

УДК 81'322
ББК 81.1

Е.М. Коваленко

КОНТРАСТИВНЫЕ ИССЛЕДОВАНИЯ В ПЕРЕВОДОВЕДЕНИИ: КОРПУСНЫЙ ПОДХОД

Рассматриваются вопросы использования корпусного подхода в профессиональной деятельности лингвиста-переводчика. Проанализированы возможности использования монолингвальных корпусов русского и английского языков для проведения контрастивных исследований в сфере переводоведения. Показано, что такой подход позволяет изучить не только различие в контекстах естественного употребления в языках в целом, но и в отдельных диалектах каждого из языков.

Ключевые слова: *корпусы текстов, корпусная лингвистика, переводоведение, коллокации, контрастивные корпусные исследования, НКРЯ, ГИКРЯ, GloWbE.*

DOI 10.23683/1995-0640-2017-3-56-65

Коваленко Елена Михайловна – докт. филос. наук, профессор кафедры перевода и информационных технологий в лингвистике Института филологии, журналистики и межкультурной коммуникации Южного федерального университета
Тел.: +7928-957-02-93
E-mail: emkovalenko@sfedu.ru

© Коваленко Е.М., 2017.

Компьютерные технологии в переводческой деятельности достаточно часто обсуждаются в контексте использования систем машинного перевода [Голов, 2011; Митрофанова, 2014]. Однако существуют и другие компьютерные ресурсы, которые могут быть использованы в работе переводчика, проведении исследований в области переводоведения, а также для обучения. К ним относятся, в первую очередь, корпуса текстов, которые обеспечивают достаточную, при условии репрезентативного объема корпуса, полноту представления языковых явлений в их естественной контекстной форме, позволяя гарантировать типичность языковых данных и объективность их изучения [Захаров, Богданова, 2011, с. 7].

Несомненный интерес для переводчика представляют корпуса параллельных текстов, которые дают возможность разрешить некоторые вопросы, связанные с трудностями перевода, например, найти контексты слов, имеющих несколько переводных эквивалентов, подобрать эквиваленты терминологических и фразеологических словосочетаний и т.п. В контексте русскоязычных переводов интерес представляют параллельные корпуса из Национального корпуса русского языка (КоПарТ), совокупный объем которых превышает 72 млн слов. Данный корпус постоянно пополняется и развивается как за счет расширения количества представленных языков, так и путём увеличения объемов двуязычных текстов. На сегодняшний день он представлен 16 подкорпусами, но,

несмотря на его активное развитие, имеет не очень большой объем двуязычных корпусов – от 25 тыс. в китайско-русском подкорпусе до 18 млн в англо-русском подкорпусе, при этом объемы текстов переводов с русского и на русский значительно разнятся (см. [НКРЯ]). В корпусе также заявлены, но текстов пока еще нет, 13 языков (чешский, хорватский, португальский, нидерландский, греческий и т.д.), что можно рассматривать как направление развития коллекции параллельных корпусов НКРЯ.

В современной корпусной лингвистике существует проблема создания сбалансированных по составу и достаточных по размеру параллельных корпусов [Потемкин, 2012, с. 139], в частности, небольшой объем параллельных текстов для каждого конкретного языка, неясность критериев отбора материала (проблема авторских прав, доступности текстов и т.п.) в КоПарТ НКРЯ ставит вопрос о его репрезентативности. Поэтому возникает необходимость обратиться к изучению возможностей монолингвальных корпусов для решения переводоведческих задач, в том числе для контрастивных исследований лексических вариаций, синонимов, коллокаций, культурно маркированных слов и фраз, типологических признаков текстов, скрытых моделей использования лексики, диахронического развития концептов. Для проведения подобных исследований можно обратиться к корпусам текстов исследуемых языков и провести сравнительный анализ полученных из корпусов результатов для изучаемого языкового явления. Для работы с русскоязычными переводами могут быть использованы одноязычные подкорпусы Национального корпуса русского языка и Генеральный Интернет-Корпус Русского Языка, а также корпусы текстов языка перевода (или оригинала). При этом проблема качественных и достаточных по объему параллельных корпусов существует и для других задач компьютерной лингвистики, например в работе [Mikolov, Le et al., 2013] рассматривается технология машинного перевода, позволяющая автоматизировать процесс создания модели перевода (таблиц слов и фраз) с использованием одноязычных корпусов текстов.

В качестве материала и инструментов исследования использованы корпусы английского языка Марка Дэвиса (corpus.byu.edu/), Национальный корпус русского языка (НКРЯ) (ruscorpora.ru), Генеральный Интернет-Корпус Русского Языка (ГИКРЯ) (www.webcorpora.ru/) и их программное обеспечение.

Среди доступных для свободного использования корпусов особый интерес для контрастивных исследований в области переводоведения представляют корпусы различных вариантов английского, испанского и португальского языков, архитектура и интерфейс для которых были разработаны профессором лингвистики Марком Дэвисом из университета Бригама Янга (США). Эта коллекция корпусов содержит вариант Британского Национального Корпуса (англ. *British National Corpus*, сокр. *BYU-BNC*) (100 млн слов), созданного в 1980-х – начале 1990-х гг. в Oxford University Press. Здесь представлены два крупнейших корпу-

са американского варианта английского языка: Корпус Современного Американского Английского языка (англ. *Corpus of Contemporary American English*, сокр. *COCA*) и Корпус Исторического Американского Английского (англ. *Corpus of Historical American English*, сокр. *COHA*), а также один из крупнейших веб-корпусов - Интернет-Корпус Английского языка (англ. *Corpus of Global Web-Based English*, сокр. *GloWbE*).

Корпус Современного Американского Английского языка (*COCA*) содержит 220 тыс. текстов и 520 млн слов, являясь самым большим и единственным общедоступным корпусом американского варианта английского языка. Корпус *COCA* регулярно обновляется (ежегодное пополнение \approx 20 млн слов), поэтому дает возможность исследовать текущие изменения в языке. В равных пропорциях представлены 5 жанров: *устный* (стенограммы спонтанной речи из 150 телепрограмм и радиопередач); *беллетристика* (короткие рассказы и пьесы из литературных, детских и популярных журналов, первые главы первого издания книг с 1990, сценарии кинофильмов); *популярные журналы* (статьи из 100 журналов различной тематики – новости, здоровье, финансы, религия, спорт и т.п.); *газеты* (статьи из различных разделов 10 крупнейших американских газет); *научные журналы* (статьи из 100 рецензируемых научных журналов, представленных по научным направлениям согласно системе каталога библиотеки Конгресса США) [*COCA*].

Корпус Исторического Американского Английского языка (*COHA*) содержит 115 тыс. текстов и более 400 млн слов, являясь самым большим корпусом «исторического» американского варианта английского языка. Корпус сбалансирован по жанрам и поджанрам с разделением на десятилетние периоды, что позволяет, по мнению создателей корпуса, изучать реальные исторические изменения в языке. В корпусе представлены 4 типа текстов, сбалансированных по составу в соответствии с каталогом библиотеки Конгресса США, причем период 1990 – 2010 гг. представлен текстами соответствующего жанра из корпуса *COCA*: *художественная литература* (тексты из Project Gutenberg, исторические книги проекта Making of America Корнельского университета, театральные и киносценарии); *журналы* (тексты из журналов проекта Making of America, журналы); *газеты* (не менее пяти газет для каждого десятилетия); *публицистика* (тексты из Project Gutenberg, archive.org, книги) [*COHA*].

Интернет-Корпус Английского языка (*GloWbE*) – один из крупнейших корпусов английского языка (около 2 млрд слов, собранных автоматически с веб-страниц 340 тыс. сайтов из 20 различных англоязычных стран мира, \approx 60% информации из блогов), который предоставляет возможность изучать различия между 20 диалектами английского языка. Например, если сравнить частоту употребления лексики для разных диалектов в корпусе, то такие формы, как *fortnight*, *rained off*, *on holiday* чаще встречаются в британском, ирландском, австралийском диалектах; *trousers*, *rather more* – в британском варианте английского; а *at university*, *[be] different to* – в британском и австралийском встречаются чаще, чем

в американском и ирландском диалектах. Корпус позволяет также проводить сравнительные исследования между группами стран, например, в Южной Азии распространены такие формы, как *out of station*, *be elder to*, *keep in view* (см. [GloWbE]).

Национальный корпус русского языка (НКРЯ) находится в свободном доступе в сети Интернет (ruscorpora.ru), его объем составляет более 600 млн слов, и помимо КоПаРТ он включает коллекцию одноязычных корпусов (Основной (более 265 млн слов), Синтаксический, Газетный, Диалектный и ряд других). НКРЯ предназначен как для лингвистов, так и для специалистов в других областях знания – тех, кто проводит исследования, основанные на изучении текстов (например, в области социологии, психологии, истории, маркетинга и т.п.). Корпус существует уже более 10 лет, достаточно активно используется и в научных исследованиях, и в обучении языку, и в компьютерных технологиях, например, на базе Основного корпуса НКРЯ был создан «Новый частотный словарь русской лексики» [Ляшевская, Шаров, 2009]. Этот корпус также используется в задачах машинного обучения, например, для обучения языковой модели на основе нейронных сетей *RusVectōrēs* (инструмент изучения отношений между словами в дистрибутивных моделях [*RusVectōrēs*]), причем хорошие результаты обучения подтверждают репрезентативность корпуса, его сбалансированность для основных языковых уровней, несмотря на его небольшой объем [Kutuzov, Andreev, 2015, p. 143].

Генеральный Интернет-Корпус Русского Языка (ГИКРЯ), созданный при технологической и организационной поддержке компании АВВУУ, представляет собой мегакорпус текстов (более 20 млрд слов), собранных посредством «полностью автоматической технологии сбора и разметки текстов из Рунета» [ГИКРЯ]. Этот корпус является коллекцией материалов крупнейших ресурсов российских социальных медиа: ВКонтакте, Живого Журнала, Блогов Мейл.ру, Журнального Зала, а также новостных порталов.

В связи с развитием сети Интернет и демократизацией процесса создания контента резко возрос поток неофициальных текстовых сообщений, что привело к появлению феномена социальных медиа, распространение которых открывает новые перспективы для изучения обыденного сознания. Поэтому ГИКРЯ в качестве инструмента для обработки и анализа такого разнообразного и объемного материала имеет огромное значение и для лингвистов, которые получили возможность изучать обыденную речь в естественном контексте, и для социологов, психологов, историков и других исследователей, получивших инструмент для мониторинга происходящих в массовом сознании процессов [Боровикова, 2012, [http](http://)]. При этом алгоритмы, связывающие метаязыковые параметры текстов (пол, возраст, место нахождения автора, дата написания текста, жанр) с их лингвистическими характеристиками, дают возможность изучать распространение языковых явлений в разных возрастных, социальных, региональных группах, следить за их ди-

намикой в сети Интернет, проводить различные диахронические исследования на материале социальных медиа.

Рассмотрим возможность проведения контрастивных исследований в сфере русско-английского и англо-русского перевода на материале рассматриваемых корпусов текстов; особенности проведения исследований в корпусах английского и русского языка.

По утверждению разработчиков, рассматриваемые корпуса английского языка позволяют проводить *диахронические* исследования, изучать *различия* между *диалектами* английского языка, а также выявлять *жанровые различия* благодаря нескольким факторам: представленности в корпусах текстов достаточно широкого спектра жанров, диалектов и периодов времени, большого объема текстов, а также оригинальной архитектуры и интерфейса, которые позволяют получать результаты поиска в несколько раз быстрее, чем Sketch Engine (www.sketchengine.co.uk) – самый известный сервис для больших корпусов.

Большой объем корпуса GloWbE дает возможность, помимо лексических исследований диалектов английского языка, проводить исследования идиоматических выражений, так как их частота употребления «чувствительна» к объему корпуса. Например, изучение в корпусе выражений, связанных с «*heart*», показало, что такие идиомы, как *have a big heart*, *heart of gold*, *heart to heart*, *heavy heart* более распространены в британском и американском варианте английского языка, *heart of stone* – в британском, американском и танзанийском диалектах, *heart bleeds for (someone)* – чаще всего употребляется в британском варианте, но также достаточно часто употребляется в американском и нигерийском диалектах, *heart stands still* встретила только в филиппинском диалекте (согласно данным корпуса), см. [GloWbE]. Хотя все эти варианты идиом на лексикографических ресурсах объясняются, но особенности их использования не описаны (см. [www.native-english.ru/idioms/]).

Одним из наиболее интересных применений корпуса GloWbE является возможность сравнить частоты употребления слов и словосочетаний в разных странах. Например, согласно данным корпуса, слова *Quran* или *Allah* чаще всего встречаются в странах, исповедующих ислам (Пакистан, Бангладеш, Малайзия), и в США, а слово *Buddh* – в Индии, Шри-Ланке и Великобритании; слово *feminism* используется в светских англоязычных странах в несколько раз чаще, чем в странах Южной Азии и Африки. Во всех корпусах Марка Дэвиса представлена возможность исследовать коллокации (частотно устойчивые словосочетания) [Захаров, 2015, с. 668] для выбранного корпуса и подкорпуса, которые автоматически создаются для заданного слова или фразы (глубина контекста задается) и ранжируются по частотности употребления в данном подкорпусе. Это довольно мощный инструмент для изучения способов употребления лексики, позволяющий изучать реальные естественные контексты функционирования лингвистических явлений. В корпусе GloWbE дополнительно можно исследовать коллокации для разных диалектов английского языка, например, согласно корпусу, сочетание [*adj.*] + *book*

в азиатских странах чаще употребляется по отношению к религиозным текстам (*divine, revealed, Buddhist*), а самые частотные сочетания – *new, comic, good + book*, чаще всего используются в США; слово *wife* чаще всего сочетается с прилагательным *good, new, future, beautiful*, причем чаще всего сочетание *good wife* употребляется в США, Великобритании, Пакистане и Нигерии: в контекстах американского и британского диалектов речь идет о *хорошей жене* как любящей и доверяющей мужчине, а во многих контекстах пакистанского и нигерийского диалекта – как о социальной роли женщины по отношению к мужчине, кроме того, в светских странах слово *wife* с такими прилагательными как *chaste, temporary, obedient, Muslim, virtuous* употребляется гораздо реже, чем в странах Востока (см. [GloWbE]).

Кроме описанных корпусов коллекция Марка Дэвиса содержит еще корпус новостных текстов, автоматически собираемых в сети Интернет, корпусы текстов британского варианта Википедии, Британского парламента, сценариев американских сериалов, небольшой корпус канадского варианта английского языка и корпус на основе *Google Books* (американский, британский английский и испанский языки). Таким образом, эта коллекция дает уникальные возможности для изучения различных вариантов английского языка в их естественном контексте.

Диахронические исследования и изучение жанровых различий позволяет проводить и Национальный корпус русского языка. Для изучения диалектов русского языка можно использовать небольшой Диалектный подкорпус НКРЯ (www.ruscorpora.ru/search-dialect.html), который, однако, не позволяет изучать различные вариации русского языка подобно GloWbE, так как имеет небольшой объем (около 278 тыс. слов) и содержит только устные тексты. Основной подкорпус НКРЯ тоже не может быть использован для изучения диалектных вариаций русского языка, поскольку содержит в основном литературные тексты, и даже тексты электронной коммуникации (блоги, смс-сообщения, форумы), которые есть в подкорпусе, не имеют метаданных по локации их создания. НКРЯ имеет очень интересную и разветвленную систему запросов как по метаданным, так и морфологический, лексико-грамматический и семантический поиск, однако инструмента создания коллокаций по интересующему подкорпусу *не имеет* – инструменты N-грамм, представленные в НКРЯ, работают только с текстами Основного корпуса в целом (www.ruscorpora.ru/help-ngrams.html). Это несколько сужает возможности корпуса, так как построение частотного списка коллокаций для заданного подкорпуса требует кропотливой работы с полученными по запросу контекстами.

Очень перспективным для решения задачи исследования диалектных вариаций русского языка, построения коллокаций и проведения диахронических исследований для выбранного подкорпуса является ГИКРЯ, несмотря на некоторые ограничения по доступу (необходимость регистрации по запросу) в связи с продолжающейся активной разработкой и тестированием [ГИКРЯ]. Именно в корпусе ГИКРЯ можно про-

вести исследования (подобно исследованиям в GloWbE) для русского языка по частоте употребления слов и словосочетаний в разных регионах России и странах, где зарегистрированы русскоговорящие пользователи социальных медиа.

Исследование в корпусе, аналогичное проведенному в GloWbE, показало, что слова *Коран* и *Аллах* в российских социальных медиа (по материалам ГИКРЯ) чаще всего встречаются в сообщениях, которые отмечены локацией (местонахождением автора) – *Дагестан*, а слова *Будда* и *феминизм* – локациями *Москва*, *Россия* и *Петербург*, при этом разброс местонахождения авторов сообщений для этих слов значительно выше, чем для слов *Коран* и *Аллах*; слово *книга* употребляется в основном в светском контексте, хотя были найдены несколько контекстов, связанных с духовными книгами, но, к сожалению, локация в них была не определена, при этом чаще всего слово *книга* употребляется с прилагательным *любимая*. Слово *жена* употребляется вместе с прилагательными *целомудренная*, *послушная*, *покорная*, *добродетельная* достаточно редко в корпусе, причем в подавляющем большинстве случаев локация не выявлена, поэтому определенных выводов сделать невозможно о связи данных словосочетаний с географическим положением авторов. Изучение контекстов употребления данных словосочетаний в НКРЯ показало, что сочетание *целомудренная жена* употребляется в Основном подкорпусе НКРЯ всего 4 раза: 1 раз в религиозном тексте (1779 г.), 1 раз в публицистике (1934 г.), 1 раз в «Истории государства Российского» (том 3, 1808 – 1818) и 1 раз в художественной литературе (1814 г.); сочетание *временная жена* употребляется в Основном корпусе 8 раз, причем чаще всего в значении «фронтальная жена», *послушная жена* – 12 раз, *покорная* – 22, *добродетельная жена* – 24 (например, (1) *Домострой*» *учил, что добродетельная жена добра, трудолюбива и молчалива* [Вера Михайлова. *Дамы отжимают кавалеров* // «Русский репортер», 2014]).

Исследование в корпусах русского языка (НКРЯ и ГИКРЯ) показало, что словосочетания *добродетельная*, *целомудренная*, *покорная*, *послушная жена* для русского языка – достаточно редкое явление, а самыми употребительными сочетаниями [*прилагат.*] + *жена* в обоих корпусах являются: *молодая*, *новая*, *бывшая*, *любимая*, в отличие от исследований в корпусе GloWbE: *good*, *new*, *future*, *beautiful* (за исключением *new* и *новая*). При этом словосочетание *хорошая жена* достаточно часто встречается в корпусах русского языка, обозначая чаще всего качества женщины – добрая, приветливая, заботливая, обустроивающая жизнь мужчины, что также отличается от английских контекстов.

Изучение переводных аналогов слов и словосочетаний в английских и русских подкорпусах показало различие в контекстах естественного употребления не только в языках в целом, но и в отдельных диалектах каждого из языков, обусловленных различием в восприятии, менталитете, социальном устройстве жизни. Следовательно, корпуса текстов дают уникальные возможности для изучения различных вариантов английского и русского языка, что представляет несомненный интерес не

только для лингвиста-исследователя, но и переводчика, преподавателя языка, а также для всех, кто изучает язык.

Рассмотренные корпусы текстов представляют мощный инструмент для контрастивных исследований языка в контексте переводоведения. В то же время хочется отметить потенциал в развитии корпусов русского языка: получение коллокаций непосредственно из заданного подкорпуса с возможностью ранжирования их по частоте употребления, уточнение локаций авторов текстов для проведения исследований по вариациям языка значительно увеличили бы возможности использования корпусов русского языка для решения многих исследовательских задач.

Литература

Боровикова Е. Российские учёные разметят интернет [Электронный ресурс]. // STRF.ru – Наука и технологии России. 18.12.2012. URL: http://www.strf.ru/material.aspx?d_no=50859&CatalogId=21731#.VujLjKyNBc (дата обращения 14.03.2017).

Генеральный Интернет-Корпус Русского Языка (ГИКРЯ). О проекте. [Электронный ресурс]. URL: <http://www.webcorpora.ru/> (дата обращения 14.03.2017).

Голов А. Н. Информационные технологии как инструмент перевода иноязычного технического текста // Наука и современность. 2011. № 11 С. 244–249.

Захаров В. П., Богданова С. Ю. Корпусная лингвистика. Иркутск, 2011. 161 с.

Захаров В. П. Сочетаемость через призму корпусов // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 27–30 мая 2015 г.). Вып. 14 (21): В 2 т. Т. 1. М.: Изд-во МГГУ, 2015. С. 667–682.

Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). М.: Азбуковник, 2009. 1112 с.

Митрофанова Т. А. Роль современных ИКТ в формировании профессиональной компетентности переводчиков // Современное языковое образование: инновации, проблемы, решения: сб. науч. тр. М.: Изд-во МГГУ им. М.А.Шолохова, 2014. с. 42 – 51.

Национальный корпус русского языка (НКРЯ). [Электронный ресурс]. URL: <http://www.ruscorpora.ru/index.html> (дата обращения 14.03.2017).

Потемкин С. Б. Проблемы разработки параллельного корпуса переводов русской классики // Научно-информационный журнал «Армия и общество». 2012. № 2 (30). С. 138 – 146.

British National Corpus (BYU-BNC). [Электронный ресурс]. URL: <http://corpus.byu.edu/bnc/> (дата обращения 12.03.2017).

Corpus of Contemporary American English (COCA). [Электронный ресурс]. URL: <http://corpus.byu.edu/coca/> (дата обращения 12.03.2017).

Corpus of Global Web-Based English (GloWbE). [Электронный ресурс]. URL: <http://corpus.byu.edu/glowbe/>, <http://corpus.byu.edu/glowbe/help/dialects.asp> (дата обращения 12.03.2017).

Corpus of Historical American English (COHA). [Электронный ресурс]. URL: <http://corpus.byu.edu/coha/> (дата обращения 12.03.2017).

Kutuzov A., Andreev I. Texts in, Meaning out: Neural Language Models in Semantic Similarity Tasks for Russian, Computational Linguistics and Intellectual

Technologies. Papers from the Annual International Conference (Dialogue'2015), Issue 14, vol. 2, of 2, Moscow, 2015, pp. 133 – 144.

Mikolov T., Le Q. V., Sutskever I. Exploiting similarities among languages for machine translation. Arxiv preprint, 2013. Available at <http://arxiv.org/abs/1309.4168> (дата обращения 14.03.2017).

RusVectōrēs. О проекте. [Электронный ресурс]. URL: <http://ling.go.mail.ru/dsm/ru/about> (дата обращения 14.03.2017).

References

Borovikova E. Rossiyskiye uchenyye razmetyat internet. *STRF.ru – Nauka i tekhnologii Rossii*. 18.12.2012. Available at: http://www.strf.ru/material.aspx?d_no=50859&CatalogId=21731#.VyjLjdKyNBc (accessed 14.03.2017). (In Russian).

General'nyy Internet-Korpus Russkogo Yazyka (GIKRYa). *O proyekte*. Available at: <http://www.webcorpora.ru/> (accessed 14.03.2017). (In Russian).

Golov A. N. Informacionnyye tekhnologii kak instrument perevoda inoyazychnogo tekhnicheskogo teksta, *Nauka i sovremennost'*, 2011, no 11, pp. 244-249. (In Russian).

Zakharov V. P., Bogdanova S. Yu. *Korpusnaya lingvistika*, Irkutsk, 2011, 161 p. (In Russian).

Zakharov V. P. Sochetaemost' cherez prizmu korpusov. *Komp'uternaya lingvistika i intellektual'nyye tekhnologii: po materialam ezhegodnoy Mezhdunarodnoy konferentsii «Dialog»*, 14 (21), Tom 1, Moscow, 2015, pp. 667-682. (In Russian).

Lyashevskaya O. N., Sharov S. A. *Chastotnyy slovar' sovremennogo russkogo yazyka (na materialakh Nacional'nogo korpusa russkogo yazyka)*, Moscow, 2009, 1112 p. (In Russian).

Mitrofanova T. A. Rol' sovremennykh IKT v formirovaniy professional'noy kompetentnosti perevodchikov, *Sovremennoye yazykovoye obrazovaniye: innovatsii, problemy, resheniya, Sbornik nauchnykh trudov*, Moscow, 2014, pp. 42-51. (In Russian).

Natsional'nyy korpus russkogo yazyka (NKRYa). Available at: <http://www.ruscorpora.ru/index.html> (accessed 14.03.2017). (In Russian).

Potemkin S. B. Problemy razrabotki parallel'nogo korpusa perevodov russkoy klassiki, *Nauchno-informacionnyy zhurnal Armiya i obshchestvo*, 2012, no 2 (30), pp. 138-146. (In Russian).

British National Corpus (BYU-BNC). Available at: <http://corpus.byu.edu/bnc/> (accessed 12.03.2017).

Corpus of Contemporary American English (COCA). Available at: <http://corpus.byu.edu/coca/> (accessed 12.03.2017).

Corpus of Global Web-Based English (GloWbE). Available at: <http://corpus.byu.edu/glowbe/>, <http://corpus.byu.edu/glowbe/help/dialects.asp>. (accessed 12.03.2017).

Corpus of Historical American English (COHA). Available at: <http://corpus.byu.edu/coha/> (accessed 12.03.2017).

Kutuzov A., Andreev I. Texts in, Meaning out: Neural Language Models in Semantic Similarity Tasks for Russian, *Computational Linguistics and Intellectual Technologies, Papers from the Annual International Conference (Dialogue'2015)*, Issue 14, vol. 2 of 2, Moscow, 2015, pp. 133-144.

Mikolov T., Le Q. V., Sutskever I. *Exploiting similarities among languages for machine translation*. Arxiv preprint, 2013. Available at: <http://arxiv.org/abs/1309.4168>. (accessed 12.03.2017).

RusVectōrēs, O projekte, Available at: <http://ling.go.mail.ru/dsm/ru/about> (accessed 14.03.2017). (In Russian).

Elena M. Kovalenko (Rostov-on-Don, Russian Federation)

Contrastive Research in Translation Study: Corpus Approach

This article discusses some issues of the corpus approach application in the professional activities of a linguist-translator. It describes the collection of corpora of the English language, created by Mark Davis, Professor of Linguistics of Brigham Young University. The paper analyzes the potential of the Russian National Corpus, General Internet-Corpus of Russian (RNC, GICR) and one of the largest corpora created by Mark Davis - Corpus of Web-Based Global English (GloWbE) in solving translational problems. It also considers the applications of monolingual corpora in conducting contrastive research in the field of translation studies: some lexical variations in different English dialects in the Corpus of Web-Based Global English, as well as their translation equivalents in the Russian corpora were studied; a comparative analysis of the obtained results for the addressed linguistic phenomenon was conducted. It is shown that such approach allows us to study the difference in the contexts of natural use not only in languages in general, but also in separate dialects of each language.

Key words: *corpora, corpus linguistics, translation studies, collocations, contrastive corpus research, RNC, GICR, GloWbE.*

Elena M. Kovalenko – Ph.D. of philosophy, professor of translation/interpreting study and IT in linguistics dpt. Institute of philology, journalism and cross-cultural communication. Southern Federal University. Phone: +7-928-957-02-93, e-mail: emkovalenko@sfedu.ru