

Михнова Е. Д.,

кандидат технических наук, преподаватель Харьковского национального технического университета сельского хозяйства Харьковского гуманитарного университета «Народная украинская академия»

ОБЛАЧНЫЕ ТЕХНОЛОГИИ В СИСТЕМАХ НА ОСНОВЕ ПАМЯТИ ПЕРЕВОДОВ

Використання хмарних ресурсів для зберігання пам'яті перекладів, термінологічних словників, вихідних та цільових файлів перекладу, посібників зі стилю та довідкових матеріалів має важливе значення для організації ефективного обміну даними між перекладачами, редакторами, коректорами, менеджерами та іншими членами команди, що працюють над крупними перекладацькими проектами. У статті розглянуто ключові проблеми застосування цих технологій з точки зору кінцевого користувача. Надано аргументи на користь необхідності навчання цим технологіям при підготовці спеціалістів в галузі традуктології (перекладознавства).

Cloud resource usage for storing translation memories, term bases, initial reference files and target translations, style guides and additional materials on the topic of translation is very important to ensure efficient data exchange between translators, editors, proofreaders, managers and other stuff, working on large translation projects. In this article, key issues faced by the end users are observed in relation to these technologies application. Argumentation is provided concerning the necessity of teaching the technologies in translation studies institutions.

На сегодняшний день компьютерная лингвистика является развивающимся направлением научных исследований. Несмотря на значительный прогресс в этой области, связанный, прежде всего, с организацией баз знаний и формализацией грамматических правил, многие вопросы машинного перевода все еще остаются открытыми. И хотя существует большой выбор программных систем автоматического машинного перевода, большинство профессиональных переводчиков в своей повседневной работе используют, так называемые, системы автоматизированного перевода, САТ-инструменты (Computer Aided Translation или Computer Assisted Translation tools) [1]. В отличие от автоматического перевода, эти системы призваны помочь переводчику аккумулировать его собствен-

ную базу переводов, а не свести до автоматизма сам процесс перевода, поскольку в последнем случае качественного результата достичь невозможно.

Отсутствие приемлемого качества перевода при использовании автоматических средств демонстрируют традиционные системы Google Translate, Promt и другие. Основное назначение подобного класса программных продуктов — дать общее представление о содержании текста пользователю, который не знает исходного языка, чтобы не искать перевод каждого слова в словаре. Эти системы неприемлемы для профессионального перевода, когда в результате нужно получить действительно качественный текст, который будет читаться не просто носителем, но еще и экспертом предметной области переводимого материала.

Сложность перевода зачастую состоит именно в специфике исходного текста и его узкой ориентации на предметную область, как например, юриспруденция, телекоммуникации, медицина или тяжелая промышленность. Многозначность терминов исходного языка делает перевод полностью нечитаемым для специалиста той или иной области. А так называемую лингвистами проблему «непереводимости» совершенно невозможно решить машинными средствами, даже заложив эвристические правила передачи, не говоря о грамматических особенностях каждого языка [2].

Функциональные возможности автоматизированных систем, как правило, сводятся к накоплению базы переводов (англ., Translation Memory, ТМ) и сравнению отдельных сегментов, слов или фраз, при очередном их появлении в тексте с ранее накопленной ТМ. Каждая запись ТМ представляет собой предложение (или его часть) параллельных двуязычных текстов. Такая база данных позволяет осуществить быстрый поиск по содержимому. При обработке вновь поступившего на перевод текста система поочередно сравнивает каждое его предложение с имеющимися парами перевода, отображая процентное соответствие и предлагая подстановку перевода при высоком проценте совпадения. Переводчику придется проработать сегменты без соответствия с ТМ, отредактировать частично совпадающие и подтвердить полностью соответствующие [3].

Во многих современных системах на основе памяти переводов имеется возможность использования внешней, так называемой, терминологической базы (англ., Term Base, ТВ). Наряду с возможностью подключения нескольких файлов ТМ, в рамках работы над одним проектом можно использовать несколько терминологических баз одновременно. На рисунке 1 показаны типичные компоненты интерфейса программы на основе памяти переводов с подключенными терминологическими базами.

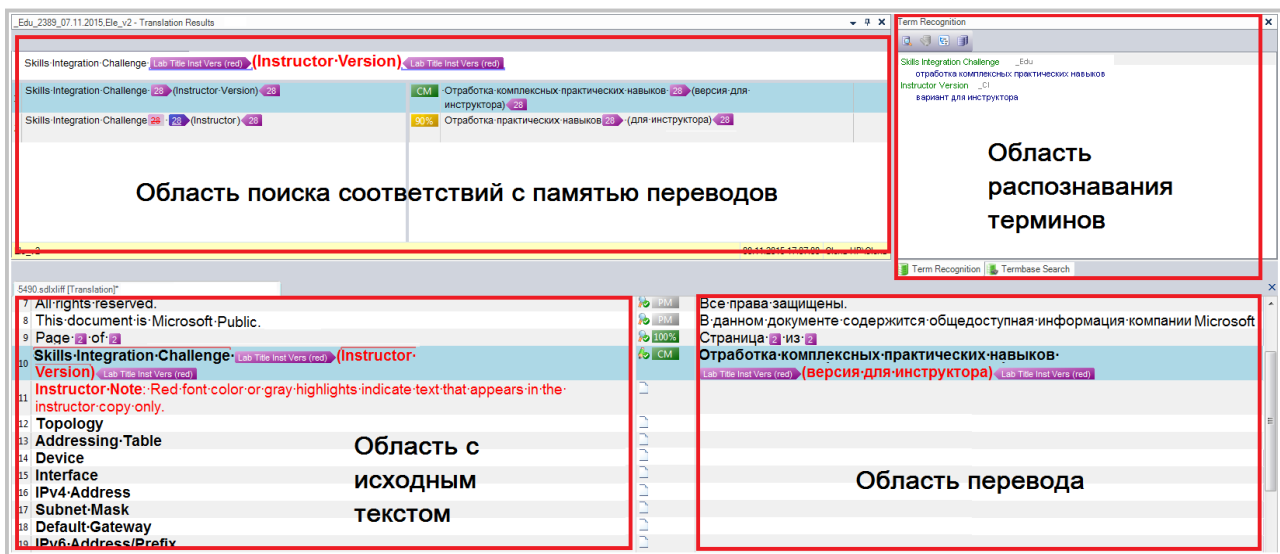


Рисунок 1 – Типичные компоненты интерфейса программы на основе памяти переводов с подключенными терминологическими базами

Таким образом, ТМ получается присущ принцип самообучения. Причем ТМ целесообразно формировать для каждой отдельной предметной области. Кроме того, переводчик получает мощный инструмент терминологического поиска, который, в отличие от обычного словаря с множеством значений одного слова, имеет узкую направленность на контекст и учитывает стилистическую специфику каждого лица, осуществляющего перевод. Большинство систем на базе ТМ используют алгоритм нечеткого соответствия (fuzzy match), позволяющий находить фразы даже с небольшим процентом совпадений, что иногда оказывается очень полезным для соблюдения однотипности терминологии.

К системам автоматизированного перевода на базе ТМ относят: SDL Trados Studio, Atril Deja Vu X, MemoQ, Star Transit, WordFast, Systran, Lionbridge, Across, AnyMem, MateCat, MemSource, XTM, OmegaT, ABBYY

SmartCAT, Alchemy Catalyst и многие другие. Безусловным лидером среди CAT-инструментов на сегодняшний день является SDL Trados.

Начиная с версии 2007 года, основным изменением, ощутимым для переводчика, стало лишь отступление от интеграции с текстовым процессором MS Word, что, по большому счету, лишь усложнило адаптацию к программе ее «старых» пользователей Trados. Принцип работы стал более походить на средство локализации SDL Passolo, когда переводимый программный интерфейс «загружается» в программу локализации, где и осуществляется автоматизированный перевод с учетом накопленной ТМ. Также версия SDL Trados Studio от 2015 года характеризуется добавлением поддерживаемых форматов файлов, однако принципиально нового подхода к аккумуляции или извлечению данных из ТМ за 8 лет так и не было предложено.

Преимущества применения CAT-инструментов сначала могут показаться не очевидными, но по мере наполнения ТМ и при условии однотипного объемного исходного текста, подсказки системы и подстановка готовых сегментов перевода будут наблюдаться все чаще, а варианты перевода будут более точными. Не говоря о текстах с большим количеством повторов, где эффективность CAT очевидна, работая над крупным проектом в условиях временных ограничений, группе переводчиков приходится согласовывать терминологию, что также можно автоматизировать с помощью CAT и свести возможные рассогласования к минимуму.

Для эффективного обмена данными по проекту в последнее время все чаще стали использоваться защищенные облачные хранилища. Группе, работающей в рамках одного проекта, выдается ссылка на защищенный паролем общий ресурс, доступ к которому возможен после разрешения администратора проекта, в роли которого обычно выступает руководитель проекта либо ответственный представитель отдела кадров. Некоторые из существующих CAT-средств основаны на облачных технологиях и предполагают наполнение базы переводов в режиме онлайн. Иными словами, имитируется работа всей группы с серверной базой переводов.

Таким образом, можно сделать вывод о целесообразности применения данного программного комплекса на основе ТМ для крупных проектов с объемами порядка десятка и более авторских листов, что приблизительно составляет объем средней книги. Использование систем на основе памяти переводов для менее емких проектов, что на сегодняшний день практикуется во многих бюро переводов, лишь усложняет работу переводчиков. Безусловно, соответствий будет найдено значительно больше, и вероятность повторов сильно возрастает, но процесс такого перевода можно сравнить со стрельбой из пушки по воробьям. Иными словами, новому небольшому проекту навязывается стилистика старых, и сохраняя однотипность в рамках агентства, теряется индивидуальность каждого конкретного проекта, который при условии высокой квалификации переводчика, мог быть передан более точно на целевой язык.

Подводя итог, следует отметить, что вышеописанные технологии широко применяются для крупных переводческих проектов в развитых западных странах, что экономит время переводчика и деньги заказчика, тогда как на территории стран СНГ показатель их использования не превышает 30 % от всех переводческих компаний. По причине дороговизны лицензионного ПО, студенты отечественных ВУЗов лишены возможности изучения передовых технологий, востребованных на международном рынке. Хотя существует целый ряд бесплатных онлайн версий САТ-средств, как и стационарных инструментов САТ с открытым исходным кодом, в Европе и США все же чаще используются именно лицензионные и более мощные системы, такие как SDL Trados Studio, Atril DeJa Vu X, MemoQ.

Несмотря на то, что в некоторых ВУЗах, обучающих переводческой специальности, внимание все же уделяется одному (или нескольким), как правило, бесплатному программному продукту подобного класса, учитывая их разнообразие (которое сейчас уже достигло порядка полусотни), актуальной проблемой становится освоение нескольких наиболее востребованных инструментов, упомянутых выше. Конечно, все они построены по одному принципу и, овладев одной программой, легко освоить другую. Однако, начиная проект в условиях

крайне сжатых сроков, у членов группы нет возможности тратить время на установку и ознакомление с новым для них ПО.

Список использованной литературы

1. Sin-wai Ch. The Routledge encyclopedia of translation technology / Ch. Sin-wai. – Abingdon, Oxon ; New York : Routledge, 2015. – 718 p.

2. Gambier Y. Handbook of Translation Studies / Y. Gambier, L. Doorslaer. – Amsterdam ; Philadelphia : John Benjamins Publishing Company, 2010. – Volume 1. – 458 p.

3. In-context exact (ICE) matching : US patent No. US 7,983,896 B2 / Ross R., Gillespie K. ; assignee SDL Language Technology. – Appl. No. 11/071,706 ; filed 03.03.05 ; published 19.07.11. – 19 p.