коректор), Адобе Експириънс Мениджър (Adobe Experience Manager<sup>26</sup> или АЕМ – цялостно решение за управление на съдържанието за изграждане на уебсайтове, мобилни приложения и формуляри), Джира (Jira<sup>27</sup> – един от най-популярните софтуери за проследяване на работата по даден проект), БиЛейзи (BeLazy<sup>28</sup> – платформа, която автоматизира получаването, управлението и доставката на проекти за превод), Гридли (Gridly<sup>29</sup> – позволява по-бърза актуализации и локализация на съдържанието на множество езици) и WS API<sup>30</sup> за интеграция с всяка платформа. 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.

---

<sup>24</sup> <https://docs.memoq.com/ggl-tst/Places/regex-assistant.html>

<sup>25</sup> <https://docs.memoq.com/ggl-tst/Places/antidote-corrector.html>

<sup>26</sup> 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>

<sup>27</sup> <https://www.memoq.com/integrations/project-management-systems/jira-integration>

<sup>28</sup> Интеграцията позволява извличането на проекти от сървъра на мемоКю, както и от тези на потребителите: <https://www.memoq.com/integrations/integration-with-belazy>

<sup>29</sup> <https://www.memoq.com/integrations/content-management/gridly>

<sup>30</sup> <https://docs.memoq.com/current/api-docs/wsapi/memoqservices/index.html>

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 (Clockify<sup>31</sup>) – приложение за проследяване на свободното време и на работните часове в проекти; Протемос (Protemos<sup>32</sup>) – безплатна система за управление на преводи, клиенти, проекти и финанси; Гугъл таскс (Google tasks<sup>33</sup>) – безплатна 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 извлича 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“<sup>34</sup>.

## **1.6. 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 преведени с тяхна помощ. 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 документите със стандарти-

---

<sup>31</sup> <https://clockify.me/>

<sup>32</sup> <https://protemos.com/>

<sup>33</sup> <https://apps.apple.com/us/app/google-tasks-get-things-done/id1353634006>

<sup>34</sup> 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-9wSWGqdfdyXJBqBGwEnLFN5otgQoaeotzTZ441ckycBGhCtCJQIWVlINZGiZH4n7gRNHF7EAF0-SJrapl20N0X7sQT\\_A](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-9wSWGqdfdyXJBqBGwEnLFN5otgQoaeotzTZ441ckycBGhCtCJQIWVlINZGiZH4n7gRNHF7EAF0-SJrapl20N0X7sQT_A)

зирано и еднотипно съдържание (каквито са административните, правните, техническите и др.) и е по-малко вероятно да се наблюдава между два различни литературни текста. 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. Паралелно с традиционната форма на превод („предаване на съдържанието на устен или писмен текст от един на друг език“<sup>35</sup>) се развиват и други процеси, като локализация, интернационализация и творческият превод, отличаващи се с адаптиране на 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. 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/Toftedahl 2020: 135). 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/Ray and Kelly 2010). Характеристики на този тип текстове са: творческата адаптация на рекламно съдържание, промяната в оригиналните термини и значения (което има за цел въздействието върху целевата аудитория) и посланието на търговската марка (Карейра/Carreira 2020: 4).

Въпреки наличието на подобни термини има учени (Сдобников/Sdobnikov 2018; 72; Тимко/Timko 2021:1), които не са съгласни с употребата им. Според тях, технологиите, използвани при превод, не променят същността на познавателната дейност на преводача и локализацията, творческият 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/Timko 2021: 2).

---

<sup>35</sup> <https://ibl.bas.bg/rbe/lang/bg/превод/>

## **2. Отношението на преводачите към технологиите за компютърноподпомогнат превод**

### **2.1. Изисквания на системите за компютърноподпомогнат превод към уменията на преводачите**

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.). 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/Sin-wai 2015: 32).

### **2.2. Профилът на преводача в съвременната епоха**

Д. Гуадек 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 (Гуадек/Gouadec 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/Gambier 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.

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 компютърноподпомогнат превод (Гуадек/Gouadec 2007: 280).

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 да ускорят, улеснят и оптимизират процеса на работа.

### **2.3. Нови професии, създадени заради технологиите за компютърноподпомогнат 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 (Гуадек/Gouadec 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“ (subskills). В случай че някое от тях липсва, неизбежна е появата на проблеми в процеса на превод. Преводаческата компетентност е свързана с нивото на професио-



нализъм 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/Аly 2017: 50).

Обхватът от подходи, 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“<sup>36</sup> 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 владене 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/Imre 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 (Гуадек/Gouadec 2007: 104). Поради огромния брой от необходими знания, умения и качества преводачите всъщност са експерти и в области като комуникация, информационни технологии и др.

Езиковите познания са от съществено значение, но в много случаи са недостатъчни в днешно време. Техническите, юридическите, счетоводните и мар-

---

<sup>36</sup> [https://commission.europa.eu/system/files/2018-02/emt\\_competence\\_fw\\_k\\_2017\\_en\\_web.pdf](https://commission.europa.eu/system/files/2018-02/emt_competence_fw_k_2017_en_web.pdf)

кетинговите познания, съчетани с пълно разбиране на тематичната област, в рамките на която се превежда, също може да са наложителни.

Както се прогнозира от Д. Гуадек: 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 (Гуадек/Gouadec 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.

#### **2.4. 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. Проучването<sup>37</sup>, проведено през 2017 г. съвместно от Представителството на Европейската комисия в Обединеното кралство, Лицензиращия 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 (Олохан/Olohan 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

---

<sup>37</sup> <https://www.ciol.org.uk/report-uk-translator-survey-2016-17>  
<http://www.ciol.org.uk/sites/default/files/UKTS2016-Final-Report-Web.pdf>

също са възпиращ фактор тяхната употреба (Кодура/Kodura 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 издания по отношение на прецизност и плавност на превода. Разликата е значителна само по отношение на скоростта (Зарей и Фумани/Zarei and Fumani 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/Alshehab 2017).

Според проучване на американската компания Сиесей Рисърч<sup>38</sup> 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 Лилт<sup>39</sup> (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 анкетиранияте 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 проучването на Сиесей Рисърч<sup>40</sup> от 2020 г. 66% от 7 363 преводачи

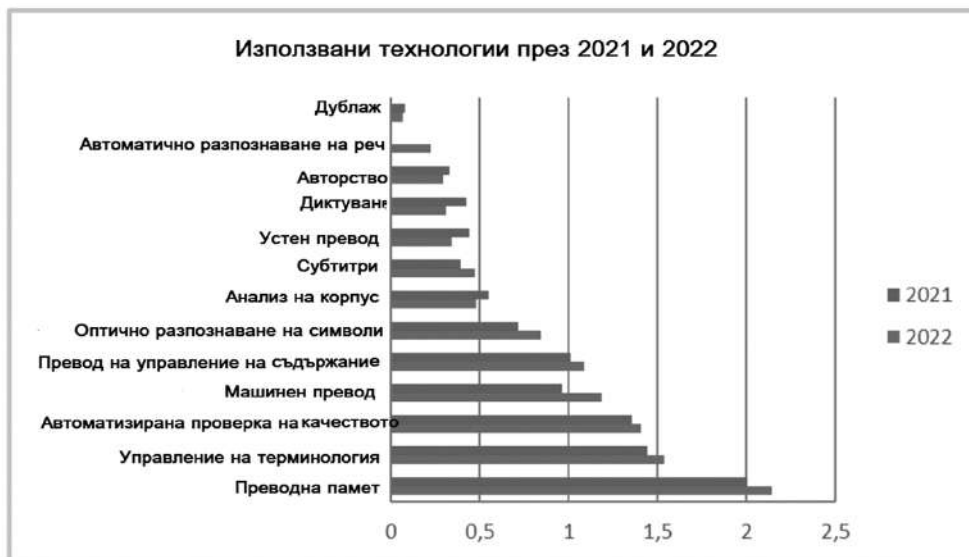
---

<sup>38</sup> Сиесей Рисърч (CSA Research) – независима американска компания за маркетингови проучвания. <https://csa-research.com/>

<sup>39</sup> Lilt <https://lilt.com/>

<sup>40</sup> <https://csa-research.com/>

използват преводна памет и инструменти за 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. Всичко казано по-горе се потвърждава от годишния доклад *Проучване на европейската езикова индустрия* (European Language Industry Survey или ELIS<sup>41</sup>) 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, възлагачи 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 закупуването на технологии, които подпомагат превода, с изключение на машинния превод (11%).



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

<sup>41</sup> <https://elis-survey.org/>

## 2.5. 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“<sup>42</sup>, 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.com<sup>43</sup>, 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 свободно<sup>44</sup>. Всички 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 подобрят.

Анкетата беше проведена в периода 01.02 – 31.03.2022 г. и в нея участваха общо 73 професионални преводачи.

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 приеме 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.

---

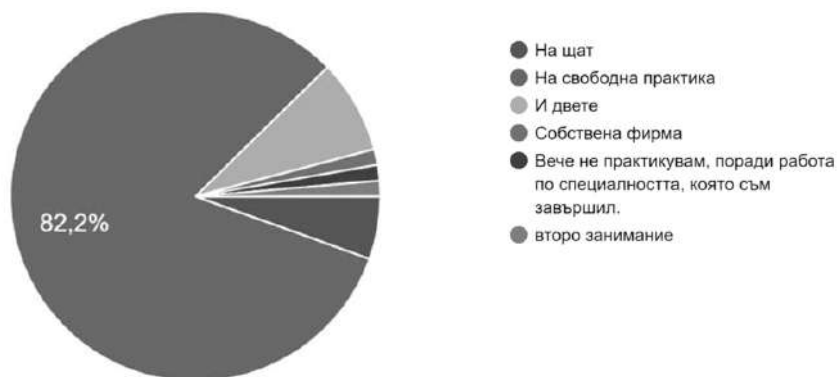
<sup>42</sup> <https://bg-bg.facebook.com/groups/pomosht.pri.prevod/>

<sup>43</sup> <https://www.proz.com/>

<sup>44</sup> <https://docs.google.com/forms/d/1YkIXezcuBVfEId9o7DqP6OrdPqFtZ6nzxZZkSRwMpPw/edit#responses>

## Каква е Вашата професионална заетост?

73 отговора



Фиг. 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.

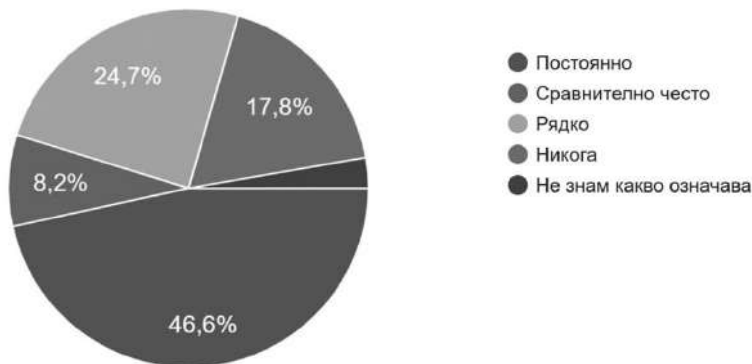
Над половината от анкетираните (54%) превеждат над 1000 страници годишно, което означава грубо по над 3 страници на ден. В съвременните условия производителността на труда на преводачите се увеличава и три страници превод на ден означава, че вероятно анкетираните се занимават и с други дейности и професионалният 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 компютърнопомогнат превод. По тази причина не са изненадващи резултатите, че 46,5% от преводачите използват системи за компютърнопомогнат 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



**Фиг. 6.** Честота на използване на инструментите за компютърноподпомогнат превод

Най-популярните системи за компютърноподпомогнат превод очаквано са Традос, Мемсоурс, мемоКю, Уърдфаст, СмартКАТ. Почти половината от анкетираните 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



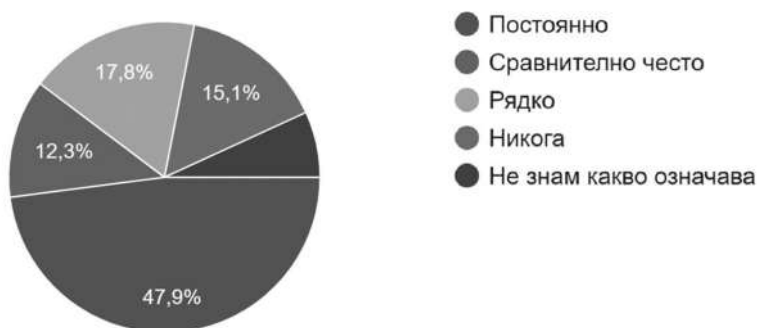
**Фиг. 7.** Оценяване на инструментите за компютърноподпомогнат превод

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 Традос, която въпреки това е прекалено тежка програма“. Отговорите показват, че (интуитивно или не) преводачите имат реалистични очаквания за развитието на системите за компютърноподпомогнат 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



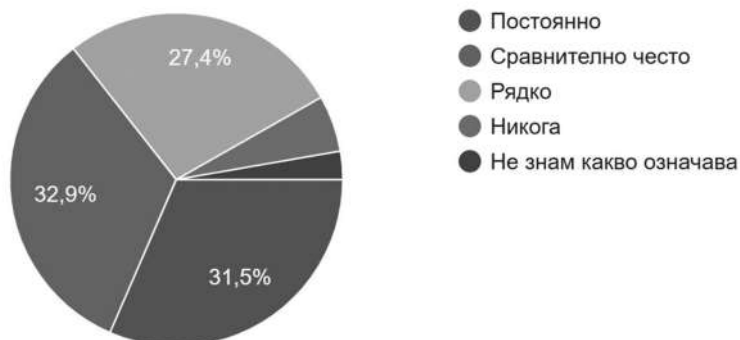
**Фиг. 8.** Отговори на преводачите относно употребата на преводна памет



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?

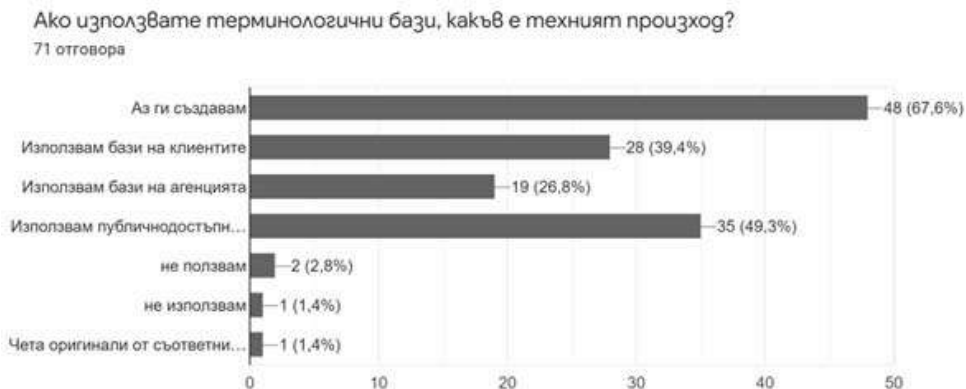
73 отговора



**Фиг. 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 публично-достъпни терминологични бази (сборът е повече от 100%, тъй като отговарящите могат да изберат повече от един отговор: **Фигура 10**).

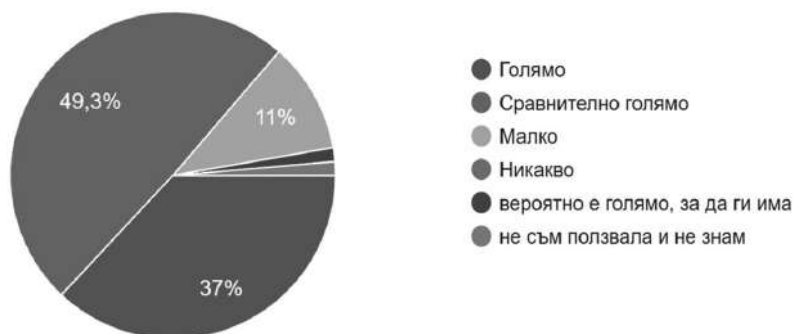
На въпроса дали използват публично-достъпни бази, 20,5% от анкетираните отговарят положително, 20,5% ги използват сравнително често, 32,9% – рядко и 26% никога не ги използват. Като най-използваните публично-достъпни терминологични бази са посочени ИАТЕ и Глосбе. Може да се направи заключението, че сравнително голям процент (41%) от участниците в анкетата използват публично-достъпни бази, възможно е, ако има по-широка популяризация на предимствата на публично-достъпните 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.



**Фиг. 10.** Произход на терминологичните бази

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



**Фиг. 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.

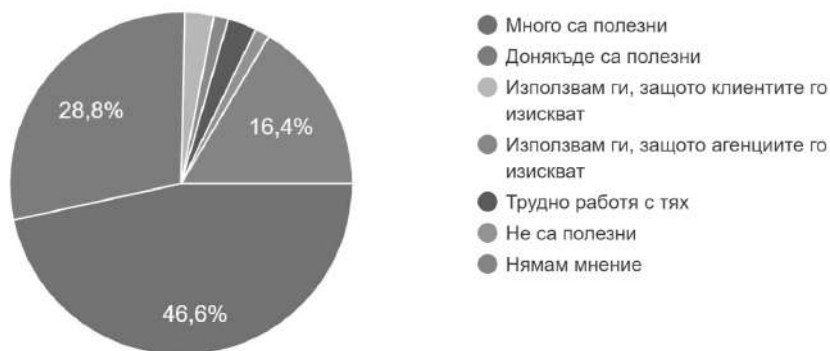
Беше включен въпросът от колко езика се състоят терминологичните бази, които преводачите използват (при условие, че ги включват в работата си). На този незадължителен въпрос отговарят 56 души: 55,4% ги използват за три езика, 8,9% за четири езика, 1,8% за пет езика и 33,9% за повече от пет езика. Отговорът 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 компютърноподпомогнат 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



Фиг. 12. Как преводачите оценяват работата с терминологичните бази

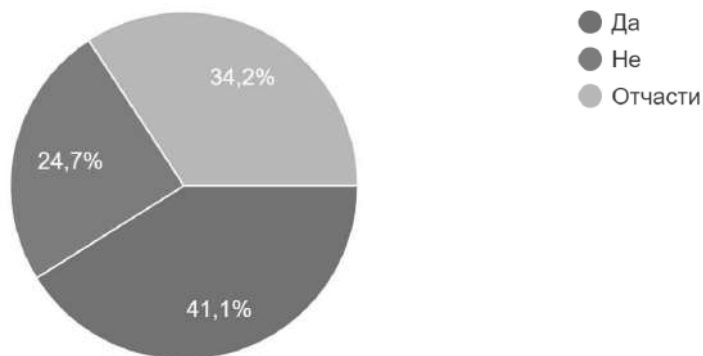
Голямо разнообразие от отговори има на въпроса „Как проверявате качеството на превода?“. Само 17,8% отговарят, че използват функционалностите за проверка на качеството на инструментите за компютърноподпомогнат превод; 9,6% си служат с външни програми, 61,6% (най-голямата група) повторно проверяват текста 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



**Фиг. 13.** Познаване на международните стандарти за качество на превода

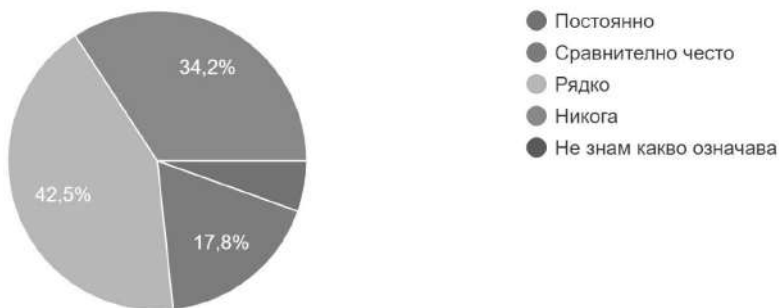
Интересни са резултатите при отговора на въпроса дали преводачите използват машинен превод – преобладават отговорите, че не се използва или че се използва рядко: 42,5% го използват рядко, а 34,2% – никога. Само 5,5% от преводачите посочват, че използват машинен превод постоянно (**Фигура 14**). Отговорите 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 свободnodостъпни 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



**Фиг. 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 подобрение 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 компютърноподпомогнат 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 пуб-

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

### **3. Терминологични бази в системите за компютърнопомогнат превод**

#### **3.1 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/Boyadzhiev 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 абривиатура SUV. Следователно както семантично, така и морфологично термините проявяват определени свойства, които ги отличават от така наречените общоупотребими думи (Уорбъртън/Warburton 2015: 646).

#### **3.2. Термините в контекста на превода**

Употребата на термини в терминологичните бази прави многозначността нежелана (въпреки че може да зависи 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/Thelen 2015: 379). Х. Сагер 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“ (Сагер/Sager 2001: 261).

А. Кис дава значително по-съкратена дефиниция: „терминът (авторът използва названието *terminus technicus*) е израз, принадлежащ към техническия

език“ (Кис/Кis 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/Imre 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/Trumble and Stevenson 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/Łukasik 2012: 98).

### **3.3. 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/De Caluwe and Van Santen 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/Łukasik 2012: 100).

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/Yang 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/Lew 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 терминологичните речници могат да бъдат включени един или повече от следните компоненти (Попова/Ророва 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 регистрацията 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.

### **3.4. 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.

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 персонализирани 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 прегрупираны 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 Арауз/Faber and Araújo 2021: 588).

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

- дефиниция (описание на значението на термина);
- източник;
- контекст на употребата на термина;
- област, към която принадлежи терминът<sup>45</sup>;
- 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.

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./Miller et al. 1990). Всички термини, синоними и преводни еквиваленти, които обозначават едно понятие,

---

<sup>45</sup> 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 взима под внимание какво означава даденият термин в други области.

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/Warburton 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/Bendana and Melby 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 IATE<sup>46</sup>) 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 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 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 (EuroVoc<sup>47</sup>) 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).

<sup>46</sup> <http://iate.europa.eu>

<sup>47</sup> <https://op.europa.eu/en/web/eu-vocabularies>

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. микро-термини.

– Електрoпeдия (Electropedia<sup>48</sup>) 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, 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 Links<sup>49</sup>) 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 главните думи на отделните речници. Полето „източник“ дава възможност за търсене по имената на авторите на речниците.

– Голям терминологичен речник (от фр. Le grand dictionnaire terminologique<sup>50</sup> или GDT) е банка от терминологични записи, създадени от Квебекския офис на френския 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.

– Термиум плус – Банка за терминология и езикови данни на правителството на Канада (Termium Plus: Government of Canada's Terminology and Linguistic Data Bank<sup>51</sup>). Вътре могат да бъдат намерени термини на англий-

---

<sup>48</sup> <https://www.electropedia.org/iev/iev.nsf/Welcome?OpenForm>

<sup>49</sup> <https://termcoord.eu/glossarylinks/>

<sup>50</sup> <http://www.oqlf.gouv.qc.ca/>

<sup>51</sup> [www.termiumplus.gc.ca](http://www.termiumplus.gc.ca)

ски, френски, испански и португалски със съответните съкращения, дефиниции и примери за употреба в широк спектър от специализирани области. Банката може да служи за разбиране на акроним, проверка на официално заглавие и намиране на еквивалент на друг език.

– Многоезиковата терминологична база на ООН (United Nations Multilingual Terminology Database<sup>52</sup> 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 (TermTerm<sup>53</sup>) 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.

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 Terminology<sup>54</sup>) или този на Майкрософт.

### **3.5. 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 преpraщане 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;

---

<sup>52</sup> <http://unterm.un.org>

<sup>53</sup> <https://termterm.kaleidoscope.online/dashboard>

<sup>54</sup> <https://term-extraction.tilde.com/>

- 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 Monitor<sup>55</sup> 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.).

Приликите обаче свършват дотук. Стандартизирана практика е на преводачите да се изпраща предварително подготвена база с термини, които са одобрени от клиента и 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 базата<sup>56</sup>).

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. По време на превод в текстовия редактор на преводача не му се налага да търси термини. Програмата сама ги визуализира.

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/Lew and de Schryver 2014).

---

<sup>55</sup> <https://languagemonitor.com/>

<sup>56</sup> Разбира се, точният начин за добавяне е строго индивидуален за всяка програма за компютърноподпомогнат превод. Тук се предлага обобщение на повечето инструменти за компютърноподпомогнат превод.

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.

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 речник<sup>57</sup>;
- Английско-български онлайн речник на PONS<sup>58</sup>;
- SA речник<sup>59</sup>.

Проблемите с тях произтичат (с малки изключения) основно от липсата на информация за тяхната надеждност и достоверност.

### **3.6. Връзка на терминологичните бази с преводната памет**

Преводната памет обикновено е интегрирана в системите за компютърноподпомогнат превод. Всяко преведено изречение се съхранява в нея. Когато преводачът срещне нов сегмент, който съответства на друг от базата, 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/Тауеbeh 2008: 97), ако са преведени добре или ако преводачът 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

---

<sup>57</sup> <https://www.rechnik-bg.com/>

<sup>58</sup> <https://bg.pons.com/>

<sup>59</sup> <http://notrial.bg/software/windows/sa-dictionary/>

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 компютърноподпомогнат 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.

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.

### **3.7. Създаване и управление на терминологичните бази в системите за компютърноподпомогнат превод**

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 импортиране<sup>60</sup>.

---

<sup>60</sup> <https://gateway.sdl.com/articles/SolutionArticles/000001192>



### 3.8. 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 са<sup>61</sup>:

– Лоджитърм (LogiTerm)<sup>62</sup> 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 (Terminava)<sup>63</sup> – 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 (Jeremy)<sup>64</sup> 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.

– Ердабълуюес Апстор (RWS AppStore)<sup>65</sup> е хранилище, където преводачите могат да търсят различни допълнителни притурки и помощни средства.

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

---

<sup>61</sup> Всички примери за такива програми са взети от списъка на Нимзи: <https://www.nimdzi.com/tbs/>

<sup>62</sup> <https://terminotix.com/index.asp?content=category&cat=4&lang=en>

<sup>63</sup> <http://www1.navcanada.ca/logiterm/addon/terminav/termino.php>

<sup>64</sup> <https://juremy.com/about>

<sup>65</sup> <https://appstore.rws.com/>

информационните технологии в своите изследвания, а резултатите от терминологичните изследвания като речници, глосари и термини се използват в специализираната комуникация между експерти (Кабре/Cabré 1998: 160).

Терминолозите 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 (Varga/Varga 2013: 28 – 29).

### 3.9. 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. Могат да бъдат както независими от системите за компютърноподпомогнат 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.

Годишният доклад на Нимзи<sup>66</sup> за 2021 отбелязва, 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

---

<sup>66</sup> <https://www.nimdzi.com/language-technology-atlas/#the-tms-arena>

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 преведени) 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 и много други.

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/Gilbert 2022).

### **3.10. 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 (Уорбъртън/Warburton 2015: 655 – 656). Така обяснен, начинът на работа с терминологията по време на превод може да изглеж-

да доста лесен. Въпреки това знанията, свързани с терминология (terminology skills) могат да се смятат за отделна компетентност (separate competence) на преводачите. 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, едноезични, многоезични, паралелни и съпоставими корпуси и да придобият умения, които ще им позволяват да използват тези ресурси ефективно (Сикора/Sikora 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 терминологията<sup>67</sup> терминологията и професионални нужди (където се разглежда ролята на терминологията в документацията, лексикографията, езиковата нормализация, езиковото обучение и лингвистичното инженерство)<sup>68</sup>, развитието на основни умения за превод и постигане на умения в разпознаването на регистри на различни езици<sup>69</sup>, разработването на изследвания в области като техническия, литературния и медийния превод, конферентния превод и управлението на терминологията<sup>70</sup>.

Освен академичните институции има редица частни учреждения, които също предлагат обучения и специализирани курсове за терминология. Пример е програмата за сертификат за управление на терминологията превод (Terminology Management Translation Certificate) на Термнет<sup>71</sup>. Организирането на и участието в конференции е друг начин за придобиване на специализирани умения в областта на терминологията. Един такъв пример е Трансфеър (TransFair<sup>72</sup>), където се срещат представители в областта на специализирания превод и на висшите учебни заведения с цел да подготвят студентите за професионалния им живот.

#### **4. Полуавтоматично създаване на многоезикови терминологични бази**

Известно е, че езикът се развива непрекъснато, като се влияе от политически, културни и икономически фактори. По същия начин и специализираната

---

<sup>67</sup> Магистърски курс по терминология на Люксембургския университет [https://www.uni.lu/studies/fhsc/master\\_in\\_learning\\_and\\_communication\\_in\\_multilingual\\_and\\_multicultural\\_contexts/programme](https://www.uni.lu/studies/fhsc/master_in_learning_and_communication_in_multilingual_and_multicultural_contexts/programme)

<sup>68</sup> Онлайн програма за терминология към Института за приложна лингвистика, изследователски център към Университета Помпео Фабра <https://www.upf.edu/en/web/terminologiaonline>

<sup>69</sup> Магистърска програма по превод и терминологични изследвания в Малтийския университет <https://www.um.edu.mt/courses/overview/PMTTFTT-2021-2-O>

<sup>70</sup> Център за изследвания на превода към Виенския университет <https://transvienna.univie.ac.at/en/about-us/>

<sup>71</sup> [www.termnet.eu](http://www.termnet.eu)

<sup>72</sup> <https://etransfair.eu/>

терминология в различни области се развива непрекъснато. П. Фабер констатира, че „ако се направи проверка на честотата на употреба на думите в публикацииите, 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/Faber 2011: 9). Една област, която привлича значително вниманието, е автоматичното извличане на термини (Килгариф и др./Kilgarriff et al. 2014; Коста и др./Costa et al. 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.

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. 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/Warburton 2015: 652).

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 (Сьогард и др./Søgaard et al. 2015: 1713). Конкретно за български език в свое проучване от 2012 г. Д. Благоева, С. Коева и В. Мурдаров констатираат 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/ Blagoeva, Koeva and Murdarov 2012: 14 – 15). 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/Koeva and Stefanova 2022: 2).

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.

#### **4.1 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./Hoisl et al. 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/Kirova 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/Nikolova 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/Nikolova 2018: 5).

При термините, които са съставни думи, най-често срещаната синтактична структура е структурата от подчинено прилагателно и главно съществително

(Коева/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 *търсеца система*), професионализми или такива, които се споделят от определена група специалисти в дадена област (например *метатърсачка*), и професионален жаргон (например *паяк*) (Кирова/Кирова 2004).

#### **4.2 Речници с компютърна терминология**

В мрежата и книжарниците могат да бъдат намерени редица речници на английски и български в областта на компютърната терминология. Те 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. При преглед на речниците, 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. 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.

#### **4.3 Автоматично извличане на термини: методи и добри практики**

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/Kageura and Umino 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 процеси за извличане на термини като характеристики и се стремят да оценят важността на всяка характеристика) (Шаятович/Šajatović 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 леми, 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/Astrakhantsev 2016).

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 термини, използвани за описание на тези нови технологии. Предложението 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./Nosari et al. 2019). За да създадат модела, авторите събират етикетиран 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./Terry et al. 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./Terry et al. 2020: 85).

#### 4.4. 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). В някои инструменти (например Мултитърм<sup>73</sup> и Кютърм<sup>74</sup>) 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 (desktop версии, за които е необходим инсталационен процес), уеббазирани (работят с браузър) и вградени в инструментите за компютърноподпомогнат превод.

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

– Файлфилтърс (fivefilters.org<sup>75</sup>) 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.

---

<sup>73</sup> <https://www.trados.com/products/multiterm-desktop/>

<sup>74</sup> <https://docs.memoq.com/current/en/memoQWeb-help/mqw-qterm.html>

<sup>75</sup> <https://www.fivefilters.org/term-extraction/> [20.07.2022]

#### 4.5. Сравнение на инструментите за автоматично извличане на термини

В проучване от 2016 А. Зарецкая и др. (Zaretskaya et al. 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 Extract<sup>76</sup>) 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 Градос. 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.<sup>77</sup>). Това е джава приложение, което извлича автоматично най-често срещаните термини (думи и съставни думи) на английски, португалски, испански, френски и руски. Предлага 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.

– Търмсут (TermSuite<sup>78</sup>) 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 Engine<sup>79</sup>) е онлайн инструмент, създаден от Лексикал компютинг (Lexical Computing Ltd<sup>80</sup>) 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 ин-

---

<sup>76</sup> <https://appstore.sdl.com/language/app/sdl-multiterm-extract/1165/> [20.07.2022]

<sup>77</sup> <https://www.dail.es/en/artificial-intelligence/> [20.07.2022]

<sup>78</sup> <http://termsuite.github.io/> [20.07.2022]

<sup>79</sup> <https://www.sketchengine.eu/> [20.07.2022]

<sup>80</sup> <https://www.lexicalcomputing.com/> [20.07.2022]

струмент осигурява възможност както за едноезиково, така и многоезиково извличане на термини. Предлагат се различни възможности за търсене. Резултатът може да бъде изтеглен в .tbx или .csv 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 локализират в паралелните документи. Поддържа български.

– Транслейтед (Translated srl<sup>81</sup>) 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.

– Терминус (Terminus<sup>82</sup>) 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 (Kea<sup>83</sup>) 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 (Rainbow<sup>84</sup>) 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 JATE<sup>85</sup>) е инструмент за разработване на и експериментиране с

---

<sup>81</sup> <https://translatedlabs.com/terminology-extraction>

<sup>82</sup> <http://terminus.iula.upf.edu/cgi-bin/terminus2.0/terminus.pl?lInt=En>

<sup>83</sup> <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]

<sup>84</sup> <https://okapiframework.org/wiki/index.php/Rainbow> [20.07.2022]

<sup>85</sup> <https://javalibs.com/artifact/uk.ac.shef.dcs/jate> [20.07.2022]

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 Мултитерм, 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 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:

*Двуетиково извличане (Bilingual extraction)* е функционалност, която се отнася до извличане на термини от изходен и преводен тест 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.

*Двуетиково съставяне на речник (Bilingual dictionaries compilation)* е процес, чрез който се създава двуетиков 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. (Zaretskaya et al. 2016: 5)<sup>86</sup>.

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.

---

<sup>86</sup> <https://wlv.openrepository.com/handle/2436/622550> [20.07.2022]

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

	ЕсДиЕл Мультигърм	Симпъл Екстрактор	ТърмСюит	Скеч Енджин	Транс- лейтед	Терминус	Киа	Рейбол	Яте
Двуезиково извличане	х		х	х					
Сряняване на контекста в изходния и целевия език	х								
Проверка на термин	х	х		х		х	х	х	х
Двуезиково 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
Честотно подреждане на термини	х	х	х	х		х	х	х	х
Поддръжка на различни езици	х		х	х		х	х	х	х
Определяне на минимален брой срещания на термин	х	х		х		х	х	х	х
Показване на лингвистична информация	х		х			х			
Определяне на минимален и максимален брой преводи			х						
Опция за изключване	х	х			х		х	х	х
Избор на минимален и максимален брой думи в състава на термин	х	х					х	х	х
Статистика	х	х	х	х	х	х	х	х	х

*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.

*Двезиково 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).

\* Настройките към 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/Warburton 2021: 163):

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

	ЕсДиЕл Мултиерм	Симпъл Екстрактор	ТермСюит	Скеч Енджин	Транс- лейтед	Терминус	Киа	Рейнбол	Яте
Срвявяване на контекста в изходния и целевия език			x						
Проверка на термин	x	x		x		x	x	x	x
Двуезиково 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	X		Y			X	X	X	
Статистика	X	X	X	X	X			X	X	X	X
Многоезиково извличане на термини	X			X							
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
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		

- Мултитерм (на Традос);
- Търмстар<sup>87</sup> (на Стар Транзит<sup>88</sup>);
- Кростърм<sup>89</sup> (на Акрос<sup>90</sup>);
- Кютърм<sup>91</sup> (на мемоКю);
- Търмбейс (на Мултитранс).

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 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.

## **5. 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 про-

---

<sup>87</sup> <https://www.star-group.net/en/products/terminology-management.html> [20.07.2022]

<sup>88</sup> <https://www.star-group.net/en/home.html> [20.07.2022]

<sup>89</sup> <https://www.across.net/en/knowledge/expert-features/how-to-use-crossterm-instances-efficiently> [20.07.2022]

<sup>90</sup> <https://www.across.net/en> [20.07.2022]

<sup>91</sup> <https://www.memoq.com/extensions> [20.07.2022]

мяна 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 към българските преводачи 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 Термокоп позволява извличането на термини от документи за български и английски<sup>92</sup>.

### 5.1. Двyezиков корпус за извличане на термини

За целите на методиката бе подбран паралелният английско-български корпус в областта на компютърната терминология КюТиЛиип (QTLear V1.2)<sup>93</sup>, 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 авторството и същия начин на споделяне (CC-BY-NC-SA). Създаден е в периода 2014 – 2015 година с европейско финансиране. Корпусът е избран, защото е двyezиков 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.

---

<sup>92</sup> <https://dcl.bas.bg/termextraction/>

<sup>93</sup> <http://metashare.ilsp.gr:8080/repository/browse/qtlearn-corpus-v12/0176c39ae9cd11e4a2aa782bcb074135ba7d767f645a48dca1d50ee3c9504253/> [20.07.2022]

## 5.2. 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.

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.

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 и 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. 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 продуктовете 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 отвори 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 P2P	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

**Фиг. 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 търговски марки, изписани на латиница. Логично несъответствие на резултатите се вижда и от честотата на появяване на термините в текста: за английски 197 пъти, а за двата опита с български по 7 и 8 пъти, думи които далеч не отговарят на статистиката за английски. Те могат да дадат първоначална представа за съдържанието на българската част от корпуса, но качеството на извлечените термини не отговаря на това на извлечените от английския текст.



ята 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. Независимо че програмата е разработена преди почти 20 години, голямо количество учени я използват в своите експерименти: например Ортего Антон и Фернандес (2015)<sup>96</sup> за експертно представяне на знание в двуезикови електронни речници; Р. М. Уюн (2018)<sup>97</sup> за ключовите думи в романите на Халед Хосейни „Ловецът на хвърчила“ и „И планината отекна“.

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 go Settings to activate Windows.
antivirus	32	69.94	antivirus	Nom
usb	31	68.81	usb	Nom

Фиг. 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).

<sup>96</sup> <https://dialnet.unirioja.es/servlet/articulo?codigo=5126623> [20.07.2022]

<sup>97</sup> <https://repository.unair.ac.id/79103/> [20.07.2022]

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 българският 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 компютърноподпомогнат 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. Тук обаче подравняването се използва за създаването на терминологична база. 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 Тилде<sup>98</sup>. За да се използва обаче, са необходими умения за програмиране, което ограничава възможността за по-широко приложение<sup>99</sup>.

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 еквивалент<sup>100</sup>. Запазва се сегментирането на ниво изречение.

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 файловете 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.

### **5.3. 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, .sdlbt. Мем-

---

<sup>98</sup> <https://www.tilde.com/>

<sup>99</sup> <https://github.com/tilde-nlp/taws>

<sup>100</sup> Подравняването може да се използва за всякакъв вид търсене на преводни еквиваленти с различна цел, например стилистична.



сорс, например, позволява импортването на терминологични бази само от формати .xlsx, .tbx и .tbx.xml (от Традос)<sup>101</sup>.

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 Convert<sup>102</sup>), Глосари Конвъртър (Glossary Converter<sup>103</sup>), Ексбенч (ApSIC Xbench<sup>104</sup>), Голдпан Тиемекс/Тибиекс Едитър (Goldpan TMX/TBX Editor<sup>105</sup>).

За настоящия експеримент бе използвана десктоп програмата Олифант (Olifant<sup>106</sup>), 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 (MateCat<sup>107</sup>) и в нея беше създаден тестов проект за превод.

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

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 компютърноподпомогнат 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.

---

<sup>101</sup> [https://www.memsource.com/uploads/2021/05/05/memsource\\_importing\\_term\\_bases.pdf](https://www.memsource.com/uploads/2021/05/05/memsource_importing_term_bases.pdf) [20.07.2022]

<sup>102</sup> <https://www.tbxconvert.gevterm.net/glossary/index.html> [20.07.2022]

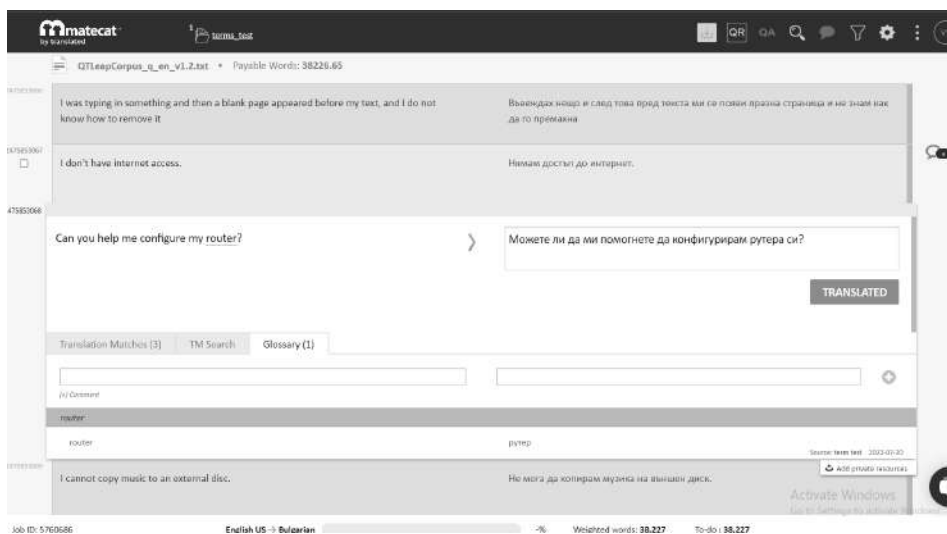
<sup>103</sup> <https://appstore.sdl.com/language/app/glossary-converter/195/> [20.07.2022]

<sup>104</sup> <https://www.xbench.net/index.php/download> [20.07.2022]

<sup>105</sup> <https://logrusglobal.com/goldpan.html> [20.07.2022]

<sup>106</sup> <http://okapi.sourceforge.net/downloads.html> [20.07.2022]

<sup>107</sup> <https://www.matecat.com/> [20.07.2022]



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

#### 5.4. 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*<sup>108</sup>), или са

<sup>108</sup> 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.

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 компютърноподпомогнат 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 *router* 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.

## Изводи

Основната цел на настоящата студия 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 изменението, 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. Направено е описание на новите функции на системите за компютърноподпомогнат превод.

Методиката за полуавтоматично създаване на многоезикова терминологична база от данни се отнася за 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 избраните 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.

## Литература

- Благоева и др. 2012:** Blagoev, D., S. Koeva and V. Murdarov. The Bulgarian Language in the Digital Age. – In: *White Paper Series*. Rehm G. and H. Uszkoreit, Springer, 2012. <<http://www.meta-net.eu/whitepapers/e-book/bulgarian.pdf>>
- Бояджиев 2012:** Бояджиев, Т. Езиковата ситуация у нас в исторически и съвременен план и европейската езикова политика. – В: *Книжовен език и диалекти. Избрани трудове*. Жобов, В. и Г. Колев (ред.), София, Университетско издателство „Св. Климент Охридски“, 2012, стр. 41–83.
- Кирова 2002:** Кирова, Л. Еволюция на българската компютърна терминология и компютърен жаргон. – В: *LiterNet*, № 2 (27), <<https://litenet.bg/publish3/lkirova/evolution.htm>>
- Кирова 2004:** Кирова, Л. Компютърната лексика – актуални процеси и тенденции. – В: *LiterNet* № 5 (54), <<https://litenet.bg/publish3/lkirova/lex-procesi.htm>>
- Коева 2007:** Koeva, 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>>
- Коева и Стефанова 2022:** Koeva, 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](https://european-language-equality.eu/wp-content/uploads/2022/03/ELE__Deliverable_D1_5_Language_Report_Bulgarian_.pdf)>
- 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](https://www.shu.bg/wp-content/uploads/file-manager-advanced/users/razvitie/as/2018/phd-D.Nikolova/avtoreferat_D_Nikolova.pdf)>
- 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>>

## References

- Aly 2017:** Aly, El-H. Workplace-Oriented Definition of Translation Competence. – In: *Translation Today*, Tariq Khan (Eds.), 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](https://www.ntm.org.in/download/ttvol/volume11-2/E-TT%20Vol_11-2_2017.pdf)>
- Alshehab 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](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)>
- Arenas 2020:** Arenas, A. G. Pre-editing and post-editing. – In: *The Bloomsbury Companion to Language Industry Studies*. E. Angelone, M. Ehrensberger-Dow and G. Massey (Eds.), 1 ed., Bloomsbury Academic <<https://research.rug.nl/en/publications/pre-editing-and-post-editing>>
- Astrakhansev 2016:** Astrakhansev, N. ATR4S: Toolkit with State-of-the-art Automatic Terms Recognition Methods in Scala. <<https://arxiv.org/abs/1611.07804>>

- Bendana and Melby 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](https://www.academia.edu/34926861/Everything_you_ever_wanted_to_know_about_translation_melby_bendana_1)>
- Bowker and Ciro 2019:** Bowker L. and J.B. Ciro. *Machine Translation and Global Research: Towards Improved Machine Translation Literacy in the Scholarly Community*, pp. 37–54.
- Bulté and Tezcan 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](https://www.researchgate.net/publication/334745562_Neural_Fuzzy_Repair_Integrating_Fuzzy_Matches_into_Neural_Machine_Translation)>
- Cabré 1998:** Cabré, M. T. *Terminology: Theory, methods, and applications*. Vol. 1 John Benjamins Publishing.
- Carreira 2020:** Carreira, O. Defining Transcreation from the Practitioners Perspective: an Interview-based Study. – In: *Translatum Nostrum. La traducción y la interpretación en el ámbito humanístico*. Comares. C. Botella Tejera, J. Franco Aixelá and C. Ilescu Gheorghiu (Eds.), <[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](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)>
- Costa et al. 2016:** Hernani, Costa., 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>>
- Faber 2011:** Faber, P. The dynamics of specialized knowledge representation: Simulational reconstruction or the perception–action interface. – In: *Terminology* 17:1, John Benjamins Publishing Company, pp. 9–29. <[https://www.researchgate.net/publication/263063290\\_The\\_dynamics\\_of\\_specialized\\_knowledge\\_representation\\_Simulational\\_reconstruction\\_or\\_the\\_perception-action\\_interface](https://www.researchgate.net/publication/263063290_The_dynamics_of_specialized_knowledge_representation_Simulational_reconstruction_or_the_perception-action_interface)>
- Faber and Araújo 2021:** Faber, P. and Araújo P. Designing Terminology Resources for Environmental Translation. – In: *The Oxford Handbook of Translation and Social Practices*, Meng, Ji. and S. Laviosa (Eds.), Oxford Academic pp. 587–615. <[https://www.researchgate.net/publication/350213272\\_Designing\\_Terminology\\_Resources\\_for\\_Environmental\\_Translation](https://www.researchgate.net/publication/350213272_Designing_Terminology_Resources_for_Environmental_Translation)>
- Gambier 2019:** Gambier, Y. Impact of technology on Translation and Translation Studies. – In: *Russian Journal of Linguistics*, Серия: Лингвистика, Т. 23. No 2. С, pp. 344–361. <[https://www.academia.edu/39818451/Impact\\_of\\_technology\\_on\\_Translation\\_and\\_Translation\\_Studies\\_1?email\\_work\\_card=title](https://www.academia.edu/39818451/Impact_of_technology_on_Translation_and_Translation_Studies_1?email_work_card=title)>
- Garcia 2015:** Garcia, I. Computer-aided translation: systems. – In: *Routledge Encyclopedia of Translation Technology*, S.-W. Chan (Ed.), University of Western Sydney, Australia, pp. 68–87. Retrieved from <http://UWSAU.ebib.com.au/patron/FullRecord.aspx?p=1843560>
- Gilbert 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](https://www.researchgate.net/publication/359586445_Solving_Terminology_Management_Pain_Points_One_Termbase_Management_System_to_Rule_Them_All/citations)>
- Gouadec 2007:** Gouadec, D. Translation as a profession. University of Rennes, Benjamins translation library, 0929-7316, v. 73, 2007 <<https://benjamins.com/catalog/btl.73>>

- Dara et al. 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>>
- De Caluve and Van Santen 2003:** De Caluve, J. and A. Van Santen. Phonological, morphological and syntactic specifications in monolingual dictionaries. – In: *A Practical Guide to Lexicography*, P. van Sterkenburg (Ed.), Institute for Dutch Lexicology, Leiden, pp. 71–82. <<https://benjamins.com/catalog/tlrp.6.10dec>>
- De Schryver 2003:** De Schryver, G.-M. Lexicographers’ dreams in the electronic-dictionary age. – In: *International Journal of Lexicography*, 16(2), pp. 143–199. <<https://academic.oup.com/ijl/article-abstract/16/2/143/925134>>
- der Meer 2020:** van der Meer, J. Translation technology – past, present and future. – In: *The Bloomsbury Companion to Language Industry Studies*. E. Angelone, M. Ehrensberger-Dow, and G. Massey (Eds.), Bloomsbury Academic, 1 ed., pp. 285–310. <<https://www.bloomsburycollections.com/book/the-bloomsbury-companion-to-language-industry-studies/ch13-translation-technology-past-present-and-future>>
- Drouin 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](https://www.researchgate.net/publication/228683045_Term_extraction_using_non-technical_corpora_as_a_point_of_leverage)>
- Eriguchi et al. 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*, Hong Kong, China, 2019, pp. 123–130. <<https://www.aclweb.org/anthology/D19-5214>>
- Hoisl et al. 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>>
- Hosarri et al. 2019:** Hosarri, M., Dev S., Kelleher J. D. TEST: A Terminology Extraction System for Technology Related Terms. – In: *ICCAE*, Perth, Australia, pp. 23–25. <<https://arxiv.org/pdf/1812.09541.pdf>>
- Imre 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](https://www.researchgate.net/publication/287911571_Term_Bases_Reloaded)>
- Imre 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](https://www.researchgate.net/publication/339289785_AN_UPDATED_VIEW_ON_TRANSLATION_STUDIES)>
- Imre and Barabás 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](https://www.researchgate.net/publication/339289691_THE_EFFICIENCY_OF_A_MULTILINGUAL_TERM_BASE)>
- Kageura and Umino 1996:** Kageura, K. and B. Umino. Methods of automatic term recognition. – In: *Terminology*, John Benjamins Publishing Co, vol 3(2), pp. 259–289. <[http://www.iro.umontreal.ca/~felipe/IFT6010-Automne2011/resources/tp3/gabriel\\_bc/Kageura\\_Umino\\_1996.pdf](http://www.iro.umontreal.ca/~felipe/IFT6010-Automne2011/resources/tp3/gabriel_bc/Kageura_Umino_1996.pdf)>
- Kanavos and Kartsaklis 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), Denver, USA. <[https://www.researchgate.net/publication/269519356\\_Integrating\\_Machine\\_Translation\\_With\\_Translation\\_Memory\\_A\\_Practical\\_Approach](https://www.researchgate.net/publication/269519356_Integrating_Machine_Translation_With_Translation_Memory_A_Practical_Approach)>

- Kilgarriff et al. 2014:** Kilgarriff, A., M. Jakubiček, 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](https://www.researchgate.net/publication/271849661_Finding_Terms_in_Corpora_for_Many_Languages_with_the_Sketch_Engine)>
- Kis 2005:** Kis, Á. Terminusalkotás: a terminológiai helyzet és a terminológiai szerep. Mindent fordítunk, és mindenki fordít” Értékek teremtése és közvetítése a nyelvészetben, pp. 105–112.
- Kodura 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>>
- Lukasik 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. 98–108. <[https://www.researchgate.net/publication/294836193\\_Terminological\\_dictionary\\_as\\_a\\_comprehensive\\_cognitive\\_and\\_linguistic\\_tool](https://www.researchgate.net/publication/294836193_Terminological_dictionary_as_a_comprehensive_cognitive_and_linguistic_tool)>
- Lew 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](https://www.researchgate.net/publication/265252133_From_paper_to_electronic_dictionaries_Evolving_dictionary_skills)>
- Lew and de Schryver 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>>
- Miller et al. 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](https://www.researchgate.net/publication/31441225_Introduction_to_WordNet_An_On-line_Lexical_Database)>
- Olohan 2020:** Olohan, M. ,Self-translation. – In: *Routledge Encyclopedia of Translation Studies*, Baker M. and G. Saldanha (Eds.), Third edition by Routledge <<https://www.routledge.com/Routledge-Encyclopedia-of-Translation-Studies-3rd-Edition/Baker-Saldanha/p/book/9781138933330>>
- Pym 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/>>
- Pym and Torres-Simón 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](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)>
- Ray and Kelly 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>>
- Ruokonen and Koskinen 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>>
- Sager 2001:** Sager J. Terminology theory. – In: *Routledge Encyclopedia of Translation Studies*. Baker, M. (Ed.), London and New York, Routledge, pp. 250–261 <[https://books.google.bg/books?id=ewBfSBo8rRsC&printsec=frontcover&hl=bg&source=gs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.bg/books?id=ewBfSBo8rRsC&printsec=frontcover&hl=bg&source=gs_ge_summary_r&cad=0#v=onepage&q&f=false)>



- Sdobnikov 2018:** Sdobnikov, V.V. New trends in translation. – In: *Kazan bulletin of young scientists*, 4, pp. 72–79.
- Shields 1999:** Shields, M. Slaves to the Computer. – In: *Institute of Translation and Interpreting Bulletin*, October 1999, pp. 4–5.
- Sikora 2018:** Sikora, I. Developing terminological competence in a specialized translation classroom-problems, resources, and tools. – In: *Częstochowa: Wydawnictwo Politechniki Częstochowskiej*, Maliszewski J. (red.), Częstochowa 2018. <[https://www.academia.edu/37422238/Developing\\_terminological\\_competence\\_in\\_a\\_specialized\\_translation\\_classroom\\_problems\\_resources\\_and\\_tools](https://www.academia.edu/37422238/Developing_terminological_competence_in_a_specialized_translation_classroom_problems_resources_and_tools)>
- Sin-wai 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>>
- Sin-wai 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>>
- Søgaard et al. 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*, Beijing, China, Association for Computational Linguistics July, pp. 1713–1722. <<https://www.aclweb.org/anthology/P15-1165.pdf>>
- Tayebeh 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](https://www.researchgate.net/publication/271602313_Translation_Memories_and_the_Translator)>
- Terryn et al. 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)*, Language Resources and Evaluation Conference (LREC 2020), pp. 85–94. <<https://biblio.ugent.be/publication/8661811/file/8661815.pdf>>
- Thelen 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](http://www.trans-kom.eu/bd08nr02/trans-kom_08_02_03_Thelen_Terminology.20151211.pdf)>
- Timko 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>>
- Toftedahl 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](https://www.ihci-conf.org/wp-content/uploads/2020/07/01_202010L017_F034.pdf)>
- Trumble and Stevenson 2002:** Trumble, W. R., and Stevenson, A. (Eds.), *Shorter Oxford English Dictionary*. Fifth Edition, North Carolina, publisher: Oxford University Press.
- Varga 2013:** Varga, C. Virtualization of Research in Terminology. Cloud-based Terminology Management Tools. – In: *Transactions on Modern Languages*, Scientific Bulletin of the Politehnica University of Timișoara, Vol. 12, No. 1-2. <[https://www.academia.edu/5631484/Virtualization\\_of\\_Research\\_in\\_Terminology\\_Cloud\\_based\\_Terminology\\_Management\\_Tools?email\\_work\\_card=title](https://www.academia.edu/5631484/Virtualization_of_Research_in_Terminology_Cloud_based_Terminology_Management_Tools?email_work_card=title)>

- Vidal et al. 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>>
- Wang and Ping 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](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)>
- Warburton 2015:** Warburton, K. Terminology management. – In: *The Routledge Encyclopedia of Translation Technology*, Sin-wai, C. (Ed), New York, Routledge, pp. 644–661. </<https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9781315749129&type=googlepdf>>
- Warburton 2021:** Warburton, K. *The Corporate Terminologist*. University of Illinois at Urbana-Champaign, John Benjamins. </<https://www.jbe-platform.com/content/books/9789027260093>>
- Wright 2001:** Wright, S. E. Data Categories for Terminology Management. – In: *Handbook of Terminology Management: Volume 2: Application-Oriented Terminology Management*. Wright S. E. and G. Budin (Eds), John Benjamins Publishing Company, pp. 552–571.
- Yang 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>>
- Zafra 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>>
- Zanettin 2015:** Zanettin, F. Concordancing. – In: *Routledge Encyclopedia of Translation Technology*. S.-W. Chan (Ed.), University of Perugia, Italy, pp. 475–487. </<https://www.taylorfrancis.com/chapters/edit/10.4324/9781315749129-40/concordancing-federico-zanettin>>
- Zarei and Fumani 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](https://www.researchgate.net/publication/324835559_Impact_of_Online_Versus_Hardcopy_Dictionaries'_Application_on_Translation_Quality_of_Iranian_M_A_Translation_Students)>
- Zaretskaya et al. 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>>

Виктория Петрова-Любенова  
Институт за български език „Проф. Любомир Андрейчин“  
Българска академия на науките  
viktoriya@dcl.bas.bg

Viktoriya Petrova-Lyubenova  
Institute for Bulgarian Language “Prof. Lyubomir Andreychin”  
Bulgarian Academy of Sciences  
viktoriya@dcl.bas.bg