



Math-Net.Ru

Общероссийский математический портал

М. Г. Кружков, Информационные ресурсы контрастивных лингвистических исследований: электронные корпуса текстов, *Системы и средства информ.*, 2015, том 25, выпуск 2, 140–159

DOI: 10.14357/08696527150209

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 94.141.255.10

19 мая 2024 г., 21:19:54



ИНФОРМАЦИОННЫЕ РЕСУРСЫ КОНТРАСТИВНЫХ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ: ЭЛЕКТРОННЫЕ КОРПУСА ТЕКСТОВ*

*М. Г. Кругжков*¹

Аннотация: Рассматриваются информационные ресурсы, используемые в контрастивных лингвистических исследованиях, и их основные характеристики. Основными видами таких информационных ресурсов являются типологические базы данных (ТБД) и электронные корпуса текстов. В рамках данной работы основное внимание уделяется электронным корпусам, среди которых наиболее релевантными для проведения контрастивных исследований являются сопоставимые корпуса, представляющие собой сбалансированные собрания оригинальных текстов на сравниваемых языках, и параллельные (переводные) корпуса, представляющие собой собрания оригинальных текстов на одном из сравниваемых языков, дополненные переводом на другой язык (или несколькими переводами на один или несколько языков). В дополнение к описанию известных информационных ресурсов для поддержки контрастивных лингвистических исследований определяется новый тип таких ресурсов: «надкорпусные кросслингвистические базы данных». Дается описание основных характеристик таких баз данных (БД) в сопоставлении с электронными корпусами текстов и обосновывается необходимость их формирования.

Ключевые слова: контрастивная лингвистика; базы данных; типологические базы данных; сопоставимые корпуса; параллельные корпуса; надкорпусные кросслингвистические базы данных

DOI: 10.14357/08696527150209

1 Введение

Данная статья является продолжением предыдущей работы [1], также посвященной рассмотрению информационных ресурсов контрастивных лингвистических исследований. Как отмечалось в работе [1], *контрастивная лингвистика* является одним из направлений современной лингвистики, которое ставит перед собой задачу сравнения языковых явлений, наблюдаемых в различных естественных языках. Контрастивные лингвистические исследования имеют как теоретическое, так и прикладное значение, позволяя выявлять общие особенности языков и специфичные характеристики отдельных языков. С одной стороны,

* Работа выполнена при поддержке РФФИ (проект 13-06-00403).

¹ Институт проблем информатики Федерального исследовательского центра «Информатика и управление» Российской академии наук, magnit75@yandex.ru

контрастивная лингвистика дает материал для языковой типологии и для выявления языковых универсалий, с другой стороны, она непосредственно связана с различными образовательными дисциплинами, такими как преподавание иностранных языков, теория и практика перевода [2], а также используется при создании систем машинного перевода.

В лингвистике традиционный ручной сбор исходных данных, которые необходимы для проведения экспериментальных исследований, часто является трудоемким и дорогостоящим этапом работы. Формирование информационных ресурсов в форме электронных корпусов и БД позволяет существенно экономить время для проведения лингвистических исследований [3–6], что в особенности актуально для задач контрастивной лингвистики, в процессе решения которых сопоставляются тексты на разных языках.

В конце прошлого века сформировалась потребность в развитии инструментария, позволяющего, с одной стороны, эффективно хранить и обрабатывать большие объемы текстов на разных языках и, с другой стороны, фиксировать результаты экспериментальных и теоретических исследований, получаемые лингвистами в процессе контрастивного анализа. Одним из наиболее эффективных инструментов реализации этих задач являются БД. При этом наиболее востребованы БД следующих двух типов:

- (1) компьютерные БД, разработанные в целях проведения типологических исследований (ТБД). Такие БД представляют собой систематизированные сведения об определенных характеристиках различных естественных языков. Этот вид информационных ресурсов был рассмотрен в работе [1].
- (2) электронные корпуса текстов, активно используемые при проведении экспериментальных лингвистических исследований. Применительно к задачам контрастивного анализа наиболее релевантными в этом плане являются сопоставимые и параллельные корпуса, которые будут рассмотрены в данной работе.

Тексты корпусов обычно *размечаются* для удобства проведения исследований, т. е. текстам и содержащимся в них языковым единицам (чаще всего словам) приписываются специальные метки. Используется, как правило, морфологическая разметка, но многие корпуса включают в себя также семантическую, синтаксическую, жанровую и другие виды разметки. Размеченные корпуса обеспечиваются специализированными поисковыми системами, реализующими как лексический, так и грамматический виды поиска.

В параллельных корпусах оригинальные и переводные тексты разбиты на фрагменты (обычно включающие в себя одно или несколько предложений или их частей). Соответствующие по смыслу фрагменты оригинального и переводного текстов геометрически выровнены друг с другом, что позволяет пользователю видеть, какой фрагмент текста на языке перевода соответствует данному отрезку текста на языке оригинала, и наоборот — какой отрезок текста на языке оригинала соответствует данному отрезку переводного текста.

Фрагмент русского текста, выровненный с тремя вариантами перевода на французский язык (из повести Н. В. Гоголя «Нос»)

Оригинал	Перевод1	Перевод2	Перевод3
К счастью, в кондитерской никого не было; мальчишки мели комнаты и расставляли стулья;	Par bonheur, la confiserie était vide; les garçons balayaient les salles et disposaient les chaises.	Le café était heureusement vide; les garçons balayaient les salles et rangeaient les chaises;	Heureusement, ul n'y avait personne dans le salon de thé; les garçons balayaient les salles et rangeaient les chaises;
некоторые с сонными глазами выносили на подносах горячие пирожки; на столах и стульях валялись залитые кофеом вчерашние газеты.	Quelques-uns, les yeux bouffis de sommeil, portaient des plateaux chargés de petits pâtés tout chauds; des journaux de la veille, tachés de café, s'éparpillaient sur les tables et sur les chaises.	d'aucuns, les yeux bouffis de sommeil, apportaient des plateaux chargé de patés châuds; les journaux de la veille, maculés de café, jonchaient les tables et les chaises,	certain, avec des yeux endormis, apportaient des pirojki chauds sur des plats; sur les tables et les chaises traînaient les journaux d'hier inondés de cafeé.
«Ну, слава богу, никого нет, — произнес он, — теперь можно поглядеть». Он робко подошел к зеркалу и взглянул.	“Grâce à Dieu il n'y a perssone! se dit Kovaliov; nous allons bien voir. . .” Il s'approcha timidement de la glace et regarda :	“Dieu merci, il n'y a personne, je vais pouvoir me regarder!” se dit Kovaliov en s'approchant d'une glace. Mais après un timide coup d'ceil:	“Dieu merci, il n'y a personne, prononça-t-il. Je peux maintenant me regarder.” Il s'approcha timidement de la glace et se jeta un coup d'ceil.
«Черт знает что, какая дрянь!» — произнес он, плюнувши. —	“Le diable sait ce que c'est! En voilà une sale histoire! prononça-t-il en crachant.	“Pouah, l'horreur! murmura-t-il, en crachant de dépit.	“Par tous les diables, c'est dégoûtant! s'écria-t-il en crachant.
Хотя бы уже что-нибудь было вместо носа, а то ничего!..»	Si au moins il y avait encore quelque chose à la place du nez. . . mais rien!..”	S'il y avait au moins quelque chose en olace de nez; mais non, rien, eien, rien!..”	Si au moins j'avais quelque chose à la place du nez. Mais alors rien!..”

Приведенная таблица наглядно иллюстрирует базовую структуру параллельного корпуса: в данном случае фрагмент русского текста выровнен сразу с тремя вариантами перевода на французский язык.

Электронные корпуса и ТБД являются эффективными средствами для проведения контрастивных лингвистических исследований, но это принципиально разные инструменты. Типологические БД включают в себя справочную информацию о свойствах и характеристиках различных языков либо отдельных языковых элементов. Электронные корпуса представляют собой, с одной стороны, (пополняемые) хранилища текстовых и языковых данных, с другой стороны, инструменты, с помощью которых могут быть получены новые лингвистические знания.

Цель данной статьи состоит в том, чтобы описать различные виды электронных корпусов, используемых в контрастивных лингвистических исследованиях:

- сопоставимые корпуса, представляющие собой сбалансированные собрания оригинальных текстов на сравниваемых языках (разд. 3);
- параллельные (переводные) корпуса, представляющие собой собрания оригинальных текстов с переводом (или с несколькими переводами) на другой язык (разд. 4).

Существует несколько разновидностей параллельных корпусов:

- классические (параграф 4.1);
- двусторонние (параграф 4.2);
- многоязычные (параграф 4.3);
- поливариантные (параграф 4.4).

Рассматриваются основные характеристики каждого из перечисленных видов электронных корпусов и решаемые с их использованием задачи.

Кроме этого, одной из важных целей статьи является описание новой категории так называемых «надкорпусных» БД, представляющих собой надстройки над корпусами, позволяющие сохранять результаты лингвистического анализа. В разд. 5 описаны их основные характеристики и обоснована необходимость их построения в процессе проведения контрастивных исследований.

2 Корпуса текстов и контрастивные лингвистические исследования

В отличие от ТБД, корпуса в явном виде не содержат информацию о характеристиках языка или языков, на которых написаны содержащиеся в них тексты¹. Однако с помощью корпусных методов исследователи могут получать такую информацию. Одним из главных преимуществ корпусных методов по сравнению с ручным анализом текстов или с использованием языковой интуиции

¹Косвенным образом такая информация может содержаться в размеченном корпусе, так как сама структура разметки включает в себя информацию, например, о частях речи данного языка, о допустимых числах, родах и т. д., но это лишь вспомогательная информация, которую можно извлечь из корпуса при наличии соответствующей разметки. Сама по себе она не позволяет судить о частотности, нормальности или аномальности тех или иных признаков и их сочетаний.

исследователя является то, что они помимо исследования известных явлений позволяют также выявлять и описывать языковые явления, о которых мы ранее имели лишь смутное представление или вовсе ничего не знали [7].

Как правило, исследователь использует корпус либо для поиска примеров определенных языковых единиц или явлений, которые он затем анализирует на качественном уровне, либо для получения статистических данных об их частотности в корпусе или в некоторой его части, *подкорпусе*. Подкорпус может формироваться по различным разрезам — языки публикаций, авторы, жанры, время создания текстов и т. п.

Как упоминалось выше, многие корпуса включают в себя разметку на уровне словоформ или более крупных языковых единиц. Чем богаче и разнообразнее разметка корпуса, тем более гибкие и эффективные поисковые интерфейсы могут быть предоставлены пользователям и тем выше его научная и образовательная ценность. Например, в Национальном корпусе русского языка (НКРЯ) [8] используется шесть типов разметки:

- (1) метатекстовая (тип текста, информация об авторе, время создания и т. д.);
- (2) морфологическая (леммы, грамматические признаки слов);
- (3) синтаксическая (синтаксическая структура предложений);
- (4) акцентная (расстановка тонов и ударений);
- (5) семантическая (семантические признаки и классы слов);
- (6) словообразовательная (данные о составе слов: префиксы, корни, аффиксы).

Корпуса могут быть проблемно-ориентированными или представлять собой универсальный инструмент, позволяющий поддерживать широкий спектр лингвистических исследований. В последнем случае особую ценность приобретают такие характеристики корпуса, как представительность и сбалансированность текстов, включенных в корпус. Сбалансированный корпус должен содержать широкий спектр различных типов письменных и устных текстов (художественные разных жанров, публицистические, научные, деловые, разговорные и т. п.), при этом каждый тип в идеале также должен быть представительным.

Применительно к компаративным лингвистическим исследованиям особую ценность имеют сопоставимые и параллельные корпуса¹. Среди параллельных корпусов можно выделить несколько подтипов, например многоязычные (с переводами на несколько языков) и поливариантные корпуса (с несколькими вариантами перевода на один и тот же язык). В следующих разделах будут подробнее рассмотрены перечисленные виды электронных корпусов и то, как они применяются в сравнительном языкознании.

¹Stig Jonansson называет параллельные корпуса переводными, а термин «параллельный корпус» расширяет и на сопоставимые, и на переводные корпуса [9]. Однако большинство исследователей под термином «параллельный корпус» понимают исключительно параллельные переводные корпуса. Поэтому в статье сопоставимые корпуса позиционируются как отдельный тип корпусов, тем более что они принципиально отличаются от переводных.

3 Сопоставимые корпуса

При проведении контрастивных лингвистических исследований часто используются так называемые *сопоставимые* корпуса [9]. Корпуса на двух или более языках называются сопоставимыми, если они составлялись на основе единых принципов из текстов одних и тех же жанров (например, речь молодежи или тексты новостей). Принципы составления включают в себя правила отбора текстов, порядок их разметки, регулирование допустимых объемов текстов и т. п.

Как правило, в рамках анализа сопоставимых корпусов исследователи хотят сравнить уровень частотности каких-либо явлений в различных языках. Так, в работе [10] сравнивается функционирование прагматических маркеров *vale* и *okay* в речи молодежи в испанском и английском языках, в работе [11] авторы сравнивают функционирование количественных аппроксиматоров в английском и французском языках, а в работе [12] авторы пытаются проследить различия в свойствах темы предложения в различных газетных жанрах в английском и испанском языках.

То, что тексты в сопоставимых корпусах не являются непосредственным переводом друг друга, имеет как недостатки, так и преимущества. С одной стороны, семантическое несовпадение содержания текстов в некоторых случаях может ставить под сомнение обоснованность их сравнения. Вместе с тем если сопоставляемые корпуса достаточно объемные, представляют близкие жанры и при их составлении использовались единые принципы формирования, то сделанные на их основе выводы, касающиеся относительной частотности тех или иных явлений в рассматриваемых языках, несомненно, заслуживают внимания.

Кроме этого, анализ сопоставимых корпусов имеет то преимущество, что тексты в этих корпусах изначально порождаются на языке оригинала, т. е. не являются переводными.

Как известно, переводные тексты отличаются от оригинальных, поскольку на них могут влиять так называемые *факторы перевода* (*translation effects*, см., например, [13]). Одним из наиболее известных факторов перевода является эффект калькирования, суть которого состоит в том, что при переводе особенности конструкций языка оригинала переносятся в конструкции языка перевода, в результате чего переводные тексты по некоторым лексическим или синтаксическим характеристикам значительно отличаются от текстов, изначально написанных на том же языке. Так, Capelle в своем анализе перевода предикатов движения с немецкого и французского показывает, что по тексту перевода можно с большой вероятностью определить, с какого языка он выполнен [14]. Использование сопоставимых корпусов позволяет исключить влияние факторов перевода, которые могут искажать результаты, полученные в результате анализа переводных параллельных корпусов.

Например, можно с определенностью утверждать, что выполнить исследование, результаты которого приведены в работе [11], с помощью переводного

параллельного корпуса было бы невозможно. Остановимся на этом подробнее. В данной работе сравнивается функционирование количественных аппроксиматоров в английском и французском, таких как *около*, *не менее*, *по крайней мере*, *почти* и т. п., и оценивается частотность употребления грамматических и семантических типов аппроксиматоров. В результате исследования, в частности, выяснилось, что во французском языке в 2–3 раза чаще указывается диапазон значений (например, *от 5 до 15 дней*), чем в английском. Очевидно, что при использовании переводного корпуса этот результат невозможно было бы получить, поскольку переводчики почти всегда переводят подобные выражения буквально (сложно представить, чтобы вышеприведенное выражение было переведено, например, как *около 10 дней*).

Таким образом, целесообразность использования сопоставимых корпусов зависит от объекта и предмета исследования: чем буквальнее осуществляется перевод конструкций данного типа, тем больше оснований предпочесть сопоставимые корпуса переводным. Степень буквальности перевода может зависеть как от самого характера рассматриваемого языкового явления, так и от жанра текста: в частности, новостные тексты обычно переводятся более буквально, чем художественная литература.

Двусторонние параллельные корпуса (см. параграф 4.2) также можно использовать как сопоставимые, если массивы оригинальных текстов в них сбалансированы. При этом возможно сопоставление как оригинальных, так и переводных текстов, а также сопоставление оригинальных и переводных текстов на одном и том же языке [9].

4 Параллельные корпуса

Несмотря на некоторые описанные выше преимущества сопоставимых корпусов, большая часть контрастивных лингвистических исследований проводится на основе параллельных (переводных) корпусов. В простейшем случае параллельные корпуса представляют собой собрание текстов на языке оригинала, снабженные одним переводом на некоторый язык («классические» параллельные корпуса). Однако существует несколько их разновидностей:

- *двусторонние параллельные корпуса* включают в себя одновременно переводы с языка *X* на язык *Y* и переводы с языка *Y* на язык *X*;
- *многоязычные параллельные корпуса* включают в себя тексты на языке оригинала, выровненные с несколькими переводами на различные языки, как правило, по одному переводу на один язык;
- *поливариантные параллельные корпуса* включают в себя тексты на языке оригинала, выровненные с более чем одним переводом на другой язык.

Далее будут рассмотрены примеры использования параллельных корпусов разных типов в контрастивных лингвистических исследованиях.

4.1 Классические параллельные корпуса

В работе [15] с помощью англо-испанского параллельного корпуса анализируются языковые средства испанского языка, использующиеся для передачи английского аффиксального отрицания. Цель исследования — проверка гипотезы о том, что, несмотря на то что между английскими и испанскими отрицательными аффиксами прослеживаются очевидные соответствия, на практике их употребление в этих языках значительно различается. Авторы разработали типологию языковых средств испанского языка, которые могут использоваться для перевода отрицательных аффиксов английского языка, и, обработав корпус, получили количественные данные по частотности использования этих средств в переводах. Полученные результаты свидетельствуют о различиях в функционировании рассматриваемых языковых средств в английском и испанском языке. Наиболее типичные переводные соответствия рассматриваются на качественном уровне с подробным разбором примеров.

В заключении своей работы авторы задаются вопросом, в полной ли мере в переводах отражаются нормы испанского языка или же переводные тексты подчиняются иным нормам, отличным от стандартных норм испанского (в последнем случае следует согласиться с выводами, сделанными, например, в работе [16], о том, что переводные тексты образуют некоторый «промежуточный код» или «диалект» языка, на который осуществляется перевод). Чтобы найти ответ на поставленный вопрос, авторы сравнивают данные, полученные из параллельного англо-испанского корпуса, с данными независимого мооязычного испанского корпуса. Проведенное сравнение свидетельствует о значительных расхождениях в употреблении языковых средств в переводных и оригинальных испаноязычных текстах, что может быть связано, например, с тем, что переводчики часто склонны выполнять перевод излишне буквально (см. также разд. 3 о проблеме факторов перевода).

В работе [17] автор ставит своей целью продемонстрировать, что параллельные корпуса позволяют выявлять семантические изменения, которые происходят в языке. Анализируя переводы французской конструкции *quand même*¹ на английский язык, автор пытается показать, что ее значение в значительной мере сместилось от традиционного противительного к реляционному. Параллельные корпуса позволяют увидеть, каким образом разные переводчики интерпретируют значение той или иной конструкции в различных контекстах.

Общее число выделенных переводных соответствий для *quand même* было относительно невелико — около 250. В связи с этим некоторые единично встречающиеся переводы сложно квалифицировать как типичные, хотя, вполне вероятно, они таковыми и являются — например, перевод *quand même* через *you have to admit* и т. п. И напротив: некоторые переводы, встречающиеся более одного раза, например *so*, представляются не вполне удачными и плохо переда-

¹Основные возможные переводные соответствия конструкции *quand même* на русском языке: *если бы даже, все-таки, несмотря ни на что, вопреки всему, во что бы то ни стало.*

ющими смысл изначального выражения. Наконец, в некоторых случаях перевод не позволяет провести различие между противительным и реляционным значениями *quand même*, поскольку в нем могут использоваться столь же многозначные единицы. Тем не менее полученные количественные данные свидетельствуют о том, что в переводах конструкций с *quand même* примерно в 50% случаев противительное значение отсутствует, что позволяет с большой степенью уверенности утверждать, что значение *quand même* с течением времени действительно сместилось от противительного к противительно-реляционному.

В заключение автор указывает на необходимость принимать во внимание роль побочных контекстных эффектов, оказывающих влияние на окончательный выбор переводчика. В этой связи он цитирует работу [18], где данной проблеме уделяется пристальное внимание:

«Перевод редко представляет собой буквальное отображение оригинала на языке перевода <...>. Очевидно, есть множество причин, которые могут заставить переводчика предпочесть тот или иной вариант перевода. Переводчики не переводят отдельные слова или конструкции; скорее они заняты поиском соответствий для употребления определенных лингвистических элементов в определенных контекстах. В результате, при определении переводного соответствия для некоторого элемента часто приходится полагаться только на интуицию эксперта».

В этой связи приобретает особое значение возможность сопоставления переводов одних и тех же текстов на один и тот же язык различными переводчиками, что становится возможным при использовании *поливариантных корпусов* (см. параграф 4.4), а также привлечения экспертов для анализа различных вариантов перевода.

4.2 Двусторонние параллельные корпуса

Двусторонние корпуса включают в себя одновременно переводы с языка *X* на язык *Y* и переводы с языка *Y* на язык *X*. Можно сказать, что двусторонний корпус представляет собой два корпуса, объединенных под одной «обложкой», однако нужно учитывать, что обычно создатели двусторонних корпусов обеспечивают сбалансированный отбор исходных текстов для обоих языков, в связи с чем они представляют для исследователей гораздо большую ценность, чем два изначально не связанных друг с другом корпуса параллельных текстов с противоположными направлениями перевода.

Двусторонние корпуса также удобны тем, что при их анализе у исследователей есть возможность сравнивать свойства переводных и оригинальных текстов на обоих языках, не обращаясь к отдельным монологичным корпусам для проведения сравнения. Например, в работе [19] авторы анализируют устойчивые сочетания слов в английском и норвежском языке. В частности, анализ двустороннего англо-норвежского корпуса позволил им установить, что наиболее частотные английские сочетания слов в 2–3 раза чаще встречаются в перевод-

ных английских текстах, чем в текстах, изначально написанных на английском языке, что является ярким примером проявления фактора перевода. Таким образом, как отмечает Johansson [9], использование двустороннего корпуса позволяет проводить различия между собственно языковыми особенностями и явлениями, обусловленными факторами перевода.

Наконец, двусторонние корпуса позволяют рассматривать и сопоставлять способы трансляции различных явлений при переводе в противоположных направлениях, что невозможно сделать ни с помощью обычных параллельных корпусов, ни с помощью сопоставимых корпусов. Например, в работе [20] рассматривается явление «экстрапозиции» клаузы-подлежащего (т. е. переноса клаузы-подлежащего из начала в конец предложения). В частности, было показано, что английские экстрапозиции примерно в 75% случаев переводятся на шведский с помощью аналогичных экстрапозиций, а шведские экстрапозиции только примерно в 60% случаев переводятся на английский с помощью аналогичных экстрапозиций. Это и ряд других подобных сопоставлений позволили авторам заключить, что экстрапозиция клаузы-подлежащего более характерна для шведского языка. Полученные количественные результаты анализируются на качественном уровне, с разбором отдельных примеров и перечислением вероятных причин выявленной тенденции. Результаты данного исследования представляют ценность как для теоретиков контрастивного языкознания, так и, например, для шведских студентов, изучающих английский язык, поскольку они помогают понять, почему даже совершенно грамотная английская речь может порой звучать не вполне естественно.

4.3 Многоязычные параллельные корпуса

Многоязычные параллельные корпуса представляют собой коллекции текстов на языке X с переводами на два или более языков Y_1, Y_2, \dots , как правило, по одному переводу для каждого языка Y_i . Многоязычные корпуса могут использоваться как для углубленного анализа языка оригинала в зеркале переводов на другие языки, так и для сравнения между собой языков перевода, а также для сопоставления всех задействованных языков — и оригинала, и перевода.

Например, в работе [21] исследуются способы выражения отношения «между» в английском и французском языке посредством сравнения переводов ряда художественных текстов с норвежского языка, содержащих в оригинале слово «*mellom*», т. е. «*между*». До широкого распространения многоязычных корпусов считалось, что идеальным материалом для контрастивного анализа являются «семантико-синтаксические эквиваленции», т. е. максимально возможные приближения к грамматически дословному переводу¹ [22]. Однако такие переводы на практике встречаются довольно редко, поэтому при использовании обычных

¹Например: *And the lovely paths among the flower beds, strewn with crushed shells! — Et les jolies sentiers entre les plates-bandes, recouverts de sable! [И прелестные дорожки между клумбами, посыпанные ракушечником!]*

параллельных корпусов постоянно возникает вопрос о правомерности сравнения неконгруэнтных конструкций. Структура многоязычного параллельного корпуса позволяет принять текст на языке оригинала (в данном случае, норвежском) в качестве основы (*tertium comparationis*) для сравнения конструкций на языках перевода, поскольку во всех случаях переводчики стремятся транслировать значение одной и той же оригинальной конструкции.

Автор проводит различие между семью возможными типами значения отношения «между». Оно может характеризовать некоторую *локацию*, некоторую часть пути при *движении*, использоваться при обозначении *размера* чего-либо или для спецификации отрезка *времени*. Кроме этого, оно может определять *взаимоотношения* или *взаимодействие* между некоторыми объектами или присутствие некоторого внешнего наблюдателя, осуществляющего *сопоставление* объектов.

Всего в корпусе было выделено около 400 случаев употребления *mellom*, при этом в 56% случаев оно одновременно переводилось на английский и французский с помощью предлогов *between* и *entre*. Из этого автор заключает, что *between* и *entre* вместе покрывают около 56% семантического поля отношения «между»¹. Важно отметить, что данный результат было бы невозможно получить без использования многоязычного корпуса. Предположим, что при анализе англо-французского корпуса мы бы увидели, что *between* переводится с помощью *entre* в *x%* случаев, а при анализе франко-английского корпуса увидели бы, что *entre* переводится как *between* в *y%* случаев — эти цифры не дали бы нам возможности оценить степень их действительного семантического пересечения.

Аналогичным образом автор анализирует области семантического покрытия и степени пересечения для других английских и французских предлогов и прочих конструкций. В конечном счете исследование показало значительную схожесть в кодировании отношения «между» в английском и французском языке. В большинстве случаев в обоих языках используются предложные группы со стандартными предлогами *between/entre*, а использование альтернативных предлогов наблюдается относительно редко, за исключением кодирования пространственных смыслов (локация и движение). Кроме этого, беспредложные конструкции значительно чаще используются для кодирования отношения «между» во французском языке, чем в английском.

В работе [23] многоязычный корпус используется для проверки и дополнения результатов, полученных ранее с использованием моноязычных корпусов. Исследуется семантика английских «эссенциальных» наречий *basically*, *essentially* и *fundamentally* в зеркале их переводов на голландский и французский языки. Используемый в этом исследовании корпус относительно невелик, поэтому полученные результаты надо рассматривать с осторожностью, однако автор считает,

¹Кроме этого *between* и *entre* независимо друг от друга покрывают еще соответственно 23% и 18% данного семантического поля.

что полученных данных достаточно для решения главной поставленной задачи, т. е. верификации ранее полученных результатов.

В работе [24] исследуется лексический профиль шведского языка в сравнении с другими языками (английским, французским, немецким, финским, норвежским, датским, исландским и голландским) на примере глаголов движения в транспорте. Сходство или различие лексических профилей языков может рассматриваться в разных плоскостях. Например, применительно к глаголам движения языки могут отличаться: самими наборами глаголов движения; по степени обязательности использования конкретного глагола для выражения определенного типа движения; по степени расширяемости семантики отдельных глаголов. Кроме этого, во многих языках даже при наличии точного семантического эквивалента часто может отдаваться предпочтение глаголам с более общими или более частными значениями.

В многоязычном корпусе были выделены шведские глаголы движения в транспорте *åka*, *fara* и *köra*, затем была собрана статистика по способам их перевода на различные языки, после чего полученные данные были проанализированы на качественном уровне. В результате была построена таблица стандартных соответствий глаголов движения на транспорте для всех рассматриваемых языков. Из нее, в частности, видно, что из числа рассматриваемых языков только в шведском существует глагол, использующийся исключительно для описания движения в качестве пассажира транспортного средства *åka* (и его менее частотный синоним *fara*).

4.4 Поливариантные параллельные корпуса

Поливариантный параллельный корпус представляет собой собрание текстов на языке оригинала, выровненных с несколькими вариантами перевода на один и тот же другой язык. Поливариантные корпуса на сегодняшний день представляют собой относительно слабо развитое направление в корпусной лингвистике. Существует поливариантный русско-немецкий корпус, созданный в Австрийской академии наук на материале выполненных в разное время переводов романа Достоевского «Идиот» [6], но он не находится в открытом доступе. Кроме этого, в русско-французском параллельном корпусе НКРЯ несколько текстов имеют от двух до четырех вариантов перевода. Соответственно, очень невелико и число собственно контрастивных лингвистических исследований, выполненных на базе таких корпусов. Между тем потенциально поливариантные корпуса открывают новые перспективы для контрастивных исследований. Например, анализ переводов, выполненных в разное время, дает возможность исследовать диахронические изменения языка перевода. Кроме этого, сравнительный анализ переводов позволяет исследовать проблему «трудностей перевода», которые возникают, в частности, при переводе культурно-специфичной лексики и прочих лингвоспецифичных языковых единиц [5, 6].

5 «Надкорпусные» базы данных

При анализе корпусного языкового материала исследователи извлекают из корпусов нужную им для лингвистических исследований информацию и могут сохранять ее. Например, после выполнения поискового запроса в НКРЯ пользователь может сохранить несколько найденных результатов в одном из трех форматов: Excel, OpenOffice Calc или XML (<http://www.ruscorpora.ru/search-main.html>). Однако найденная в корпусе информация может иметь достаточно сложную структуру, при этом ее объем может оказаться слишком большим, чтобы впоследствии можно было бы легко находить нужные примеры или получать количественные данные по встречаемости примеров различных типов. Таким образом, для хранения и обработки извлеченной из корпуса информации часто необходимо использовать отдельную БД: либо независимую, либо представляющую собой «надстройку» над структурой корпуса.

В работе [1] рассматривались специализированные ТБД, позволяющие сохранять структурированную информацию об отдельных экземплярах рассматриваемых языковых явлений (см., например, [25]). Однако ТБД не позволяют делать обобщенных статистически обоснованных выводов о свойствах текстов или корпусов, из которых они были извлечены, поскольку они извлекались не посредством последовательной обработки всех текстов корпуса, а посредством целенаправленной выборки отдельных интересующих исследователей примеров. Такой подход вполне оправдан, когда перед создателями ТБД стоит задача проиллюстрировать те или иные языковые явления примерами из корпуса. Однако часто требуется не только найти отдельные примеры, но и оценить уровень частотности определенных языковых явлений в некоторой совокупности текстов корпуса или в корпусе в целом. В такой ситуации возникает необходимость в разработке специализированной технологии, позволяющей *последовательно обрабатывать все полные тексты корпусов* (или подкорпусов) и фиксировать в отдельной БД *каждое появление* исследуемого языкового явления. Можно сказать, что такая БД является в некотором роде продолжением корпуса или «надстройкой» над ним. Ввиду отсутствия устоявшегося термина, в рамках данной работы для них вводится термин «*надкорпусные*» БД.

В идеальном случае надкорпусные БД могли бы встраиваться непосредственно в инфраструктуру существующих корпусов, что позволило бы сделать их доступными для широкого круга пользователей этих корпусов, но они могут разрабатываться и независимо. Последний подход целесообразен на текущем этапе, когда методология создания надкорпусных БД проходит опробование в ряде лингвистических проектов [26–28]. Далее будут рассмотрены первые опыты по созданию надкорпусных БД.

В рамках проекта, представленного в работе [29], проводился анализ переводных соответствий сложносоставных слов в публицистическом дискурсе. Автор работы подчеркивает, что, несмотря на то что многие современные корпуса имеют широкие поисковые возможности, их часто оказывается недостаточно для

проведения лингвистических исследований. В частности, на момент выполнения работы [29] НКРЯ не позволял задавать параметры для поиска сложных слов. В результате автор пришел к выводу о необходимости создания собственного исследовательского массива параллельных текстов современной публицистики, который был им самостоятельно выровнен и в котором были выявлены английские сложносоставные слова и их русские переводные соответствия. При этом часть операций была автоматизирована, а часть выполнялась вручную¹.

На основе массива параллельных текстов была построена надкорпусная БД, в которую записывались переводные соответствия, полученные в результате обработки выровненных текстов. Было выделено 209 английских сложносоставных слов, принадлежащих к различным частям речи и имеющих различную внутреннюю структуру. Всего было выделено более 20 видов структуры английских сложносоставных слов, наиболее распространенными из которых являются *noun + noun*, *adjective + noun*, *noun + verb* и *adjective + adjective*. Соответствующие конструкции в русском языке также были разбиты по частям речи и по виду их внутренней структуры (если в переводе также использовались сложносоставные слова).

В работах [26–28] описан проект по исследованию глагольных форм русского языка в зеркале их переводов на французский язык. Исполнители проекта также пришли к необходимости создания надкорпусной БД, расширяющей функции, доступные в обычных параллельных корпусах, в данном случае НКРЯ. Основным логическим элементом этой БД является переводное соответствие, т.е. кортеж, объединяющий глагольную конструкцию оригинала с ее функционально эквивалентным фрагментом в тексте перевода². При этом каждая глагольная конструкция соотносится с некоторым базовым видом из типологии (системы) временных форм глагола в русском языке и может специфицироваться рядом дополнительных признаков (см. http://a179.ipi.ac.ru/corpora_dynasty/main.aspx).

Описанная БД включает в себя около 5 тыс. переводных соответствий, кроме этого она оснащена двуязычным лексико-грамматическим поисковым интерфейсом. Данный интерфейс позволяет искать соответствия интересующих пользователя типов, задавая различные сочетания грамматических признаков и/или лексики для русских глагольных конструкций и соответствующих им функционально эквивалентных фрагментов, а также получать информацию о частотности этих типов (подробнее о назначении, функциях и структуре данной надкорпусной БД см. работы [26–28]).

Разработанная технология формирования надкорпусных БД легко адаптируется для исследования разноплановых лингвистических явлений. Помимо

¹Во многих случаях при анализе параллельных текстов невозможно обойтись без ручных процедур, для выполнения которых приходится привлекать экспертов, свободно владеющих обоими языками.

²В работах [26–28] для обозначения описанного переводного соответствия используется термин *моноэквивалентность* (термин *полиэквивалентность* используется для обозначения более сложной структуры, включающей в себя более одного варианта перевода исходной конструкции).

описанного выше проекта по исследованию глагольных форм сегодня она уже используется для контрастивного анализа лингвоспецифичных языковых единиц и коннекторов.

6 Заключение

В данной работе, а также в работе [1], продолжением которой она является, было описано несколько категорий информационных ресурсов для поддержки контрастивных лингвистических исследований. В работе [1] были рассмотрены ТБД, которые можно разделить на ТБД широкого профиля и специализированные ТБД. В данной работе были рассмотрены электронные корпуса нескольких видов: сопоставимые, параллельные, двусторонние, многоязычные и поливариантные.

Было введено новое понятие — «надкорпусные» БД, представляющие собой надстройки над инфраструктурой параллельных корпусов. При обработке данных параллельных корпусов надкорпусные БД позволяют сохранять информацию о результатах контрастивного лингвистического анализа в виде формализованных описаний соответствий конструкций оригинального текста их функционально эквивалентным фрагментам в одном или нескольких переводных текстах. Данные описания формируются лингвистами-экспертами в результате последовательной обработки текстов параллельных корпусов.

Были даны примеры использования надкорпусных БД при проведении контрастивных исследований, а также выделены основные элементы сходства и различия для функционально и структурно близких БД и корпусов. В частности, было отмечено, что, несмотря на некоторое сходство специализированных ТБД и надкорпусных БД (и те и другие содержат примеры исследуемых конструкций или ссылки на них), принципиальное отличие между ними состоит в том, что в специализированных ТБД примеры специально подбираются, чтобы проиллюстрировать парадигмы рассматриваемых языковых явлений, в то время как надкорпусные БД содержат в себе *все* экземпляры исследуемых языковых явлений вместе с их переводными эквивалентами. При этом данные в надкорпусных БД генерируются посредством последовательной обработки параллельных текстов, а потому на их основе можно делать статистически обоснованные выводы и обобщения.

Последовательный анализ языковых объектов определенных категорий или видов в корпусе — всегда трудоемкий процесс, поэтому возможность сохранить результаты этой работы в отдельной БД имеет особую ценность. Возможность просматривать сохраненные переводные соответствия в различных разрезах и аспектах, делая срезы по различным грамматическим признакам, типам конструкций и/или лексическим единицам, позволяет исследователям проверять свои гипотезы на основе всего массива обработанных данных, а также выявлять и описывать языковые явления, которые не были ими замечены при выборочной обработке отдельных фрагментов корпуса.

Кроме этого, в тех случаях, когда задачей обработки корпуса одновременно занимается несколько экспертов-лингвистов, что неизбежно при проведении крупных проектов, надкорпусная БД обеспечивает возможность интегрировать плоды их совместной работы, не говоря уже о возможности предоставить другим исследователям доступ и к полученным результатам, и ко всем использованным исходным данным, что значительно упрощает процессы верификации полученных результатов и позволяет многократно использовать созданный информационный ресурс в дальнейших исследованиях.

Литература

1. *Кружков М. Г.* Информационные ресурсы контрастивных лингвистических исследований: типологические базы данных // Системы и средства информатики, 2015. Т. 25. № 1. С. 198–212.
2. *Гак В. Г.* О контрастивной лингвистике // Новое в зарубежной лингвистике. Вып. XXV. Контрастивная лингвистика. — М.: Прогресс, 1989. С. 5–17.
3. *Добровольский Д. О., Кретов А. А., Шаров С. А.* Корпус параллельных текстов // Научная и техническая информация. Сер. 2. Информационные процессы и системы, 2005. № 6. С. 16–27.
4. *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.
5. *Андреева Е. Г., Касевич В. Б.* Грамматика и лексика (на материале англо-русского корпуса параллельных текстов) // Национальный корпус русского языка: 2003–2005. — М.: Индрик, 2005. С. 297–307.
6. *Добровольский Д. О.* Корпус параллельных текстов в исследовании культурно-специфичной лексики // Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. — СПб.: Нестор-История, 2009. С. 383–401.
7. *Stubbs M.* Words and phrases: Corpus studies of lexical semantics. — Oxford: Blackwell, 2002. 287 p.
8. Национальный корпус русского языка. <http://www.ruscorpora.ru/index.html>.
9. *Johansson S.* Seeing through multilingual corpora. On the use of corpora in contrastive studies // Studies in Corpus Linguistics. — Amsterdam: Benjamins, 2007. Vol. 26. 378 p.
10. *Stenström A.-B.* Youngspeak: Spanish vale and English okay // Advances in corpus-based contrastive linguistics. — Amsterdam: Benjamins, 2013. P. 127–137.
11. *De Cock S., Goossens D.* Quantity approximation in English and French business news reporting // Advances in corpus-based contrastive linguistics. — Amsterdam: Benjamins, 2013. P. 139–155.
12. *Lavid J., Arús J., Moratón L.* Thematic variation in English and Spanish newspaper genres // Advances in corpus-based contrastive linguistics. — Amsterdam: Benjamins, 2013. P. 261–286.
13. *Toury G.* Descriptive translation studies and beyond. — Amsterdam: Benjamins, 1995. 311 p.

14. *Capelle B.* Show me your translation and I'll tell you what source language it comes from: Motion verbs in English translated from French vs. German // *Societas Linguistica Europaea, SLE 2011 — 44th Annual Meeting: Book of Abstracts.* — Logroño: Universidad de La Rioja, 2011. P. 53–54.
15. *Rabadan R., Izuierdo M.* A corpus-based analysis of English affixal negation translated into Spanish // *Advances in corpus-based contrastive linguistics.* — Amsterdam: Benjamins, 2013. P. 57–82.
16. *Øverås L.* In search of the third code: An investigation of norms in literary translation // *Meta*, 1998. Vol. 43. No. 4. P. 557–570.
17. *Beeching K.* A parallel corpus approach to investigating semantic change // *Advances in corpus-based contrastive linguistics.* — Amsterdam: Benjamins, 2013. P. 103–125.
18. *Aijmer K., Foolen A., Simon-Vandenberghe A.-M.* Pragmatic markers in translation: A methodological proposal // *Approaches to discourse particles.* — Amsterdam: Elsevier, 2006. P. 101–114.
19. *Ebeling J., Ebeling S. O., Hasselgård H.* Using recurrent word-combinations to explore cross-linguistic differences // *Advances in corpus-based contrastive linguistics.* — Amsterdam: Benjamins, 2013. P. 177–199.
20. *Herriman J.* The extraposition of clausal subjects in English and Swedish // *Advances in corpus-based contrastive linguistics.* — Amsterdam: Benjamins, 2013. P. 233–259.
21. *Egan T.* Tertia comparationis in multilingual corpora // *Advances in corpus-based contrastive linguistics.* — Amsterdam: Benjamins, 2013. P. 7–24.
22. *Krzeszowski T. P.* Contrasting languages: The scope of contrastive linguistics. — Berlin: Mouton de Gruyter, 1990. 286 p.
23. *Simon-Vandenberghe A. M.* English adverbs of essence and their equivalents in Dutch and French // *Advances in corpus-based contrastive linguistics.* — Amsterdam: Benjamins, 2013. P. 83–102.
24. *Viberg Å.* Seeing the lexical profile of Swedish through multilingual corpora // *Advances in corpus-based contrastive linguistics.* — Amsterdam: Benjamins, 2013. P. 25–56.
25. *Brown D. P., Tiberius C., Chumakina M., Corbett G. G., Krasovitsky A.* Databases designed for investigating specific phenomena // *The use of databases in cross-linguistic studies. Empirical approaches to language typology.* — Berlin – New York: Mouton De Gruyter, 2009. Vol. 41. P. 117–154.
26. *Бунтман Н. В., Зализняк А. А., Зацман И. М., Кружков М. Г., Лощилова Е. Ю., Сичинава Д. В.* Информационные технологии корпусных исследований: Принципы построения кросслингвистических баз данных // *Информатика и её применения*, 2014. Т. 8. Вып. 2. С. 98–110.
27. *Zatsman I., Buntman N., Kruzikov M., Nuriev V., Zalizniak A. A.* Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery // *15th Conference (European) on Knowledge Management Proceedings.* — Reading, MA, USA: Academic Publishing International, 2014. P. 1063–1071.
28. *Kruzikov M. G., Buntman N. V., Loshchilova E. Ju., Sitchinava D. V., Zalizniak A. A., Zatsman I. M.* A database of Russian verbal forms and their French translation equivalents // *Компьютерная лингвистика и интеллектуальные технологии*, 2014. Вып. 13(20). С. 284–297.

29. Мамонтова В. В. Корпус параллельных текстов и база данных как инструменты исследования переводческих соответствий // Вестник Пятигорского государственного лингвистического ун-та, 2008. № 3. С. 124–127.

Поступила в редакцию 05.03.15

INFORMATION RESOURCES FOR CONTRASTIVE STUDIES: ELECTRONIC TEXT CORPORA

M. G. Kruzhkov

Institute of Informatics Problems, Federal Research Center “Computer Science and Control,” Russian Academy of Sciences, 44-2 Vavilov Str., Moscow 119333, Russian Federation

Abstract: This article presents information resources used in contrastive linguistic studies and their principle features. There are two main types of such information resources: typological databases and electronic text corpora. This paper is focused on the latter. There are two types of corpora, which are particularly relevant for contrastive studies: comparable corpora — balanced collections of original texts in the languages compared and parallel (translation) corpora — collections of original texts in one of the compared languages aligned with their translations into other compared language(s). In addition to description of the existing information resources of contrastive linguistic studies, this paper introduces a new type of such resources, which are termed here as corpus extension databases. The article outlines features of such databases in comparison to electronic corpora and justifies the necessity for creating them.

Keywords: linguistic studies; databases; typological databases; comparable corpora; parallel corpora; corpus extension databases

DOI: 10.14357/08696527150209

Acknowledgments

The work was performed with financial support of the Russian Foundation for Basic Research (grant No. 13-06-00403).

References

1. Kruzhkov, M. G. 2015. Informatsionnye resursy kontrastivnykh lingvisticheskikh issledovaniy: Tipologicheskie bazy dannykh [Information resources for contrastive studies: Typological Databases]. *Sistemy i Sredstva Informatiki — Systems and Means of Informatics* 25(1):198–212.
2. Gak, V. G. 1989. O kontrastivnoy lingvistike [On contrastive linguistics]. *Novoe v zarubezhnoy lingvistike. Vyp. XXV. Kontrastivnaya lingvistika* [New developments in foreign linguistics. Vol. XXV. Contrastive linguistics]. Moscow: Progress. 5–17.

3. Dobrovol'skiy, D. O., A. A. Kretov, and S. A. Sharov. 2005. Korpus parallel'nykh tekstov [Corpus of parallel texts]. *Nauchnaya i Tekhnicheskaya Informatsiya* [Scientific and Technical Information]. Ser. 2: Informatsionnye protsessy i sistemy [Informational processes and systems]. 6:16–27.
4. Dobrovol'skiy, D. O., A. A. Kretov, and S. A. Sharov. 2005. Korpus parallel'nykh tekstov: Arkhitektura i vozmozhnosti ispol'zovaniya [Corpus of parallel texts: Architecture and applications] *Natsional'nyy korpus russkogo yazyka: 2003–2005* [Russian National Corpus: 2003–2005]. Moscow: Indrik. 263–296.
5. Andreeva, E. G., and V. B. Kasevich. 2005. Grammatika i leksika (na materiale anglo-russkogo korpusa parallel'nykh tekstov) [Grammar and lexis (based on the English-Russian parallel corpus)]. *Natsional'nyy korpus russkogo yazyka: 2003–2005* [Russian National Corpus: 2003–2005]. Moscow: Indrik. 297–307.
6. Dobrovol'skiy, D. O. 2009. Korpus parallel'nykh tekstov v issledovanii kul'turno-spetsifichnoy leksiki [Using a parallel corpus for investigation of culture-specific lexis] *Natsional'nyy korpus russkogo yazyka: 2006–2008. Novye rezul'taty i perspektivy* [Russian National Corpus: 2006–2008. New results and prospects]. St. Petersburg: Nestor-Istoriya. 383–401.
7. Stubbs, M. 2002. *Words and phrases: Corpus studies of lexical semantics*. Oxford: Blackwell. 287 p.
8. Natsional'nyy korpus russkogo yazyka [Russian National Corpus]. Available at: <http://www.ruscorpora.ru/index.html> (accessed March 5, 2015).
9. Johansson, S. 2007. Seeing through multilingual corpora. On the use of corpora in contrastive studies. *Studies in corpus linguistics*. Amsterdam: Benjamins. Vol. 26. 378 p.
10. Stenström A.-B. 2013. Youngspeak: Spanish vale and English okay. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 127–137.
11. De Cock, S., and D. Goossens. 2013. Quantity approximation in English and French business news reporting. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 139–155.
12. Lavid, J., J. Arús, and L. Moratón. 2013. Thematic variation in English and Spanish newspaper genres. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 261–286.
13. Toury, G. 1995. *Descriptive translation studies and beyond*. Amsterdam: Benjamins. 311 p.
14. Capelle, B. 2011. Show me your translation and I'll tell you what source language it comes from: Motion verbs in English translated from French vs. German. *Societas Linguistica Europaea 44th Annual Meeting in Logroño: Book of Abstracts*. Logroño: Universidad de La Rioja. 53–54.
15. Rabadan, R., and M. Izquierdo. 2013. A corpus-based analysis of English affixal negation translated into Spanish. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 57–82.
16. Øverås, L. 1998. In search of the third code: An investigation of norms in literary translation. *Meta* 43(4):557–570.
17. Beeching, K. 2013. A parallel corpus approach to investigating semantic change. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 103–125.

18. Aijmer, K., A. Foolen, and A.-M. Simon-Vandenberg. 2006. Pragmatic markers in translation: A methodological proposal. *Approaches to discourse particles*. Amsterdam: Elsevier. 101–114.
19. Ebeling, J., S.O. Ebeling, and H. Hasselgård. 2013. Using recurrent word-combinations to explore cross-linguistic differences. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 177–199.
20. Herriman, J. 2013. The extraposition of clausal subjects in English and Swedish. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 233–259.
21. Egan, T. 2013. Tertia comparationis in multilingual corpora. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 7–24.
22. Krzeszowski, T. P. 1990. *Contrasting languages: The scope of contrastive linguistics*. Berlin: Mouton de Gruyter, 1990. 286 p.
23. Simon-Vandenberg, A. M. 2013. English adverbs of essence and their equivalents in Dutch and French. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 83–102.
24. Viberg, Å. 2013. Seeing the lexical profile of Swedish through multilingual corpora. *Advances in corpus-based contrastive linguistics*. Amsterdam: Benjamins. 25–56.
25. Brown, D. P., C. Tiberius, M. Chumakina, G. G. Corbett, and A. Krasovitsky. 2009. Databases designed for investigating specific phenomena. *The use of databases in cross-linguistic studies. Empirical approaches to language typology*. Berlin–New York: Mouton De Gruyter. 41:117–154.
26. Buntman, N. V., A. A. Zaliznyak, I. M. Zatsman, M. G. Kruzhkov, E. Yu. Loshchilova, and D. V. Sichinava. 2014. Informatsionnye tekhnologii korpusnykh issledovaniy: Printsipy postroeniya krosslingvisticheskikh baz dannykh [Information technologies for corpus studies: Underpinnings for cross-linguistic database creation]. *Informatika i ee Primeneniya — Inform. Appl.* 8(2):98–110.
27. Zatsman, I., N. Buntman, M. Kruzhkov, V. Nuriev, and A. A. Zalizniak. 2014. Conceptual framework for development of computer technology supporting cross-linguistic knowledge discovery. *15th Conference (European) on Knowledge Management Proceedings*. Reading, MA: Academic Publishing International. 1063–1071.
28. Kruzhkov, M. G., N. V. Buntman, E. Ju. Loshchilova, D. V. Sichinava, A. A. Zalizniak, and I. M. Zatsman. 2014. A database of Russian verbal forms and their French translation equivalents. *Komp'yuternaya Lingvistika i Intellektual'nye Tekhnologii* [Computer Linguistics and Intellectual Technologies] 13(20):284–297.
29. Mamontova, V. V. 2008. Korpus parallel'nykh tekstov i baza dannykh kak instrumenty issledovaniya perevodcheskikh sootvetstviy [Parallel corpus and database as tools for investigating translation correspondences]. *Vestnik Pyatigorskogo Gosudarstvennogo Lingvisticheskogo Un-ta* [Newsletter of Pyatigorsk State Linguistic University] 3:124–127.

Received March 05, 2015

Contributor

Kruzhkov Mikhail G. (b. 1975) — leading programmer, Institute of Informatics Problems, Federal Research Center “Computer Science and Control” of Russian Academy of Sciences, Moscow 119333, 44-2 Vavilov Str., Russian Federation; magnit75@yandex.ru