

УДК 81'42

Зубов А.В.

Минский государственный лингвистический университет

ИСПОЛЬЗОВАНИЕ ПАРАЛЛЕЛЬНОГО КОРПУСА ТЕКСТОВ ДЛЯ ПЕРЕВОДА НЕСВОБОДНЫХ СЛОВСОЧЕТАНИЙ

Zubov A.

Minsk State Linguistic University

APPLICATION OF PARALLEL CORPUS OF TEXTS FOR TRANSLATION OF BOUND WORD COMBINATION

Аннотация. В предлагаемой статье автором анализируются методы использования параллельных корпусов текстов для перевода иноязычных текстов. Показывается различие между системами переводов, работающими по принципам Translation Memory и Memory System. Автором предлагается новый подход к переводу несвободных словосочетаний (терминологических словосочетаний, идиом, метафор) с использованием параллельных текстов. Приводится один из вариантов нового подхода на примере перевода терминологических словосочетаний с английского языка на русский язык.

Ключевые слова: перевод, текст, слово, метод, параллельный, алгоритм, терминологический.

Abstract. In the present article we analyze the methods of application the parallel texts for translation of texts. The author fixes his attention on differences between systems Translation Memory and Memory System. Our new method directed to translate terminological word combinations, idioms and metaphors. The author gives the examples that demonstrate algorithm of translation of English terminological word combinations into Russian.

Key words: translation, text, word, method, parallel, algorithm, terminological.

В книге Л.Л. Нелюбина «Введение в технику перевода» [9, 43–61] подробно рассматриваются основные теории и модели перевода. Касаясь здесь теории машинного перевода, он, вслед за Ю.Н. Марчуком [7, 295–299], отмечает, что сейчас важно осознать и преодолеть определенное отставание как в теории прикладной (компьютерной) лингвистики, так и в практике сопоставительного описания переводимых соответствий в конкретных языковых парах. Практическую реализацию этих идей можно обнаружить в системах машинного перевода работающих совершенно по новому принципу [5; 12; 17]. В отличие от известных систем машинного перевода, работающих на лингвистических правилах [4; 7; 9], эти новые системы используют корпуса ранее переведенных текстов. Впервые идея использовать для перевода текстов определенные фрагменты уже переведенных текстов была высказана в 1975 году русским ученым Г.Г. Белоноговым и опубликована в предисловии к книге Д.А. Жукова «Мы – переводчики» [4]. Такой принцип перевода был назван «фразеологическим машинным переводом». Позже он был реализован в виде двух известных систем машинного перевода ERTRANS и RETRANS [2, 2].

В более развернутом виде идея использования для перевода уже сохраненных корпусов переведенных текстов была предложены в 1981 году японским ученым Макато Нагао. Несколько позже он опубликовал статью [16], где свой подход к автоматическому переводу назвал «машинный перевод на основе заложенных примеров или машинный перевод по аналоговому принципу» и обосновал его следующим образом [16, 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 большие специальные 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 [5, 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 [5, 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 году Питер Эртерн. Суть такого подхода была им сформулирована следующим образом [15, 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 [17, 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 [5, 131–132; 17, 9–26]. К сожалению, пока нет работ, посвященных использованию параллельных текстов для автоматического перевода несвободных словосочетаний. Рассуждая о возможностях использования параллельных текстов для перевода художественных произведений, Д.И. Добровольский выделяет следующие основные трудности таких текстов [3, 128]: 1) нестандартное употребление лексических единиц; 2) нерегулярные синтаксические конструкции; 3) излюбленные слова автора; 4) несвободные словосочетания; 5) идиомы и конвенциональные метафоры; 6) авторские метафоры; 7) культурно-специфичные элементы (названия, имена, титулы и др.). Если вспомнить основные трудности перевода научно-технических текстов, то в этот перечень можно добавить и трудности в переводе терминологических словосочетаний.

Переходя к рассмотрению возможностей использования параллельных текстов для перевода несвободных словосочетаний, будем считать такими словосочетаниями цепочки слов текста, смысл которых не определяется суммой значений, составляющих это словосочетание слов. В первую очередь отнесем к числу таких текстовых единиц *терминологические словосочетания, идиомы и метафоры*. Следующая специфика предлагаемого метода использования параллельных текстов для перевода несвободных словосочетаний связана с тем, что перевод словосочетания берется 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 [8; 11, 180–181].

Основная идея этого метода основывается на следующих предположениях: 1) в каждом несвободном словосочетании предложения исходного переводимого текста можно выделить его опорное («ядерное», «узловое») слово, являющееся основой словосочетания; 2) в рамках взятого из параллельного корпуса текстов перевода исходного предложения можно выделить несвободное словосочетание, опорным словом которого является один из переводных эквивалентов опорного слова исходного несвободного словосочетания или наиболее значимое с лингвистической точки зрения слово этого словосочетания; 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 несвободного словосочетания; б) заданы выявленные на предварительных этапах лингвистических исследований текстовые признаки, определяющие границы несвободных словосочетаний 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 [6, 4]: 1) метафора – это не только лексико-семантическая категория, но и формальная, синтаксическая; 2) метафоры, созданные не макроконтекстом, а микроконтекстом, имеют четкие формальные параме-

тры; 3) русским формализованным метафорическим структурам соответствует система коррелятов в ПЯ; 4) на семантическом уровне формализованные структуры способны организовываться в лексико-семантические группы (далее ЛСГ).

Продемонстрируем приведенные выше условия успешного перевода несвободных словосочетаний с помощью параллельного корпуса текстов на примере перевода с использованием параллельного корпуса текстов англоязычных терминологических словосочетаний проблемной области «Вычислительная техника» на русский язык [13;14]. 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.

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

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

Принципиальный алгоритм перевода терминологического словосочетания с опорой на эту базу данных приведен на рис. 2 [14, 204].

Созданная на основе такого алгоритма компьютерная программа в более 90% случаев нашла верные переводные эквиваленты английских терминологических словосочетаний. При переводе метафор и идиом в большинстве случаев в несвободном словосочетании переведенного предложения не будет

опорного слова, являющегося одним из переводных эквивалентов опорного слова исходного несвободного терминологического словосочетания. Тогда в базе данных, подобной приведенной выше в табл.1, необходимо выделить в переводе соответствующего словосочетания наиболее значимое для него слово. Например, при переводе по рассматриваемому методу английских метафор на русский язык может выглядеть следующим образом (табл.3):

Таблица 2

I	опознавание английского опорного термина во вновь пришедшем переводимом предложении и выделение его в тексте
II	опознавание английского ТС с найденным опорным термином в переводимом предложении и выделение его из текста
III	поиск выделенного переводимого словосочетания в английской части корпуса параллельных текстов и выделение предложения, в котором оно употреблено
IV	опознавание квазиосновы русского опорного термина словосочетания- эквивалента переводимого английского ТС в соответствующем предложении русской части корпуса параллельных текстов
V	нахождение переводного эквивалента английского ТС в соответствующем предложении русского параллельного текста и печать его на экране дисплея

Таблица 3

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 – за- теряться во тьме веков
V' + pron + prep + A + N	V' + prep + A + N	See one through darkest hours – Справляться с тяжелыми испытаниями

ЛИТЕРАТУРА:

1. Белоногов Г.Г., Зеленков Ю.Г., Новоселов А.П. и др. Системы фразеологического машинного перевода. Состояния и перспективы развития // НТИ. Сер. 2. Информационные процессы и системы. 1998. № 12. С. 14-21.
 2. Белоногов Г.Г., Хорошилов А.А., Хорошилов А.А. Автоматизация составления англо-русских двуязычных фразеологических словарей по массивам двуязычных текстов (биллингв) // НТИ. Сер. 2. Информационные процессы и системы. 2010. № 5. С. 1-8.
 3. Добровольский Д.О. Корпус параллельных текстов как инструмент анализа литературного перевода // Компьютерная лингвистика и интеллектуальные технологии. М.: Наука, 2003. С. 126-131.
 4. Жуков Д.А. Мы – переводчики. М.: Знание, 1975. 124 с.
 5. Кво Ч.К. Технологии перевода. М.: Академия, 2008. 252 с.
 6. Кузин О.А. Русская метафора и практика её перевода на узбекский язык: автореф. дис. ... канд. филол. наук. Ташкент, 1980. 20 с.
 7. Марчук Ю.Н. Компьютерная лингвистика. М.: Восток–Запад, 2007. 317 с.
 8. Михайлов М.Н. Стыковка параллельных текстов в автоматическом режиме: иллюзии и перспективы // НТИ. Сер. 2. Информационные процессы и системы. 2003. С. 18-26.
 9. Нелюбин Л.Л. Введение в технику перевода. М.: Наука, 2009. 208 с.
 10. Нелюбин Л.Л. Очерки по введению в языкознание. М., 2007. 199 с.
 11. Серебряная Л.В., Романюк Г.Э., Демидович В.С. и др. Статистическая система машинного перевода // Информационные системы и технологии (IST'2009): Междун. науч. конф., Минск, 2009. С. 179-183.
 12. Соловьева А. Профессиональный перевод с помощью компьютера. М.: ПИТЕР, 2008. 158 с.
 13. Степанова Д.В. Лингвистические аспекты перевода на русский язык английских терминологических словосочетаний с использованием корпуса параллельных текстов: дис. ... канд. филол. наук. Минск, 2006. 141 с.
 14. Степанова Д.В. Построение принципиального алгоритма автоматического выделения английских терминологических словосочетаний и их перевода на русский язык с помощью корпуса параллельных текстов // Актуальные проблемы прикладной лингвистики: сб. науч. статей. В 2 ч. Минск: МГЛУ, 2008. Ч. 2. С. 200-209.
 15. Arthern P. Aids unlimited: The scope for machine aids in a large organization // Aslib Proceeding. 1981. 33. P. 309-319.
 16. Nagao M. A Framework of a Mechanical Translation between Japanese and English by Analogy Principle // Artificial and Human Intellegence. – Amsterdam: North–Holland. 1984. P. 173-180.
 17. Somers Harold, Fernandez Diaz Gabriela. Translation Memory versus Example-based MT: What is the Difference? // International Journal of Translation.. July – Dec., 2004. Vol. 16. № 2. P. 4-33.
-