



**ДА СЕ ДОВЕРЯВАМЕ ЛИ НА МАШИНИЯ ПРЕВОД?
(ЕЛЕКТРОННИ ЕЗИКОВИ РЕСУРСИ, ПОДПОМАГАЩИ
СПЕЦИАЛИЗИРАНИЯ ПРЕВОД)**

Красимира Петрова
krapetrova@slav.uni-sofia.bg

**Софийски университет „Св. Климент Охридски“
София, бул. “Цар Освободител“ №15
РЕПУБЛИКА БЪЛГАРИЯ**

Ключови думи: специализиран превод, машинен превод, електронни езикови ресурси

Резюме: Съвременното глобализиращо се многоезично общество с активен обмен на информация се нуждае от бързо превеждане на огромен обем информация. Преводаческата дейност, особено за специални цели, е компютърно подпомогната от различни електронни езикови ресурси – речници, справочници, бази данни, преводаческа памет, едноезични и паралелни корпуси, специализирани програми. В приложенията са посочени някои надеждни линкове за такива онлайн ресурси.

Важно за качествено и интердисциплинарно обучение на студентите е запознаването им с електронни езикови ресурси на родния и изучавани чужди езици, създаването на умения за ползването им, решаването на практически задачи чрез тях.

Чрез конкретни задачи за сравняване на резултатите от превода на един текст от различните системи за машинен превод, анализ на неточностите, редактиране на резултата, т.е. благодарение на „учене чрез правене“ студентите търсят възможни отговори на въпроса „Да се доверяваме ли на машинния превод?“ Отговорът е положителен, ако отчитаме целите на превода, жанра, тематиката, спецификата на текста, надеждността на системите за машинен превод. Запознаването с основни електронни езикови ресурси и развиване на умения за работа с тях, с необходимата критична оценка и редакторска намеса на различни етапи от превода биха повишили качеството и ефективността на специализираните преводи в различни тематични сфери.

Засиленото използване на ИКТ (информационни и комуникационни технологии) в епохата на глобализиращо се и интернационализиращо се образование налага познаването на съществуващите електронни езикови ресурси, изисква умения за ползването им, решаването на практически задачи чрез тях от студенти и специалисти, работещи със специализирани текстове. Тези компетентности са полезни за бъдещата им професионалната дейност като изследователи, преподаватели, преводачи, редактори, коректори или в други професии, за повишаването на мотивацията на студентите и качеството на тяхното интердисциплинарно обучение.

Съвременната преводаческа дейност, особено за специални цели, е компютърно подпомогната от различни електронни езикови ресурси – речници, справочници, бази данни, преводаческа памет, онлайн ресурси, специализирани програми.

Обработката на естествения език обединява постиженията на компютърните технологии и математически формализми, подпомогнали развитието на нови научни дисциплини (за историята и развитието на тази научна област вж. [1: iv]). Усъвършенстването на обработката на езика се развива в две насоки: 1) подходи, основани върху правила (rule-based systems), като се използват постиженията на структурната лингвистика, предимно генеративна и депendentна граматика, и 2) емпирични подходи, основани върху статистически, вероятностни методи и обработка на данни (data-driven), които могат да бъдат прочетени от машините. Тези две тенденции предопределят и разработваните компютърни модели, модули, средства, системи.

Обработката на естествен език се гради върху тясната връзка между теоретични и приложни изследвания, което е ярка проява на интердисциплинарността в науката. За създаването на електронни езикови ресурси се комбинират познания от:

- ◆ когнитивните науки (лингвистически адекватно моделиране на функционирането на човешкия език);
- ◆ генеративната лингвистика (трансфер на знания и модели от теорията към приложения и средства за обработка на езика);
- ◆ изкуствен интелект (създаване на компютърни системи, симулиращи определени аспекти на когнитивното поведение на човека, включително овладяването на езика, процеса на комуникация);
- ◆ компютърна лингвистика.

Именно последната дисциплина исторически е свързана с машинния превод, а сега е пресечна зона между 3 области: 1) лингвистика, като компютърът се ползва като средство за тестване на различни лингвистични теории; 2) компютърни науки, където въз основа на взаимодействие между естествени и формални езици се работи върху разпознаване на езици, синтактичен анализ, структуриране на данни и т.н.; 3) изучаване на литературата с обработка на огромни корпуси от литературни текстове, създаване на хипертекстове, установяване на авторството на анонимни текстове и др.

Обработката на естествен език е тясно свързана с високите информационни технологии, като техният синтез се прилага в автоматичния превод, обучението, библиотечното дело, различни търсачки в интернет, телефонното, финансово и бизнес обслужване, туризма, електронната търговия и други области.

Съвременното ни общество е повлияно от глобализацията, предполага ползването на информация на различни езици. Въпреки многоезичието на голяма част от населението, неоченима помощ са компютърните средства за обработка на езика.

От мечтите и първите опити за автоматизиран превод през 40-те години на XX в. до съвременните мигновени системи за машинен превод с високо качество има огромен напредък в тези технологии. Обзор за историята и съвременното състояние на машинния, автоматичния или машинно подпомогнатия превод (термините се ползват като синонимни) е представен в [2, 3].

Речниците в електронен формат са необходими за професионалната дейност на преводача или ползвателя на специализирани текстове, както и в ролята им като важен модул, „захранващ“ машинния превод. Сред предимствата на електронните речници пред традиционните хартиени са постоянният, бърз достъп и търсене в тях, възможността за попълване, актуализиране, превръщането им в мултимедийни продукти чрез добавяне на изображения, озвучаване, илюстративни примери. Сред опасностите са постоянната поддръжка на сайтовете, където са базирани, надеждността

на поместената информация в отворения тип източници, където потребители могат да добавят данни. В Приложение 1 се съдържа ограничен списък на надеждни речници в електронен формат, които биха били полезни при четенето на специализиран текст. Те са или утвърдени академични речници, представени в електронен вид, или по-нови речници, енциклопедии, специализирани терминологични речници.

Обемните терминологични речници, големи корпуси от преведени текстове, които се използват като тренировъчни, подобряват многократно качеството на съвременните системи за машинен превод. Затова другият важен ресурс, необходим за създаването на добри системи за машинен превод, са корпусите на даден език – едноезични и паралелни – виж Приложение 2. Както термини от речниците, така и готово преведени фрази, клишетата се натрупват от потребителя в т.нар. преводаческа памет в зависимост от личните занимания и предпочитания.

Днес има огромно разнообразие от системи за автоматичен или машинен превод, или машинно подпомогнат превод – някои от тях са представени в Приложение 3. Ние дори можем да не осъзнаваме, че те са автоматично включени в нашите търсачки, където получаваме резултати на няколко езика, превод на сайтове, съобщения в социалните мрежи и др. Когато съзнателно прибъгваме до използването на системи за превод и очакваме значителна точност на резултата, е добре да направим редакция на изходния текст и пост-редакция на получения преводен текст. За да се подобри резултатът от автоматичен превод, изреченията трябва да бъдат къси, да са граматически правилни, да се избягват многозначни думи и сложни граматични конструкции. Качеството на превода се подобрява, ако се подават за превод не отделни думи или изречения, а свързан текст, поне един абзац, за да се осигури необходимия контекст и се сравни със запаметени готови фрази. В специализирани текстове е препоръчително да се използват утвърдени термини и понятия, които присъстват в речниците или са предварително въведени в речника на системата за машинен превод. Важна е и прагматичната информация, която се формира от знанията ни за света, но и от конкретния текст и ситуация. Така например въпросът на улицата *Имате ли часовник?* означава *Колко е часът?* Помните ли шеговития отговор „Да, имам.“? ;-) Важно е и спазването на установените фрази от речевия етикет: *Въ не скажете, где ближайшая станция метро?* В руския език въпросително-отрицателното изречение се смята за по-любезно. В българския език *Вие няма ли да ми кажете, къде е най-близката станция на метрото?* звучи емоционално, с упрек, нервно вместо обичайното *Бихте ли ми казали...? Може ли да ми кажете....? Извинете, къде е ... ?*, защото за българския речев етикет са типични положителните въпросителни изречения. Или например надписите на врата следват възприети клишетата: *От себя. / К себе. – Бутни. / Дръпни.* Табелка с предупреждение има определена формула: *Caution. Wet Floor. – Осторожно! Скользяй пол. / Влажный пол. – Внимание! Хлъзгав под.* за разлика от буквалния превод в друг свободен текст – *мокър под*. За усъвършенстването на автоматичния превод подобни фрази се включват допълнително.

Използването на системи за машинен превод днес е неизбежна необходимост и улеснение, но трябва да сме запознати с възможните грешки, неточности, пропуски. Затова в свободно избираема дисциплина по „Приложна лингвистика (работа с електронни езикови ресурси)“ (СУ, ФСлФ – [4]) към темата, свързана с машинен превод, се включва учебна задача да се сравнят резултатите от няколко машинни преводачки (за руски и български език с превод в двете посоки). Един и същ текст е преведен от няколко системи за машинен превод, като студентите отбелязват различните типове грешки, коментират ги като експерти-лингвисти, предлагат свои варианти за превод, изтъкват предимствата и недостатъците на различните системи за машинен превод.

Електронните ресурси за различни езици – речници, корпуси, програми за обработка на езика и за редактиране и др. са необходимост или самостоятелно, или като основа за машинен превод.

Така студентите търсят възможни отговори на въпроса „**Да се доверяваме ли на машинния превод?**“ чрез конкретна задача за „учене чрез правене“. Има доста фактори, от които зависи отговорът.

С какви цели превеждаме? За лично ползване, за ориентиране в темата, за получаване на информация – можем да се доверим на електронните преводачки, но за официални документи не е препоръчително.

Какви „преводачки“ използваме? Професионални платени програми за сериозни теми и документи, или безплатни общодостъпни за ежедневно, битово, научно-популярно ползване. Необходима ли е допълнителна обработка на текста? Да, редактиране на изходния текст, пост-редактиране на преведения текст. Нужно ли е да знаем добре езиците, от и на които превеждаме? Зависи от целите на превода. За официално използване на текста трябва да имаме много високо ниво на владение на езика, да сме по-добри от машината, за да редактираме допуснатите неточности. Ако просто искаме да получим представа от текст на език, който не владеем изобщо, такава компетенция не е нужна. Дали да ползваме машинен превод зависи и от сферата на употреба на езика – официално-делово общуване за професионалисти с професионални програми (от типа на Trados, Prompt, SkyCode и др.) или за лично ползване с общодостъпни свободни програми (като Google Translate, Yandex, Bing и др.), които също са достатъчно надеждни. Наличните програми и ресурси могат да бъдат особено полезни при превод на специализиран текст за натрупване на терминология, преводаческа памет, в създаването на многоезични ресурси.

Съвършеният машинен превод, макар да е една неосъществима мечта, с усилията на много специалисти вече има доста добри резултати, доближаващи се до дейността на най-добрите преводачи. Автоматичният превод пести време и усилия на преводача преди окончателната редакция на текста. По пътя към нейното осъществяване са развити много полезни приложения за обработка на езика. Общодостъпността, скоростта, високото качество на машинния превод улесняват нас като потребители, стига да преценяваме целите, типа текст, адресата на машинния превод.

ЛИТЕРАТУРА:

[1] Dale R., H.Moisl, H.Somers. Handbook of Natural Language Processing. “Marcel Dekker Inc.”, New York-Basel, 2000.

[2] <https://prevodachi.eu/mashinen-prevod-istoriya-v-1-tchast/>

[3] <HTTPS://WWW.LANGUST.RU/ETC/HISTORY.SHTML>

[4] <https://www.slav.uni-sofia.bg/subjects>,

https://www.slav.uni-sofia.bg/images/stories /documents/sid/Krasimira_Petrova.pdf

Приложение 1. Речници и Електронни ресурси (списъкът е отворен)

1.1. Е-речници (англо-български):

- SA Dictionary - <https://sa-dictionary.software.informer.com/>
- Dictionary.NET - <http://www.dictionary.net/>
- Eurodict - <http://www.eurodict.koralsoft.com/> (7 езика)
- Oxford Dictionary - <https://www.oed.com/> <https://www.oxfordlearnersdictionaries.com/>
- Lingvosoft - <http://www.lingvosoft.com/> (50 езика)

- Glosbe - <https://glosbe.com/en/bg>
- Pons - <https://en.pons.com/translate>

1.2. Е-речници (руски, руско-български)

- Бернштейн С.Д. Болгарско-русский словарь. (разл.изд., 3-е изд. 1986). Москва.
<https://djvu.online/file/kkuXSY22B39hU>
- Тимонина Б.В., Мартинова Г.А. Болгарско-русский и русско-болгарский словарь. 2009, Дрофа, ISBN: 978-5-9576-0524-9
<https://djvu.online/file/vCvUZKyrKM4Sk>
- Болгарско-русский словарь <https://gufo.me/dict/bgru>
- Русско-болгарский словарь <https://gufo.me/dict/rubg>
- Ожегов С., Н.Ю. Шведова. Толковый словарь русского языка. М.: ИТИ Технологии. 2006 и др. <http://www.ozhegov.org/>
- Ушаков Д.Н. – Толковый словарь русского языка: В 3 т., М., Мир кн., 2001.
<http://ushakovdictionary.ru/>
- <https://kartaslov.ru/>
- Словари и энциклопедии -Бизнес, экономика и юриспруденция, Техника и технологии <https://gufo.me/>
- Словари <https://www.uchiyaziki.ru/> -
- Болгарский словарь. Polyglot 7. Многоязычный.
http://bulgarianavsegda.com/bolgarskii_slovar
- Мультимедийный лингвострановедческий словарь «Россия». На сайта на ИРЕ „А.С. Пушкин”. <https://ls.pushkininstitute.ru/lsslovar/index.php?title>
- Справочно-информационный портал ГРАМОТА.РУ – русский язык для всех
<http://gramota.ru/slovari/>

1.3. Специализирани Е-речници

- Словари и энциклопедии на Академике. – разделы Техника, технологии, механизмы, Экономика, юриспруденция - <https://dic.academic.ru/>
- Англо-русский дорожно-транспортный словарь https://traffic_en_ru.academic.ru/
- Технический железнодорожный словарь. Васильев Н.Н. и др.(ред).
<https://dic.academic.ru/contents.nsf/railway/>

Приложение 2. Корпуси на българския и руския език (списъкът е отворен)

- Национальный корпус русского языка <https://ruscorpora.ru/>
- Параллельный корпус (болгарский) <https://ruscorpora.ru/new/search-para.html?lang=bul>
- Български национален корпус <http://dcl.bas.bg/bulnc/>
- Успореден корпус на руски и български текстове <http://rbcorpus.com/contact.php>

Приложение 3. Е-преводачи (списъкът е отворен)

- Google Translate - <https://translate.google.com/>
- Bing Microsoft Translator - <https://www.bing.com/translator>
- iTranslate - <https://itranslate.com/webapp>
- DeepL - <https://www.deepl.com/>
- Reverso - <https://www.reverso.net/text-translation>
- WebTrance - <http://webtrance.skycode.com/?setlang=en>
- Яндекс Переводчик <https://translate.yandex.ru/>
- TextGrabber - <https://pdf.abbyy.com/products/mobile/text-grabber/en/>
- KantanMT - <https://www.kantanai.io/>

- Systran Translate - <https://www.systran.net/en/translate/>
- Canopy Speak - <https://www.withcanopy.com/canopy-speak/>
- Lingua Custodia - <https://www.linguacustodia.finance/en/home/>
- SDL Trados Studio - <https://www.trados.com/products/trados-studio/>
- Smartcat - <https://www.smartcat.com/>
- <https://translate.academic.ru/>
- www.translate.ru
- <https://www.translate.eu/>
- www.bing.com
- <https://bg.glosbe.com/>
- <https://translate.yandex.ru/>
- www.online-translator.com/
- www.translator.ru/
- translate.google.com/
- translate.google.bg/

SHOULD WE TRUST MACHINE TRANSLATION? (ELECTRONIC LANGUAGE RESOURCES SUPPORTING SPECIALIZED TRANSLATION)

Krasimira Petrova

krapirova@slav.uni-sofia.bg

*Sofia University „St. Kliment Ohridski“
Sofia, 15 Tsar Osvoboditel Blvd
THE REPUBLIC OF BULGARIA*

Key words: *specialized translation, machine translation, electronic language resources*

Abstract: *Today's globalizing, multilingual society with an active exchange of information needs rapid translation of a huge volume of information. Translation activity, especially for special purposes, is computer-assisted by various electronic language resources – dictionaries, reference books, databases, translation memory, monolingual and parallel corpora, specialized programs. Some reliable links to such online resources are listed in the appendices.*

It is important for the quality and interdisciplinary education of students to familiarize them with electronic language resources of their native and studied foreign languages, to create skills for their use, to solve practical tasks through them.

Through concrete tasks of comparing the results of the translation of a text from different machine translation systems, analysis of inaccuracies, editing of the result, i.e. thanks to „learning by doing“, students search for possible answers to the question „Should we trust the machine translation?“ The answer is positive if we take into account the goals of the translation, the genre, the subject matter, the specificity of the text, the choice of reliable machine translation systems. Acquaintance with basic electronic language resources and developing skills for working with them, with necessary critical evaluation and editorial intervention at various stages of the translation, would increase the quality and efficiency of specialized translations in various thematic areas.