

**ПРЕВОД И ИНТЕРАКТИВЕН МАШИНЕН ПРЕВОД,  
БАЗИРАН НА ПРИМЕРИ. ПРОГРАМАТА *TREFL***

*Руси Николов, Малина Дичева*  
*Пловдивски университет „Паусий Хилендарски“*

**TRANSLATION AND INTERACTIVE MACHINE TRANSLATION  
BASED ON EXAMPLES EXCERPTED FROM *TREFL***

*Roussi Nikolov, Malina Dicheva*  
*Paisii Hilendarski University of Plovdiv*

*TREFL* is a portable and easy-to-use application for Windows with powerful NLP functionality. It takes its name – *Translation REFerence Library* – from its database management capabilities and has been developed as a multilingual reading, writing and translation aid tool. It combines selected characteristics from Translation Memory Systems and Internet/Desktop Search Engines, plus some elements of semantic search. It is intended to be used as a versatile and effective reading, writing and translation aid tool capable of managing very large sets of data.

**Key words:** Translation Theory, Reading, Writing and Translation Aid Tool

Пълното или частичното автоматизиране на превода е научна област, която се развива с оглед на решаването на две важни практически задачи: 1) предварителен достъп – макар и частичен и ненадежден – до информация, кодирана с помощта на непознат за потребителя естествен език, и 2) улесняване, ускоряване и хомогенизиране на специализирания превод въз основа на *повторяемостта* на преводните съответствия – на лексикално, фразеологично и синтактично равнище.

Значимостта на първата задача лесно може да се оцени косвено, като се отчетат съответните интелектуални и материални инвестиции. Това развитие на утилитарния подход в превода, продиктувано от доказана практическа необходимост, води от своя страна до качествени подобрения, които разширяват възможностите на машинния превод до такава степен, че вместо да се противопоставя на системите за преводаческа памет (Косева 2010: 111), той се превръща в полезно до-

пълнение и неделима част на тези професионални инструменти за превод. Този процес може да се обясни и с обстоятелството, че с нарастването на относителната тежест на „суровите“ преводачески данни в съотношението данни/правила, т.е. с изпреварващото нарастване на броя и на дължината на конкретните преводни единици, характеристиката „*машинен*“ в понятието „*машинен превод*“ придобива във все по-голяма степен условен смисъл. Пример: (англ.) *Actions speak louder than words.* → *Действията говорят повече от думите*<sup>1</sup>. Изкуственият интелект се свежда в случая до автоматично извличане и организиране на информация, която е продукт на човешка дейност.

Твърдението за „*хуманизиране*“ на машинния превод при базираните на данни подходи е логически строго и практически може да се илюстрира с многобройни примери, но то изразява само една тенденция, защото общата статистическа оценка остава негативна – машинният превод на идиоматични изрази и поговорки е резултат най-често на прилагането на елементарни трансферни правила – лексикални и граматически – или на преводни съответствия, извлечени от буквален, непрофесионален човешки превод. Пример: (англ.) *smart as a bag of hammers* → *умна като торба с чукове*<sup>2</sup>.

По-деликатно е да се коментира *повторяемостта* (или *аналогията*) – ключова дума за всички подходи за пълно или частично автоматизиране на превода. Повторяемостта не накърнява творческото начало и произтичащото от него авторство в превода (толкова скъпи за специалистите по художествен превод), защото повторяемостта е неотменим съпътстващ компонент на всяка интелектуална дейност<sup>3</sup>. Противопоставянето на художествения и специализирания превод е традиционно, но двете дейности могат да се обединят в следната обща дефиниция на превода:

**Превод:** прекодиране на текст от един на друг естествен език, при което се запазва *качеството на информацията*, т.е. нейният потенциал да променя по определен начин когнитивното и емоционалното състояние на човека<sup>4</sup>.

---

<sup>1</sup> translate.google.com.

<sup>2</sup> www.bing.com/translator.

<sup>3</sup> „Like all human activities requiring intelligence, translation too has repetitive components and creative components“. (Батачария 2015: XIX)

<sup>4</sup> „l'information est ce qui provoque un changement d'état psychique chez un être humain et, peut-être, une action; c'est aussi ce qui provoque un changement dans un système inanimé“ (Мера 1973 : 373).

Понятието *качество на информацията* е заимствано от кибернетиката. Използваното там по-общо понятие *състояние на система* е конкретизирано тук с *когнитивното и емоционалното състояние на човека*.

Твърдението, че „*превеждането не е просто прекодиране*“ (Димова 2000: 23), е широко разпространено. В интерпретативната теория „да преведеш, не означава да транскодираш, а да разбереш и да изразиш отново“ (Селескович, Ледерер 2001: 19)<sup>5</sup>. Въпреки че очевидно противоречи на горната дефиниция, това твърдение е приемливо и полезно, но с необходимото уточнение, че понятието *прекодиране* е използвано от цитираните автори очевидно в смисъл на „формално прекодиране“, т.е. промяна на кода на съобщението чрез прилагането на система от краен брой стриктни правила. Така можем да прекодираме един текст от *Unicode* в *ASCII* например, но не и от един естествен език на друг. От друга страна, всеки текст – изходен или целеви – е кодирано съобщение (кодирано с помощта на естествен език) и в този смисъл, като се вземе предвид естеството на „суровината“ и на „крайния продукт“, може да се твърди, че *превеждането е просто прекодиране*; отделен е въпросът, че такова прекодиране не е никак просто, то може да се формализира само частично и следователно не подлежи на пълно автоматизиране.

От горната дефиниция може да се изведе и абстрактна дефиниция на идеално преведен текст: това е текст, който променя когнитивното и емоционалното състояние на идеални билингви така, както би го променил оригиналният текст. Обратно, ако приемем за постулат интуитивното понятие *идеален превод*, можем да изведем теоретичен критерий за идентифициране на идеални билингви.

В дефиницията за превод понятието *качество на информацията* е предпочетено пред по-популярното, но различно понятие *съдържание*, защото само първото отчита значението и на *формата на текста*: така както въздейства музиката, така и формата – прозаична или поетична – на текстовото съобщение може да катализира въздействието му върху емоционалното състояние на човека – слушател или читател. Езиковата форма на текста влияе и на когнитивното състояние

<sup>5</sup> Вж. още (Ледерер 1994: 11): „La théorie interprétative [...] a établi que le processus [de la traduction] consistait à comprendre le texte original, à déverbaliser sa forme linguistique et à exprimer dans une autre langue les idées comprises et les sentiments ressentis“.

на човека чрез отчитането например на връзката между ритъм на речта и вербална памет<sup>6</sup>.

Идеята за еквивалентността на въздействието върху реципиента като критерий за преводна еквивалентност не е нова, но се подценява нейният потенциал за кратка и същевременно строга дефиниция на понятието „еквивалентност“ (и следователно на самото понятие „превод“). В своя релевантен анализ на развитието на преводознанието, А. Димитрова (2007: 188) отбелязва, че „*то* [понятието „еквивалентност“] *е дефинирано и типологизирано по различни начини* [...] като авторите докъм 80-те години на ХХ век] *търсят пътища към постигането на максимално съответствие между оригинала и превода по отношение на комуникативната цел, информацията и въздействието върху реципиента*“. Комуникативната цел обаче не може да бъде различна от целта на всяко едно съобщение: дори когато съобщението е въпросително, подбудително или възклицателно изречение, първичната, непосредствена цел е да предаде информация, която има някаква стойност за реципиента, а тази стойност се измерва именно с психологическото въздействие върху него на когнитивно и емоционално равнище. Изброените три критерия за еквивалентност се свеждат следователно до един. Същият автор посочва още (Димитрова 2007: 188 и сл.), че „*от 70-те години на ХХ век насам започват да се развиват нови комуникативно-прагматични схващания в преводознанието, които се противопоставят на убеждението, че еквивалентността е необходима и осъществима в превода*“. Тези схващания вероятно представляват интерес с оглед на специфичната проблематика на художествения и в частност на стихотворния превод. Липсата на еквивалентност обаче като единствен обективен критерий за качество на превода, и по-точно отричането на стремежа към еквивалентност поощрява свободната и субективна интерпретация, което изолира преводознанието, отдалечавайки го от възможността за ползотворно интердисциплинарно взаимодействие с точни науки като компютърната лингвистика.

Макар и на друго интерпретационно равнище, *повторяемостта* се оказва ключова дума не само за автоматизирания превод, но и за всеки превод, с уточнението, че *повторяемостта* засяга не само частичните и формални преводни съответствия, но и текста като цяло, и по-специално функционирането му като инструмент за въздействие. Задачата на преводача в чист вид е да приеме целта на автора за своя и да я постигне с различни средства. В този смисъл не може да се го-

---

<sup>6</sup> „Les textes rythmés sont plus facilement appris que les textes non rythmés“ (Фрес 1974: 175).

вори за равнопоставеност между автор и преводач: авторът е автор на целта и неин изпълнител, а преводачът е само изпълнител на същата цел, но с различни средства. Различието на средствата не се свежда единствено до различие в използваните езици. Еднакво добри преводи на един текст, направени от различни преводачи на един и същ целеви език, имат в една или друга степен различна езикова форма. Естественният език е редувант и предоставя алтернативни изразни средства. Изборът им зависи от менталната езикова „база от данни“ на преводача. С оглед на това възниква следният въпрос: при коя от следните две ситуации съществуват по-добри предпоставки за творческа изява на преводача: при равни други условия 1) преводачът разчита единствено на своя когнитивен потенциал и 2) преводачът ползва и колективния ум на преводачи от цял свят, материализиран под формата на свръхголеми структурирани паралелни езикови ресурси, управлявани от свръхефективни и интелигентни информационни системи за достъп до релевантна информация? Очевидният отговор на този въпрос е основание да определим такива софтуерни системи от езикови ресурси и програми за тяхното управление като помощни инструменти не само за превод, но и за творческо писане. Именно такъв инструмент е представената по-долу програмата *TREFL*<sup>7</sup>.

Технологията за превод *TREFL* е интерактивен и опростен вариант на подхода за машинен превод, базиран на примери (англ.: *EBMT, Example-based machine translation*). Опростяването е алгоритмично (софтуерно), но то се компенсира чрез интерактивното действие, т.е. чрез поставяне на по-високи изисквания към човека преводач. При този подход се реализират в известен смисъл етапите *девербализация – ревербализация* (Ледерер 1994) под формата съответно на компресиране и декомпресиране на информацията в процеса на превода. На практика девербализацията има само количествен характер – компресирането на информацията се основава на езиковата редувантност и на възможността да представим абстрактно една проста идея с помощта на ключови думи, което е свързано обикновено с частична загуба на информация, но позволява да се филтрират автоматично оптимален брой кандидати за преводни еквиваленти; информацията се възстановява интерактивно при окончателното филтриране на междинните резултати.

Подходът за машинен превод *EBMT* е концептуално строг и интелектуално привлекателен, какъвто е и подходът, базиран на правила

<sup>7</sup> <http://web.uni-plovdiv.bg/rousni/>

(*RBMT*). И двата подхода обаче, приложени самостоятелно, а не в хибридни системи, са еднакво неуспешни в практическите им приложения (в съпоставка със статистическия машинен превод). Естествените езици трудно се поддават на моделиране чрез системи от краен брой формални правила. Преводът *EBMT* е базиран на сурови данни – двуезични корпуси от паралелни текстове, но едновременно с това използва правила за оценка на подобие (или аналогията, формално – разстоянието) между сегмента за превод, от една страна, и сегментите на същия език в данните на програмата, от друга, което е основание подходът да се определи от някои автори като хибриден – базиран на данни и на правила (Батачария 2015: XIX)<sup>8</sup>. Тук обаче правилата не просто допълват и компенсират закономерната (т.е. задължителната) липса на достатъчно данни, напротив, тяхното несъвършенство би компрометирало напълно резултата от машинния превод независимо от количеството и качеството на наличните паралелни текстове.

Като се изхожда от това, вместо да се търси най-близката аналогия на изречение за превод в базата данни чрез формални, автоматични, но несъвършени алгоритми, в *TREFL* се използва следният метод: преводачът/пишещият „компресира“ входната информация под формата на ключови думи, например „катастрофа“ и „жертва“, и избира опцията семантично търсене, например в преводната памет *български-френски*. Получените резултати (вж. Фиг. 1) подсказват различни начини на изразяване на френски език на идеята „катастрофата/инцидентът/бедствието... взе жертви, отне/коства живота на...“, която се изразява на френски език със съвсем различни изрази, в случая лексикални средства (напр. *катастрофата направи жертви* в буквален превод).

---

<sup>8</sup> „EBMT tries a combination: data supplies translation parts that rules recombine to produce translation.“

[катастрофа] [пътуващи]

L'accident a tué les 7 occupants de l'appareil, tous Américains, en partance pour Dubaï.

**Бедствието** взе около 2 000 **жертви**.

Ce désastre fit environ 2 000 victimes.

автобусна **катастрофа** с много **жертви**

accident de bus avec de nombreuses victimes

[катастрофа] [вземам **жертви**] [отнемам/коствам живота на]

L'accident a fait deux victimes.

Un accident a coûté la vie, ce lundi à Venère, à une mère de famille âgée de 49 ans, domiciliée à Valay.

Най-малко 100 са **загиналите** при самолетна **катастрофа** в Мадрид.

Un accident d'avion fait au moins 100 morts à Madrid.

*Фиг. 1. Екранна снимка на резултатите от семантично търсене в базата данни български-френски с ключови думи „катастрофа“ и „жертва“*

Тази полезна информация при превод от български на френски (или просто при писане на френски език) е труднодостъпна с помощта на традиционни и престижни средства като лексикални или фразеологични речници. Факт е обаче, че машинният преводач на Гугъл превежда точно, например:

*Катастрофата взе три жертви. → L'accident a fait trois victimes.*

Семантичното търсене с *TREFL* чрез ключови думи предлага, за разлика от машинния превод, селекция от варианти. Естествено възниква въпросът за семантичното търсене в глобалната мрежа като приоритетна възможност. Отговорът на този въпрос е положителен, но тази възможност не е безалтернативна; нещо повече – тя не е оптимална с оглед на следните дадености:

- В интернет се индексират текстове, а не изречения, което означава, че две „ключови“ думи може да не са свързани в проста мисъл, т.е. едната да се намира в началото, а другата в края на дълъг текст.
- Текстовете в интернет са двуезични и паралелни по изключение. (От друга страна, разбира се, ключовите думи могат да се въведат директно на целевия език. Освен това, използвайки опции за разширено търсене, потребителят може да определи статистически типичните за целевия език колокации, например: „une \* bavure“ → „une grosse bavure“




- В интернет потребителят няма контрол върху индексираните данни, не може да ги обогатява и персонализира. И както е известно, „прекалено голямата информация убива информацията“.
- Семантичното търсене в интернет е с утилитарна насоченост, но не и на лингвистично равнище.

Програмата *TREFL* е локален аналог на интернет търсачка като *Гугъл* по своите основни и разширени опции за търсене, по своето бързо действие и поддържането на съвременните *html* стандарти за мултимедийно съдържание. Локалният характер на *TREFL* ограничава възможностите за управление на данни в мрежа, но по същата причина отпадат ограниченията (свързани с мрежовата сигурност) за взаимодействие между *html* елементи и локално съдържание. Програмата позволява създаването на хипервръзки от *html* страницата с резултатите от търсенето към цитирания мултимедиен документ, съхраняван на компютъра, например към конкретния сегмент от филм, съдържащ съответната реплика в резултатите на търсенето (вж. Фиг. 2).



Фиг. 2. Снимка на част от екрана с резултати от търсене (car wreck), илюстриращи възможни хипервръзки



- Бутоните със символи ,  и  отварят съответния локален документ; при аудио- или видеофайл документът се отваря за съответния на репликата времеви интервал.
- Снимката с автомобила е в действителност анимация, съответстваща на локален .gif файл.
- Подчертаните сегменти в текста от резултатите са специални връзки (при задържане на курсора на мишката върху тях) към прозорец с обяснителни бележки или снимка/видео.
- Стандартни хипервръзки към интернет сайтове.
- Връзки за всевъзможни операции в операционната система *Windows*; в примера – за автоматично въвеждане на данни в регистъра.

Практическата целесъобразност на набора от функции на една приложна програма или на какъвто и да е инструмент не е достатъчно условие да определим инструмента като полезен. Допълнително условие е ергономичността – икономичност на жестовете и на усилията за тяхното усвояване и запомняне особено когато броят на функциите е голям. Това изискване няма концептуална връзка с лингвистичните и чисто преводачески аспекти на програмата, но неизпълнението му би ги компрометирало или обезсмислило напълно. Кратките указания за работа с програмата в придружаващата я документация и нейните характеристики показват стремеж за търсене на оптималното съотношение между комплексност (мултифункционалност), висока производителност (бързо действие, слабо зависимо от количеството на данните) и ергономичност (удобство за работа). Придобиването на умения за ефективна работа с програмата изисква първоначални интелектуални инвестиции – незначителни в техническо отношение, но професионално предизвикателни на лингвистично равнище, както е показано в наличните в базата данни на програмата примери.

## ЛИТЕРАТУРА

**Батачария 2015:** Bhattacharyya, P. *Machine Translation*. London, New York: CRC Press, Boca Raton, 2015.

**Димитрова 2007:** Димитрова, А. *Играта в творчеството на Херман Хесе – естетически, езикови и преводни аспекти*. [Dimitrova, A. *Igrata v tvorchestvoto na Herman Hese – esteticheski, ezikovi i prevodni aspekti.*] Велико Търново: Фабер, 2007.

- Димова 2000:** Димова, А. *Увод в теорията на превода*. [Dimova, A. *Uvod v teoriyata na prevoda.*] Шумен: УИ „Епископ Константин Преславски“, 2000.
- Косева 2010:** Косева, Е. *Технологията Преводна Памет*. [Koseva, E. *Tehnologiyata Prevodna Pamet.*] // *Научни трудове на Русенския университет* – 2010, том 49, серия 3.2, 111 – 115.
- Ледерер 1994:** Lederer, M. *La traduction aujourd’hui*. Hachette Paris: F.L.E., , 1994.
- Мера 1973:** Mérat, R. *Informatique*. // *Encyclopédie des sciences industrielles*. Paris: Quillet, 1973.
- Селескович, Ледерер 1984:** Seleskovitch, D., Lederer, M. *Interpréter pour traduire*. Paris: Coll. Traductologie, No 1, Didier Erudition, 1984.
- Фрес 1974:** Fraisse, P. *Psychologie du rythme*. Paris: Presses Universitaires de France, 1974.