Мамонтова В. В.

КОРПУС ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ И БАЗА ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ ПЕРЕВОДЧЕСКИХ СООТВЕТСТВИЙ: ПРОБЛЕМЫ И ПРОЦЕДУРЫ ФОРМИРОВАНИЯ

Адрес статьи: www.gramota.net/materials/1/2008/8-2/50.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по рассматриваемому вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2008. № 8 (15): в 2-х ч. Ч. II. С. 117-118. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2008/8-2/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

17. Сокращенная история Русской Церкви Филарета, архиепископа Черниговского. - СПб: В Синоидальной Типографии, 1869.
18. Федотов. Г. Святые Древней Руси. - М.: Моск. рабочий, 1990. - 269 с.

КОРПУС ПАРАЛЛЕЛЬНЫХ ТЕКСТОВ И БАЗА ДАННЫХ ДЛЯ ИССЛЕДОВАНИЯ ПЕРЕВОДЧЕСКИХ СООТВЕТСТВИЙ: ПРОБЛЕМЫ И ПРОЦЕДУРЫ ФОРМИРОВАНИЯ

Мамонтова В. В.
Северо-Кавказский государственный технический университет, г. Ставрополь

Использование корпусов в исследовании переводческих проблем может проходить разными способами в зависимости от возможностей корпуса и целей исследования. В контексте анализа воспроизведения сложносоставных слов в англо-русском переводе предполагается решение следующих проблем.

Во-первых, это тип корпуса и состав текстов. Непосредственным задачам такого исследования отвечает прежде всего параллельный корпус. Однако существующие корпусы не могут быть использованы в работе непосредственно для сбора нужного эмпирического материала - сложносоставных слов в публицистических текстах. Первой причиной является состав корпуса. Так, англо-русский и русско-английский корпус - подразделение Национального корпуса русского языка (http://ruscorpora.ru), имеет в своём составе английскую и американскую художественную литературу XIX-XX веков, а также русскую классическую литературу XIX века [Добровольский, Кретов, Шаров 2005]. Таким образом, современная публицистика в нём отсутствует, равно как и современная русская литература.

Во-вторых, это поисковые возможности корпуса. Поисковые возможности современных корпусов, несомненно, велики. НКРЯ позволяет выбирать грамматические признаки (часть речи, падеж, число, род, одушевлённость, наклонение/форма, время, лицо, залог, вид, переходность, степень/краткость), различные семантические характеристики. Большой набор параметров предусмотрен и для словообразовательной выборки. В частности, для предметных имён существительных имеется возможность выборки диминутивов, аугментативов, сингулятивов, nomina agentis и nomina feminina. Для непредметных существительных можно выбирать между отглагольными и отадъективными именами. В группе имён собственных доступны гипокористические и стяжённые формы, аббревиатуры [Национальный корпус русского языка]. Наиболее развёрнутые формы запроса разработаны для имени прилагательного, однако и в данном случае нет возможности задавать параметры для поиска сложных слов.

Всё сказанное выше заставляет искать иные пути использования корпуса в целях переводческого анализа сложносоставных слов и их соответствий. Одним из таких путей нам видится формирование собственного исследовательского корпуса параллельных текстов современной публицистики.

Первая проблема на пути формирования корпуса - это наличие исходных текстов на английском языке и их переводов. Для этих целей могут быть использованы ресурсы сайтов inosmi.ru, inopressa.ru и подобных, предоставляющих переводы на русский язык текстов зарубежной публицистики и дающие ссылки на оригинальные источники. В последнее время сайт inopressa.ru представляется более эффективным ресурсом, поскольку он регулярно даёт ссылки на доступные источники, тогда как источники сайта inosmi.ru часто оказываются платными сервисами. Таким образом, на первом этапе у исследователя оказывается английский текст и его русский перевод, хотя в поисковом режиме последовательность обратная: русский перевод и английский оригинал. В качестве примера рассмотрим текст «Дипломатия по заветам Дона Корлеоне» (The Washington Post, США) (русский перевод www.inossmi.ru, оригинал www. washingtonpost.com). Предварительно документы трансформируются из формата html в текстовый формат.

Следующим шагом является выравнивание текстов. Выравнивание текстов представляет собой непростую задачу. Вот как её описывают Д. О. Добровольский, А. А. Кретов и С. А. Шаров применительно к корпусу параллельных текстов НКРЯ. Корпус параллельных текстов обслуживается программой ПарТекс, которая имеет на входе два параллельных текста - оригинал и перевод: «Выравнивание осуществляется на уровне предложений. Программа выдаёт на выходе синхронизированный текст, в котором последовательно за каждым предложением оригинала следует соответствующее предложение перевода» [Добровольский, Кретов, Шаров 2005: 265]. Мы использовали программу ParaSent, размещённую на сайте словаря Мультитран [Мультиран]. На вход также подаются два текста в формате текстового документа (используется стандартная программа «Блокнот»). Затем программой производится выравнивание, в итоге получается файл Excel с параллельными предложениями на английском и русском языках. Данный пример иллюстрирует результат выравнивания на материале одного из текстов (приведены несколько фрагментов в начале текста):

Corleone Diplomacy

"Дипломатия по заветам Дона Корлеоне ("The Washington Post", США)"

By Jim Hoagland

Джим Хогланд (Jim Hoagland), 02 июля 2007

Vladimir Putin's Russia pursues power and treasure by constantly calling attention to its ability to disrupt U.S. goals abroad.

Открыто бравируя собственной возможностью срывать зарубежные планы США, Россия Владимира Путина стремится получить власть и богатство. Для сравнения:

Hu Jintao's China works to soften the effects of its accelerating economic and military rise and to embed itself ever more deeply in the U.S. economy.

Китай под властью Ху Цзиньтао (Hu Jintao) постоянно работает над тем, чтобы смягчить влияние своего укоряющегося экономического и военного подъема и как можно более тесно встроиться в структуру американской экономики.

Цитируемые выше исследователи следующим образом видят перспективы использования полученного материала: «В таком синтезированном тексте поиск интересующих пользователя слов может осуществляться штатными средствами обычных текстовых редакторов, например, таких, как «Майкрософт Ворд». В программе ПарТекс для поиска может быть «подстрока в строке» и на выход подается текстовый файл, в который входят все пары предложений оригинала и перевода, содержащие заданную последовательность символов. Возможен поиск как английских, так и русских слов или словосочетаний» [Добровольский, Кретов, Шаров 2005: 265].

Использование штатных средств текстовых редакторов для поиска слов, конечно же, не может вызвать возражения. Однако это пока не решённая компьютерными средствами проблема. Для её решения необходимо провести семантическую (словообразовательную) разметку, с помощью которой можно было бы маркировать сложносоставные слова. Между тем маркирование переходных случаев с помощью машинных средств было бы крайне затруднительно. Именно по этой причине подобные разметки не присутствуют в доступных корпусах.

Исходя из описанной ситуации в формируемом корпусе выявление сложносоставных слов проводится вручную. Однако мы не остановились на этом и продолжили исследование в направлении создания базы данных. Пример записи базы данных сложносоставных слов на материале текстов современной публицистики выглядит следующим образом (программа Microsoft Access).

Код

{1}

Источник

{washingtonpost.com Corleone Diplomacy}

Предложение_А

{The political skies over the Maine resort may be cloud-filled when Bush and Putin publicly describe their talks Monday}

Исходная форма

{may be *cloud-filled*}

Словарь_А

{}

Предложение_Р

{В понедельник, когда Путин с Бушем будут рассказывать общественности о результатах своих переговоров, в политическом небе над Мэном вполне могут сгуститься тучи**}**

Перевод

{могут *сгуститься тучи*}.

Данная база банных может быть затем расширена путём добавления как «информационных» (типа **Словарь_Р**, **Источник_Р**, так и «исследовательских» (**Грамматика_А**, **Структура_А**, **Грамматика_Р**, **Структура_Р** и других) полей. Завершать базу могут поля, содержащие выводы и мнения исследователя относительно рассматриваемых явлений, например, **Комментарий** и другие.

В таком виде (плюс сами исходные тексты на английском и русском языках, которые в случае необходимости также могут быть добавлены в базу) полученная база данных вполне может рассматриваться в качестве исследовательского корпуса для сопоставительного лингвопереводческого анализа сложносоставных слов в англо-русских публицистических текстах.

Список использованной литературы

1. **Добровольский Д. О., Кретов А. А., Шаров С. А.** Корпус параллельных текстов: архитектура и возможности использования // Национальный корпус русского языка: 2003-2005. - М.: Индрик, 2005. - С. 263-296.
2. **Мультитран. Система для переводчиков** [Электронный ресурс]. - http:// www.multitran.ru.
3. **Национальный корпус русского языка** [Электронный ресурс]. - http://ruscorpora.ru.

ОБ ОСОБЕННОСТЯХ КУЛЬТУРНЫХ ИКОНИЧЕСКИХ ЗНАКОВ

Мерзлякова Е. О.

Томский государственный педагогический университет

Данная статья посвящена изучению особенностей культурных иконических знаков, основных их параметров, позволяющих знаку иконическому «возвыситься» и приобрести культурное содержание.

Изучение различий знака и символа с позиций философии, лингвистики, культурологии и других наук позволяет исследователям считать, «что содержательной основой знака в языке является значение, а содержательной основой символа (слова как символа) - образ» [Иванов 2002: 51]; [Лосев 1993: 879]; [ЭС 1999: