

СИСТЕМА ПЕРЕВОДА «TRANSLATION MEMORY» КАК ОСНОВА ДЛЯ ПЕРЕВОДА ФРАЗЕОЛОГИЗМОВ

С использованием первых компьютеров возникла идея их использования для автоматического перевода текстов. Первый такой перевод был проведен в США в 1954 г. В этом «Джорджтаунском эксперименте» впервые компьютер перевел русские технические тексты на английский язык.

С тех пор создано большое число компьютерных программ, позволяющих переводить тексты с разных языков. Тем не менее, такие программы не позволяют решить все проблемы перевода. Это, в определенной степени, связано как с заметным отставанием в теории прикладной (компьютерной) лингвистики, так и в отсутствии практических данных по сопоставительному описанию различных пар естественных языков [1, с. 43–61; 2, с. 195–199].

Большие надежды в этом плане возлагаются на системы машинного перевода, работающие совершенно по новому принципу [3–5]. В отличие от известных систем машинного перевода, работающих на лингвистических правилах [1; 2; 4], эти новые системы используют корпуса ранее переведенных текстов. Впервые идея использовать для перевода текстов определенные фрагменты уже переведенных текстов была высказана в 1975 г. русским ученым Г.Г. Белоноговым и опубликована в предисловии к книге Д.А. Жукова «Мы — переводчики» [6]. Такой принцип перевода был назван «фразеологическим машинным переводом». Позже он был реализован в виде двух известных систем машинного перевода ERTRANS и RETRANS [3].

В более развернутом виде идея использования для перевода уже сохраненных корпусов переведенных текстов была предложена в 1981 г. японским ученым Макато Нагао. Несколько позже он опубликовал статью [7], где свой подход к автоматическому переводу назвал «машинный перевод на основе заложенных примеров или машинный перевод по аналоговому принципу» и обосновал его следующим образом [7, с. 178]: «Человек не переводит простое предложение с помощью глубокого лингвистического анализа. Скорее, человек при переводе сперва членит содержание предложения на некоторые отдельные фразы..., затем переводит эти фразы на другой язык и, наконец, надлежащим образом складывает эти фрагменты перевода в одно длинное предложение. Такой перевод каждого фрагмента предложения будет переводиться по принципу аналогового перевода, со ссылкой на надлежащие примеры».

Со временем для создания таких компьютерных систем перевода стали создаваться специальные большие базы данных, содержащие переводы двух- трехсловных фраз исходного языка и их переводов на другой язык. Такой перевод стал называться статистическим или накопительным переводом (Translation Memory).

По этому принципу сейчас работает целая серия переводческих программ [3, с. 80–81]: Omega T, Studio Basic, Translation, Aid Trans, Open Language Tools и другие. Но наиболее известной программой перевода, работающей по принципу Translation Memory, является программа TRADOS [3, с. 88–131].

В наши дни идея использовать в процессе перевода уже переведенные кем-то и когда-то тексты («параллельные тексты») реализуется в системах нового типа — системах памяти переводов (Memory System). Впервые идею о необходимости создания переводческих систем подобного типа высказал в 1978 г. Питер Эртерн. Суть такого подхода была им сформулирована следующим образом [8, с. 318]: «В действительности, должна быть возможна разработка такой программы, которая бы позволила текстовому процессору «запоминать» любую часть нового текста, введенного в него и уже переведенного, а затем извлекать ее вместе с уже выполненным переводом».

По существу системы этого типа переводят не отдельные фрагменты текста, а целые предложения. Отсюда и грамматическая правильность всего предложения. В системах, работающих по принципу накопительного перевода, переведенные фрагменты могли быть грамматически несогласованными.

Как подчеркивается в работе [5, с. 23], главное различие между системами Translation Memory и Memory System заключается в том, что вторая из них является помощником переводчика, где основанная ответственность за принятие решения лежит на пользователе, который сам обращается к некоторой базе примеров. А система Translation Memory выполняет перевод автоматически, обращаясь к созданной ранее большой базе данных.

Сейчас специалисты по переводу работают над созданием систем перевода текстов, объединяющих преимущества двух описанных выше систем, использующих параллельные тексты [3, с. 131–132; 5, с. 9–26]. Однако применение параллельных текстов требует от пар переводимых текстов выполнения двух основных условий:

1. Они должны быть максимально сравнимыми: совпадать по жанру («новости»), типу («политические»), содержанию («выборы президента») и форме («отчеты», «письма» и т.д.).

2. Такие тексты должны быть «выравнены», т.е. они должны содержать соотносимые логически связанные фрагменты, имеющие смысловое начало и конец (разделы, параграфы, главы, абзацы, предложения).

Если первое условие вполне выполнимо, то «выравнивание» переводимых текстов весьма трудоемко и требует применения специальных компьютерных программ [9; 10].

К сожалению, пока нет научных работ, посвященных использованию параллельных текстов для автоматического перевода фразеологизмов.

Фразеологизмом (фразеологической единицей) будем считать «общее название семантически связанных сочетаний слов и предложений, ко-

торые, в отличие от сходных с ними по форме синтаксических структур, не производятся в соответствии с общими закономерностями выбора и комбинации слов при организации высказывания, а воспроизводятся в речи в фиксированном соотношении семантической структуры и определенного лексико-грамматического состава» [11, с. 559]. К числу фразеологизмов лингвисты относят идиомы, метафоры, устойчивые словосочетания, терминологические сочетания и др.

Специфика предлагаемого метода использования параллельных текстов для перевода фразеологизмов связана с тем, что их переводы берутся не из некоторой базы данных заранее переведенных словосочетаний, а из конкретного перевода некоторого исходного предложения при условии, что исходный и переведенный тексты, взятые из параллельного корпуса, «выравнены» или «стыкуются» по предложениям [9; 10, с. 180–181].

Основная идея этого метода основывается на следующих предположениях:

1. В каждом фразеологизме предложения исходного переводимого текста можно выделить его опорное («ядерное», «узловое») слово, являющееся основой фразеологизма.

2. В рамках взятого из параллельного корпуса текстов перевода исходного предложения можно выделить фразеологизм, опорным словом которого является один из переводных эквивалентов опорного слова исходного переводимого фразеологизма или наиболее значимое с лингвистической точки зрения слово этого фразеологизма.

3. В базе данных предлагаемой системы перевода для текстов достаточно узкой предметной области:

а) заданы структурные схемы (хотя бы на уровне классов слов) исходного и возможного переведенного фразеологизмов;

б) заданы выявленные на предварительных этапах лингвистических исследований текстовые признаки, определяющие границы фразеологизмов в рамках исходного предложения и его перевода.

Так, по отношению к переводу на узбекский язык русских метафор, эти наиболее общие положения были сформулированы следующим образом [13, с. 4]:

1) метафора — это не только лексико-семантическая категория, но и формальная, синтаксическая;

2) метафоры, созданные не макроконтекстом, а микроконтекстом, имеют четкие формальные параметры;

3) русским формализованным метафорическим структурам соответствует система коррелятов в узбекском языке;

4) на семантическом уровне формализованные структуры способны организовываться в лексико-семантические группы.

С опорой на сформулированные выше условия в Минском государственном лингвистическом университете были созданы алгоритм и компьютерная программа для перевода англоязычных терминологических словосочетаний проблемной области «Вычислительная техника» на рус-

ский язык [12]. Для построения алгоритма перевода таких терминологических словосочетаний были изучены 600 таких словосочетаний, взятых из английской части корпуса параллельных текстов по вычислительной технике и их переводы, выделенные из русской части данного корпуса и из современных специальных словарей [14].

На следующем шаге анализа все выделенные английские терминологические словосочетаний были разделены на группы, состоящие из 2-х, 3-х, 4-х слов. В каждом из них было определено опорное слово. Далее в русских переводных эквивалентах этих терминологических словосочетаний также выделены опорные слова. В итоге была построена база данных переводных соответствий. Фрагмент ее приведен в табл. 1. Апостроф здесь обозначает место опорного слова.

Созданная на основе такого алгоритма компьютерная программа в более 90% случаев нашла в текстах по вычислительной технике верные переводные эквиваленты английских терминологических словосочетаний.

Таблица 1

Структурная модель английского ТС	Структурная модель русского переводного эквивалента	Пример
$N1 + N'$	$A + N'$	Background image — фоновое изображение
$N1 + N2 + N'$	$N' + N1 + N2$	Message queuing service — служба очереди сообщений
$A + N1 + N2 + N'$	$A + N' + N1 + N2$	Trivial file transfer protocol — простейший протокол передачи файла

При переводе метафор и идиом в большинстве случаев в во фразеологизме переведенного предложения не будет опорного слова, являющегося одним из переводных эквивалентов опорного слова исходного переводимого фразеологизма. Тогда в базе данных, подобной приведенной выше в табл.1, необходимо выделить в переводе соответствующего фразеологизма наиболее значимое для него слово. Например, при переводе по рассматриваемому методу английских метафор на русский язык база данных переводных соответствий может выглядеть следующим образом (см. табл.2):

Таблица 2

Структурная модель английской метафоры	Структурная модель русского переводного эквивалента	Пример
$V' + art + N + prep.$	$V' + prep + N$	Close the door on — покончить с прошлым
$V' + pron + N$	$V' + N$	Show one's hand — раскрывать карту
$V' + prep. + art + A + prep. + N$	$V' + prep + N + N$	Lose in the most of time — затеряться во тьме веков
$V' + pron + prep + A + N$	$V' + prep + A + N$	See one through darkest hours — Справляться с тяжелыми испытаниями

ЛИТЕРАТУРА

1. Марчук, Ю.Н. Компьютерная лингвистика: учеб. пособие. — М., 2007.
2. Нелюбин, Л.Л. Введение в технику перевода: учеб. пособие. — М., 2009.
3. Кво, Ч.К. Технологии перевода. — М., 2008.
4. Соловьева, А. Профессиональный перевод с помощью компьютера. — М., 2008.
5. Somers, Harold, Fernandez, Diaz Gabriela. Translation Memory versus Example-based MT: What is the Difference? // International Journal of Translation. Vol. 16. № 2. July — Dec., 2004. — P. 4–33.
6. Жуков, Д.А. Мы — переводчики. — М., 1975.
7. Nagao, M. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle // Artificial and Human Intellegence. — Amsterdam. 1984. — P. 173–180.
8. Arthern, P. Aids unlimited: The scope for machine aids in a large organization // Aslib Proceeding. — 1981. — № 33. — P. 309–319.
9. Михайлов, М.Н. Стыковка параллельных текстов в автоматическом режиме: иллюзии и перспективы // НТИ. Сер. 2. Информационные процессы и системы. — 2003. — № 10. — С. 18–26.
10. Серебряная, Л.В. Статистическая система машинного перевода // Информационные системы и технологии (IST'2009): междунар. науч. конф. — Минск, 2009. — С. 179–183.
11. Лингвистический энциклопедический словарь. — М., 1990. — 684 с.
12. Степанова, Д.В. Лингвистические аспекты перевода на русский язык английских терминологических словосочетаний с использованием корпуса параллельных текстов: дис. ... канд. филол. наук: 10.02.21. — Минск, 2006.
13. Кузин, О.А. Русская метафора и практика её перевода на узбекский язык: автореф. ... канд. наук: 10.02.01 — русский язык. — Ташкент, 1980.
14. Hajczuk, Roman, Zubow, Aleksander. English–Polish–Russian–Belarusian dictionary on informatics and computer technology. — Bialystok, 2003.