

**О. Ю. Инькова**

*Женевский университет; ИПИ ФИЦ ИУ РАН  
(Женева, Швейцария; Москва, Россия)  
Olga.Inkova@unige.ch*

## **СИСТЕМА СВЯЗУЮЩИХ СРЕДСТВ РУССКОГО ЯЗЫКА: СОСТАВ И МЕТОДЫ ОПИСАНИЯ**

В статье рассматриваются вопросы определения состава связующих средств русского языка (коннекторов), которые представляют собой динамическую систему как в количественном, так и в качественном аспекте. Автор рассматривает существующие подходы к решению данной проблемы, в частности, то, как решается проблема неоднословных коннекторов в аннотированных корпусах русского языка на примере ХАНКО и «Корпусного словаря неоднословных лексических единиц» НКРЯ. В новом информационном ресурсе — надкорпусной базе данных (НБД) коннекторов — единицей аннотирования и анализа является *речевая реализация* (РР), т. е. та форма, в которой коннектор употреблен в конкретном контексте. При фиксации РР в НБД входящие в состав неоднословных коннекторов элементы приписываются к перекрестным кластерам. Такой подход позволяет не только зафиксировать реально встречающиеся формы показателей связи, но и наглядно представить возможности сочетаемости того или иного элемента, входящего в их состав, т. е. системные связи, существующие между элементами неоднословных показателей. Эти данные могут быть использованы для описания потенциала языковых средств, которым располагает говорящий, предсказывая возможность того или иного сочетания элементов в одной РР. Статистические возможности НБД позволяют, в свою очередь, составить представление о количестве и частоте употребления того или иного показателя связи в тексте.

*Ключевые слова:* коннекторы, надкорпусные базы данных, формальное варьирование, русский язык, корпусная лингвистика, количественные методы анализа.

Связующие средства русского языка, как бы их ни называли: союзы, аналоги союзов, скрепы, коннекторы, ... — представляют собой динамическую систему, как в количественном, так и в качественном аспектах. Что касается количественного аспекта, то эта система постоянно пополняется за счет новых языковых единиц, тогда как некоторые единицы, наоборот, исчезают. Так, в последнее время широкое распространение получили коннекторы с формантом *при*, особенно *при этом*

и *притом что*; они входят в словари только в XX веке [Инькова 2015, 2018а]. Другие единицы выходят из употребления или приобретают стилистическую окраску, ограничивающую их использование; ср. *коли, доколи, диви бы, зане* и др.

Что касается качественного аспекта, то динамичность системы обусловлена во многом тем, что связующие средства русского языка характеризуются высокой степенью формальной вариативности. Многие из них состоят более чем из одного графического слова или представляют собой 'разрывные' языковые единицы (*хотя... но, не только... а даже* и др.). Они строятся как бы из кирпичиков, которые могут вступать друг с другом в самые разнообразные сочетания<sup>1</sup>; ср., например, *да и, да притом и, да вдобавок и, да притом еще и, да еще притом* и т. д. или группы коннекторов непосредственного следования, которые образуют как многоэлементные коннекторы (*только, лишь только, как только, едва лишь только* и др.), так и двухкомпонентные (*как только... так сразу, едва лишь только... тут же* и др.)<sup>2</sup>.

Такая подвижность системы связующих средств вызывает немалые трудности для ее описания, в первую очередь, в том, что касается определения ее состава. В своей пионерской работе 1987 г. «Очерки по синтаксису сложного предложения» М. И. Черемисина и Т. А. Колосова отмечали, что «[д]о последнего времени не был зафиксирован даже грубо ориентировочный список показателей связи, оставалось неизвестным даже приблизительное их количество» [Черемисина, Колосова 2010: 123]. Эта проблема приобретает особую актуальность в связи с разработкой аннотированных корпусов русского языка и систем автоматической обработки и генерации текстов. В рамках корпусных подходов создаются списки связующих средств с целью облегчить их распознавание и разметку, а потому включающие, как правило, только неоднословные показатели.

Наиболее известные списки для русского языка: «Полный список служебных фразем» Хельсинского корпуса ХАНКО, к сожалению, удаленный со стартовой страницы корпуса [<http://h248.it.helsinki.fi/hanco/index.html>]<sup>3</sup>, и «Корпусной словарь неоднословных лексических единиц» НКРЯ [<http://www.ruscorpora.ru/old/obgrams.html>].

Список ХАНКО был составлен на основе «Толкового словаря служебных частей речи» Т. Ф. Ефремовой [2001], «Толкового словаря словосочетаний, эквивалентных слову» Р. П. Рогожниковой [2003] и сводных таблиц С. И. Богданова и Ю. В. Рыжовой «Русская служебная лексика. Сводные таблицы» [1997], в которых сравнивается наличие той или иной служебной лексемы в целом ряде словарей, в том числе и в БАС, а также в РГ-80. Раздел «Союзы» ХАНКО, включающий только неоднословные единицы, насчитывал 323 единицы.

<sup>1</sup> Т. М. Николаева [Николаева 1985; 2008], описывая структуру русских дискурсивных маркеров, использует метафору калейдоскопа, показывая, как многие русские дискурсивные маркеры, часто воспринимаемые как неразложимые в современном языке, строятся из одних и тех же элементов («блуждающих частиц»), но каждый раз в разных сочетаниях.

<sup>2</sup> Подробнее о коннекторах этого семантического класса и их формальных особенностях см. работы И. М. Кобозевой [Кобозева 2016, 2017].

<sup>3</sup> Список был доступен в 2014 году по ссылке <http://www.helsinki.fi/venaja/russian/e-material/hanco/index.htm>.

Список «Корпусного словаря», как указано на его стартовой странице, составлен на основе базы данных частотных коллокаций НКРЯ с дополнениями из того же словаря Р. П. Рогожниковой и МАС. Раздел «Обороты в функции союза и союзного слова» включает 159 единиц.

Оба списка построены по частеречному принципу. Однако, как известно, определение частеречной принадлежности связующих средств связано с немалыми трудностями, и решения, к которым приходят исследователи, часто не совпадают. По-видимому, по этой причине часть показателей связи, квалифицируемых РГ-80 как «аналоги союзов», находится в «Корпусном словаре» в разделе «Вводные обороты» вместе с единицами, которые не имеют связующей функции; например, *таким образом, иными словами, одним словом*, которые считаются аналогами союзов в ряде своих употреблений, соседствуют с *черт возьми, сохрани боже и слава аллаху*, которые ни в одном из своих употреблений таковыми не являются. Исследователь должен в каждом конкретном случае сам принимать решение о функциональных возможностях той или иной языковой единицы. Поэтому, используя данный словарь, представить себе количество языковых единиц, способных выполнять связующую функцию, не представляется возможным.

Наиболее полный на сегодняшний день список связующих средств русского языка представлен в РГ-80. Он дан в «Предметном указателе» под рубрикой «Союзы и аналоги союзов» (поэтому, строго говоря, не может считаться полноценным списком связующих средств) и насчитывает 660 языковых единиц.

Даже самый поверхностный анализ этих трех списков позволяет увидеть, что они в значительной степени различны не только по количеству включенных в них единиц, но и по принципам их отбора. Результаты их сопоставительного анализа представлены в Таблице 1.

Таблица 1

**Критерии отбора языковых единиц для включения в список связующих средств**

Критерии	ХАНКО	НКРЯ	РГ-80
Неоднословность языковой единицы	+	+	–
Возможность выступать в функции показателя связи	+	+	–
Учет вариантов с запятой	+	–	–
Учет вариантов с усеченной гласной	+	+	–
Возможность дистантного расположения составляющих	–	–	(+)
Наличие двух- и многокомпонентных вариантов показателей связи	+	–	+

Как мы уже упоминали, в списки ХАНКО и НКРЯ включаются только неоднословные показатели связи, в отличие от списка РГ-80, который, в силу специфики его жанра — предметный указатель — отличает и тот факт, что он содержит языковые единицы, которые только упоминаются в разделах, посвященных союзам и их аналогам, как единицы, могущие входить в их состав, но которые не могут выполнять самостоятельно связующей функции. К ним, на наш взгляд, относятся *даже, еще, всё* и др. Однако граница между этими двумя видами употребления языковых единиц — употребление в составе связующего средства и возможность самостоятельного употребления

в связующей функции — проводится недостаточно четко. Вряд ли, например, можно считать, как это предлагает РГ-80 [И: § 2083], *даже* аналогом союза в таком примере:

- (1) Она с тоской ждала восклицаний, тягостных объятий, *даже* упреков [Леон.; пример РГ-80, § 2083].

*Даже* стоит здесь перед вторым членом создаваемого им градационным ряда. Следуя этой логике, придется считать аналогом союза *очень* в (2), где также представлен ряд с градационным значением:

- (2) А было ему здесь хорошо, *очень* хорошо. [Елена Мушкина. Ёлка из кондитерской // «Родина», 2007]

С другой стороны, *даже* и его устаревший и разговорный вариант *ажно* могут употребляться в структурах с причинными отношениями между пропозициями. Но их трактовка как «союзов» и «аналогов союза» следствия (обусловленность следствия чрезмерной степенью выявления признака) также вызывает сомнения:

- (3) Вдруг наверху, над моей головой, кто-то крикнул по все горло, я *ажно* при села [А. Остр.; пример РГ-80, § 3064]

Семантика следствия возникает здесь на основе смыслового соотношения частей: если убрать частицу *ажно* из (3), отношение следствия между двумя положениями вещей сохранится. Как в (2), так и в (3), *даже/ажно* придают отношению между частями высказывания градационную семантику, но само отношение выводится на основе смыслового соотношения между ними.

Во все три списка включены языковые единицы *что касается, что до* и др., для которых статус показателя связи и еще в большей степени статус коннектора требует теоретического обоснования.

Только список ХАНКО содержит варианты с запятой для составных союзов, которые могут иметь два написания, причем как с изменением условий употребления (для *потому что, оттого что* и др.), так и с одинаковыми условиями употребления (*вместо того чтобы* и др.).

Варианты с усеченной гласной (*если бы* и *если б* и др.) не представлены только в списке РГ-80, хотя в нем и фигурируют схожие по своей природе формы *или* и *иль, ли* и *ль*. И, наоборот, возможность дистантного расположения элементов неоднословного союза упоминается только в РГ-80, в частности для *да и*, но это скорее исключение, чем правило, и в списке для других единиц, которые допускают такую возможность, вариант с дистантным расположением не указан. Ср. возможность такого варианта для *хотя и* в (5), где *хотя* и *и* разделены словом *закон*, в отличие от (4), где они расположены контактно.

- (4) Понимание этого время от времени *хотя и* появляется в человеческой голове («хоть горшком назови...»), по природе своей является, вероятно, идеей «рецессивной». [Михаил Арапов. Когда текст обретает смысл // «Знание — сила», 2003]

- (5) И клирики, кстати, имеют на это право: *хотя* закон и запрещает им агитировать с амвона, он не может запретить этого в частных разговорах. [Александр Верховский. Между паствой и электоратом (2003) // «Еженедельный журнал», 2003.04.08]

Наконец, двух- и многокомпонентные показатели связи отсутствуют в списке НКРЯ, хотя для многих логико-семантических отношений это единственно возможный показатель; ср. в этом отношении *если... то* сопоставления в (6) и *как... так* и аналогии (7):

- (6) *Если* авторы авантюрного романа, вводя трущобы, каторги и больницы, действительно подготовили путь социальному роману, *то* перед Достоевским были образцы подлинного социального романа <...> [М. М. Бахтин. Проблемы поэтики Достоевского (1963)]
- (7) *Как* сел на стул, придя к Соне, *так* и сидел, не двигаясь, до нужной секунды: ни раньше, ни позже. [Юрий Трифонов. Дом на набережной (1976)]

Заметим также, что в НКРЯ в неоднословные лексемы необоснованно попадают сочетания, где составляющие не являются единым показателем связи; ср. *и хотя*, которое дается в списке неоднословных союзов, с отсылкой к такому примеру:

- (8) Несмотря на то, что уже два года на бумаге существует федеральная программа по ликвидации ветхого и аварийного жилья, по оценке Владимира Яковлева, она реализуется слабо. *И хотя* Астрахань находится в тройке лидеров по объёмам «аварийки», в России есть ещё полтора десятка регионов, где нам придётся ломать голову над решением подобной проблемы. [Леонид Барков. Вице-премьера позвали в трущобы (2003) // «Время МН», 2003.08.06]

Эта структура должна анализироваться следующим образом: *р. И [хотя r, s]q*. Иначе говоря, *и* и *хотя* соединяют здесь разные фрагменты текста, и их соседство объясняется лишь линейной структурой текста.

Все три списка, несмотря на указанные различия, отличает, тем не менее, определенная субъективность при отборе языковых единиц и алфавитный порядок их представления. Это не позволяет увидеть ни механизма образования неоднословных показателей связи, в частности того, что один и тот же «конкретизатор» может сочетаться с разными союзами, ни реального состава связующих средств русского языка.

Иной подход к анализу состава связующих средств русского языка представлен в уже упоминаемой нами работе М. И. Черемисиной и Т. А. Колосовой [Черемисина, Колосова 2010]. Для того чтобы дать наглядное представление о способности союзов сочетаться с конкретизаторами, авторы составили таблицу [Там же: 150–152], в левой части которой указаны союзы, как сочинительные, так и некоторые подчинительные (*если, когда, потому что, хотя* и др.), а в верхней части — конкретизаторы (*в особенности, в результате, в то же время, вдобавок* и др.). Несмотря на условность терминологии, такой способ представления позволяет увидеть продуктивность того или иного союза или конкретизатора в образовании

неоднословных показателей связи. Но, поскольку таблица создана в докорпусную эпоху (авторы работали с картотекой, содержащей 1000 примеров), то она, конечно, не может отразить реальную картину. Кроме того, в ней отсутствуют двух- и много компонентные союзы, которые также характеризуются высокой степенью формальной вариативности.

Надкорпусная база данных (НБД) коннекторов, созданная в ИПИ ФИЦ ИУ РАН именно с целью анализа их формальной вариативности, в значительной степени построена с учетом подхода, разработанного М. И. Черемисиной и И. А. Колосовой, и с учетом недостатков существующих списков показателей связи, а также с использованием возможностей современных информационных технологий. НБД коннекторов функционирует на основе текстов параллельных (французского и итальянского) подкорпусов НКРЯ, общий объем которых составляет 9,2 млн словоупотреблений, и позволяет лингвистам аннотировать употребления коннекторов по заданной ими системе параметров и хранить результаты этой работы в виде структурированных двуязычных аннотаций. Несмотря на ограниченный размер корпуса, НБД позволяет получить представительную статистику, характеризующую структуру коннекторов или тот или иной аспект их функционирования (подробнее о концепции НБД и ее статистических возможностях см. [Инькова 2018б]; [Inkova, Popkova 2017]).

В предлагаемом нами подходе единицей аннотирования и последующего анализа коннекторов становится речевая реализация (РР), т. е. та форма, в которой коннектор употребляется в конкретном контексте. При этом учитываются, т. е. фиксируются как разные РР, варианты с запятой (например, *потому что* и *потому, что*; *после того как* и *после того, как*), варианты с усеченной гласной (*если бы* и *если б*, *для того чтобы* и *для того чтоб*), варианты с дистантным расположением элементов (*да и* и *да|и*, *и в частности* и *и|в частности*), с разным порядком следования элементов (*лишь только* и *только лишь*). Таким образом, при аннотировании мы идем не от словаря к тексту, а от текста к списку связующих средств, максимально полно отражающему их состав в русском языке. На 12.05.2019 в НБД зафиксировано 1287 РР, т. е. уже в два раза больше, чем в наиболее полном списке РГ-80, и это притом, что наш список и список РГ-80 совпадают лишь частично, поскольку некоторые семантические классы коннекторов в НБД пока не аннотировались или аннотировались частично.

На первом этапе аннотирования, каждая регистрируемая РР приписывается к двум типам кластеров.

#### 1. Кластеры, учитывающие состав РР

В четырех кластерах этого типа учитывается:

- а) количественный состав РР (один или более элементов, входящих в ее состав),
- б) наличие показателя ЛСО в одном или в каждом из соединяемых фрагментов текста.

По признаку а) все РР разделяются на однословные и неоднословные. По признаку б) все РР разделяются на те, что маркируют один из соединяемых фрагментов, и те, что маркируют каждый из соединяемых фрагментов, которых может быть два или более. Части РР, которые находятся в каждом из соединяемых

фрагментов текста, предлагается называть *компонентами РР*. Соответственно, все РР разделяются на *однокомпонентные* и *n-компонентные*, а *n-компонентные* — на *двух-* и *многокомпонентные*.

В предлагаемой терминологии — однокомпонентные, двухкомпонентные и многокомпонентные РР — можно увидеть параллель с традиционной классификацией, которая разделяет союзные образования на одноместные, двухместные и многоместные. Эти термины кажутся нам менее удачными, поскольку используются также в другой сфере лингвистических исследований — теории валентностей. Кроме того, в разряд многоместных, а в нашей терминологии, многокомпонентных, русская грамматическая традиция включает только симметричные союзные соединения (*и... и, или... или, не то... не то...* и др.), тогда как в НБД зафиксированы и асимметричные РР, т. е. такие, где повторяется либо первый, либо второй компонент, либо оба; ср. (9) с РР *не только||не только||не только||а еще и||и*<sup>4</sup>:

- (9) И *не только* потому, что размещался он в двух больших залах со сводчатыми потолками, расписанными лиловыми лошадьми с ассирийскими гривами, *не только* потому, что на каждом столике помещалась лампа, накрытая шалью, *не только* потому, что туда не мог проникнуть первый попавшийся человек с улицы, *а еще и* потому, что качеством своей провизии Грибоедов бил любой ресторан в Москве, как хотел, *и* что эту провизию отпускали по самой сходной, отнюдь не обременительной цене [М. А. Булгаков. Мастер и Маргарита (ч. 1) (1929–1940)]

Такие РР остаются, как правило, вне поля зрения исследователей, но именно такого рода образования показывают, что форму, в которой коннектор употребляется в тексте, его РР, следует рассматривать как единицу не языка, а речи. Эта форма создается каждый раз *ex novo* в зависимости от коммуникативного намерения говорящего, синтаксической и семантической структуры высказывания и целого ряда других факторов.

Показательно распределение РР по этим четырем структурным кластерам:

- одноэлементные РР (*и, а, но, так* и др.) — 4,62 %;
- многоэлементные РР (*и особенно, а|наоборот, да вот хоть|например* и др.) — 48,5 %;
- двухкомпонентные РР (*во-первых||а во-вторых, если бы даже||то и тогда бы, если бы даже||то|все-таки* и др.) — 30,65 %;
- многокомпонентные РР (*или||или||или, зато||зато||зато, в дальнейшем||а затем||и, наконец* и др.) — 4,62 %<sup>5</sup>.

<sup>4</sup> Используемая при обозначении РР коннекторов двойная вертикальная черта «||» указывает на то, что коннектор состоит из двух или более расположенных дистантно компонентов, каждый из которых вводит фрагмент текста; одиночная вертикальная черта «|» говорит о том, что языковые единицы (элементы), составляющие коннектор или его компонент, разделены текстом, но находятся в пределах одного и того же фрагмента.

<sup>5</sup> Совпадение количества одноэлементных и многокомпонентных РР в НБД является случайным. По мере наполнения НБД число этих РР может измениться.



Интересно, однако, не только число РР, но и число употреблений РР разных структурных типов, т. е. количество аннотаций, в которых были зарегистрированы эти РР. По этому критерию распределение РР следующее:

- одноэлементные РР — 46,58 %;
- многоэлементные РР — 37,38 %;
- двухкомпонентные РР — 15,32 %;
- многокомпонентные РР — 0,72 %.

Эти данные говорят о том, что наиболее частотными являются одно- и многоэлементные РР, тогда как многокомпонентные РР можно считать периферийным явлением в системе связующих средств русского языка. Кроме того, почти все они зафиксированы с единичными употреблениями, что вполне объяснимо.

## 2. Перекрестные кластеры

Четыре кластера первого типа, отражая важный классификационный признак коннекторов (их количественный состав), не учитывают тот факт, что одна и та же языковая единица может входить в различные сочетания. Поэтому описание состава многоэлементных, двух- и многокомпонентных РР, а также сочетаемости составляющих их компонентов или элементов решается в НБД через систему «перекрестных», или «пересекающихся», кластеров второго типа. Они позволяют разметчикам одновременно помещать аннотируемую РР в разные кластеры и пополняются в процессе аннотирования новых контекстов употребления РР. См. Таблицы 2 и 3.

Такой подход позволяет не только зафиксировать реально встречающиеся формы показателей связи, но и наглядно представить возможности сочетаемости того или иного элемента, входящего в их состав. В качестве примера приведу союз *или*. В списке НКРЯ для него не дано ни одной многоэлементной РР; в списке ХАНКО — только формы *будь то... или* и *или...*, *или (иль..., иль)*. В РГ-80 дано 14 форм: *будь то... или; или; или вернее; или иначе; или иначе говоря; или наоборот; или по крайней мере; или точнее; или... или; или... но; или/или... но; иль; ли... или; либо... или; может быть... или; может быть... или, может быть*.

Создается впечатление, что, по сравнению с другими сочинительными союзами, в частности союзами *и* и *а*, союз *или* в меньшей степени способен образовывать неоднословные РР. Однако в НБД кластер *или* содержит уже 64 РР, включающие элемент *или*: *а то||или; будь то||или; будь то||или же; или; или вернее; или во всяком случае; или вообще; или все же; или гораздо правильнее будет сказать; или даже; или даже еще лучше; или еще; или же; или же просто; или же, к примеру; или же, наконец; или же|просто; или лучше; или лучшие сказать; или по крайней мере; или по меньшей мере; или просто; или просто и; или скорее всего; или точнее; или уж; или хотя бы; или хотя бы просто; или, в крайнем случае; или, в противном случае; или, вернее; или, выражаясь точнее; или, говоря adv; или, еще лучше; или, иначе говоря; или, лучше; или, лучше сказать; или, может; или, может быть; или, наоборот; или, например; или, напротив; или, попросту; или, скажем иначе; или, скорее; или, так сказать; или, точнее; или||а не то; или||или;*



Таблица 2

**Примеры РР с указанием кластеров первого и второго типов, в которые они входят**

	Речевые реализации (РР)		
	<i>а впрочем</i>	<i>а, наоборот, скорее</i>	<i>или а не то</i>
Кластеры РР	«Многоэлементные»	«Многоэлементные»	«Двухкомпонентные»
	«а»	«а»	«а»
	«впрочем»	«наоборот»	«а не то»
		«скорее»	«или»
			«не»
		«то»	

Таблица 3

**Примеры кластеров второго типа и входящих в них РР**

	Кластеры РР		
	« <i>между тем</i> »	« <i>значит</i> »	« <i>зато</i> »
РР, входящие в указанные кластеры	<i>а между тем</i>	<i>а значит</i>	<i>зато</i>
	<i>а между тем</i>	<i>если  значит</i>	<i>зато  зато</i>
	<i>меж тем как</i>	<i>если  то значит</i>	<i>зато  зато  зато</i>
	<i>между тем</i>	<i>если только  значит</i>	<i>зато при этом</i>
	<i>между тем как</i>	<i>значит</i>	<i>конечно  зато</i>
	<i>но между тем</i>	<i>значит  коли</i>	<i>но зато</i>
	<i>но между тем</i>	<i>и, значит</i>	<i>хоть  да зато</i>
	<i>хотя  а между тем</i>	<i>коли  то значит</i>	<i>хоть и  зато</i>
	<i>хотя  но между тем</i>		<i>хоть  зато  зато  зато</i>
	<i>хотя  и  но между тем</i>		<i>хоть  но зато</i>
			<i>хоть и  зато</i>
			<i>хоть и  но зато</i>

*или||или||или; или||или же; или||или||или, наконец; или||или||но, во всяком случае; или же||или||или; или|все-таки; или|например; иль; иль, может быть; ли||или; ли||или, наоборот; ли|ли|ли|ли|ли||или; либо||или; например, или; то есть, или еще лучше сказать.*

РР *или||но*, приводимая РГ-80, в НБД пока не зафиксирована, а *или||или||но* зафиксирована в форме *или||или||но, во всяком случае*.

Система перекрестных кластеров позволяет также увидеть, какие РР являются общими для двух или нескольких кластеров РР. На рисунке 1 мы приводим эту информацию для кластеров «да» и «еще», которые в НБД имеют 10 общих, т. е. включающих оба элемента, РР.



**Рисунок 1. Общие элементы для кластеров «да» и «еще» в НБД.**

Семантический анализ элементов, входящих в состав РР, позволяет, в свою очередь, установить, какие элементы сочетаются между собой. Так, в РР с союзом *или* присутствуют элементы с семантикой разделительности (*а то и, либо, ли*), предположительности (*будь то, может быть*), аддитивности (*еще*), градации (*даже, в крайнем случае, по крайней мере*), противопоставления (*наоборот*), спецификации (*например*) или генерализации (*вообще*), комментария выбора формулировки (*лучше сказать, иначе говоря, вернее, точнее, скорее*), но отсутствуют, например, показатели причинно-следственных или временных отношений. Эти данные могут быть использованы для описания потенциала языковых средств, которым располагает говорящий, предсказывая возможность того или иного сочетания элементов в одной РР.

Таким образом, предлагаемый подход позволяет фиксировать все реально встретившиеся формы, в которых употребляются связующие средства русского языка, а разработанный информационный лингвистический ресурс — НБД коннекторов — является достаточно эффективным инструментом, позволяющим это сделать. Кроме того, система перекрестных кластеров дает возможность увидеть системные связи между элементами, образующими неоднословные показатели, а статистические возможности НБД — составить представление об их количестве и частоте употребления в тексте. Разработанный метод описания структуры показателей связи приближает нас к цели, обозначенной в работе [Черемисина, Колосова 2010]: создание списка связующих средств русского языка во всем их многообразии. Представляется, однако, очевидным, что при такой динамичности системы показателей связи речь может идти только об открытом списке, состав которого будет меняться, с одной стороны, по мере регистрации новых РР, а с другой — по мере исчезновения из языка некоторых языковых единиц. Подчеркнем также, что РР являются не единицами языка, т. е. заслуживающими отдельной словарной статьи, а единицами речи, создающимися говорящим в соответствии с его коммуникативным заданием.

## Литература

БАС — Чернышев В. И. (ред.). Словарь современного русского литературного языка. В 17 т. М./Л.: Изд-во АН СССР, 1948–1965.

Богданов С. И., Рыжова Ю. В. Русская служебная лексика. Сводные таблицы. СПб: СПбГУ, 1997. 168 с.

Ефремова Т. Ф. Толковый словарь служебных частей речи русского языка. М.: Астрель-Аст, 2004. 814 с.

Инькова О. Ю. *При том что*: от морфологии и синтаксиса к семантике // Типология морфосинтаксических параметров. Материалы международной конференции «Типология морфосинтаксических параметров 2015». Вып. 2, М.: МПГУ, 2015. С. 124–149.

Инькова О. Ю. Коннекторы русского языка с формантом *при*: корпусное исследование // Russian Linguistics. 2018(a). № 2. С. 159–190.

Инькова О. Ю. Лингвоспецифичность коннекторов: методы и параметры описания // Семантика коннекторов: контрастивное исследование / О. Ю. Инькова (ред.). М.: Торус Пресс, 2018(б). С. 5–23.

Кобозева И. М. Когнитивно-семантический подход к описанию средства связи предложений (на примере коннекторов со значением непосредственного следования) // Труды Института русского языка им. В. В. Виноградова. 2016. № 10. С. 120–133.

Кобозева И. М. Коннекторы контактного предшествования во французском и русском языках по данным параллельного корпуса // Съпоставително езиковзнание. 2017. № 4. С. 48–62.

МАС — Словарь русского языка: В 4 т. / Под ред. А. П. Евгеньевой. М.: Русский язык, 1981.

Николаева Т. М. Функции частиц в высказывании. М.: Наука, 1985. 170 с.

Николаева Т. М. Непарадигматическая лингвистика (история «блуждающих частиц»). М.: Языки славянских культур, 2008. 375 с.

РГ-80 — Русская грамматика / Под ред. Н. Ю. Шведовой. М.: Наука, 1980.

Рогожникова Р. П. Толковый словарь сочетаний, эквивалентных слову. М.: Астрель-АСТ, 2003. 416 с.

Черемисина М. И., Колосова Т. А. Очерки по теории сложного предложения. Изд. 2-е. (1-е изд. Новосибирск: Наука, 1987). М.: УРСС, 2010. 226 с.

Inkova O., Popkova N. Statistical data as information source for linguistic analysis of Russian connectors // Информатика и ее применение. 2017. № 3. С. 123–131.

**Olga Yu. Inkova**

*University of Geneva / Institute of Informatics Problems FRC CSC RAS  
(Geneva, Switzerland; Moscow, Russia)  
Olga.Inkova@unige.ch*

## **SYSTEM OF CONJUNCTIVE TOOLS IN RUSSIAN: STRUCTURE AND DESCRIPTION**

The paper analyses the formal structure of conjunctive tools (connectors) in Russian, units acting as elements of a dynamic system both quantitatively and qualitatively. The author examines how the problem of annotating multiword connectors is addressed in corpus projects, including two annotated corpora of Russian language: HANCO (the Helsinki Annotated Corpus), and “Corpus Dictionary of Multiword Lexical Units” of the Russian National Corpus. The new corpus tool called a supracorpora database (SCDB) of connectors considers discursive realization (DR), e.g. the form in which a connector has been used in a certain context, as an annotation unit. During the DR annotation process in the SDC, elements of the multiword connector’s structure are attributed to cross-clusters. This approach not only allows one to register the *ad hoc* form of a given conjunctive tool, but also gives an idea of the combinatorial capacity of multiword connectors’ components and therefore reveals systematic ties between them. These data can be used to describe the range of linguistic units available to the speaker, who is free to combine some of them to produce a DR. Diverse statistic data extractible from the SDC, on the other hand, help to ascertain the number of occurrences and the frequency of a given conjunctive tool in a text.

*Keywords:* connectives, supracorpora databases, formal diversity, Russian, corpus linguistics, quantitative methods.

### **References**

BAS — Chernyshev V. I. (ed.) *Slovar' sovremennogo russkogo literaturnogo yazyka*. V 17 t. [Dictionary of Modern Russian literary language. In 17 t.]. Moscow / Leningrad, USSR Academy of Sciences publishing house, 1948–1965. (In Russ.)

Bogdanov S. I., Ryzhova Yu. V. *Russkaya sluzhebnyaya leksika. Svodnye tablitsy* [Russian grammatical words dictionary. Summary tables]. St. Petersburg, St. Petersburg State University, 1997. 168 p. (In Russ.)

Efremova T. F. *Tolkovyi slovar' sluzhebnykh chastei rechi russkogo yazyka* [Explanatory dictionary of the Russian auxiliary parts of speech]. Moscow, «Astrel'-Ast» Publ., 2004. 814 p. (In Russ.)

Inkova O. Yu. [Pri tom chto: from morphology and syntax to semantics]. *Tipologiya morfosintaksicheskikh parametrov. Materialy mezhdunarodnoi konferentsii «Tipologiya morfosintaksicheskikh parametrov 2015»* [Typology of morphosyntactic parameters. Proc. of the international conference “Typology of morphosyntactic parameters 2015”]. Moscow, Moscow State Pedagogical University. 2015, no. 2, pp. 124–149. (In Russ.)

Inkova O. Yu. [The connectors of Russian with the element *pri*: a corpus-based study], *Russian Linguistics*, 2018(a) no. 2, pp. 159–190. (In Russ.)

Inkova O. Yu. [Language-specificity of connectors: methods and description parameters]. *Semantika konnektorov: kontrastivnoe issledovanie* [Connective semantics: a contrastive study], O. Yu. Inkova ed. Moscow, «TORUS PRESS» Publ., 2018b, pp. 5–23. (In Russ.)

Inkova O., Popkova N. Statistical data as information source for linguistic analysis of Russian connectors. *Informatika i ee primeneniye*, 2017, no. 3, pp. 123–131.

Kobozeva I. M. [Cognitive-semantic approach to the description of the means of linking of sentences (on the example of connectives with the meaning of direct follow)], *Trudy Instituta russkogo yazyka im. V. V. Vinogradova*, 2016, no. 10, pp. 120–133. (In Russ.)

Kobozeva I. M. [Contact precedence connectives in French and Russian according to the parallel corpus]. *Sъpostavitelno ezikoznanie*, 2017, no. 4, pp. 48–62. (In Russ.)

MAS — *Slovar' russkogo yazyka* [Dictionary of the Russian language]: In 4 vol. Evgen'eva A. P. (ed.). Moscow, Russkii Yazyk, 1981. (In Russ.)

Nikolaeva T. M. *Funktsii chastits v vyskazyvanii* [Functions of particles in the utterance]. Moscow, «Nauka» Publ., 1985. (In Russ.)

Nikolaeva T. M. *Neparadigmaticheskaya lingvistika* (istoriya «bluzhdayushchikh chastits») [Non-paradigmatic linguistics (history of the «wandering particles»)]. Moscow, «Yazyki Slavyanskikh Kul'tur» Publ., 2008. (In Russ.)

RG-80 — *Russkaya grammatika* [Russian grammar]. Shvedova N. Yu. (ed.). Moscow, «Nauka» Publ., 1980. (In Russ.)

Rogozhnikova R. P. *Tolkovyi slovar' sochetanii, ekvivalentnykh slovu* [Explanatory dictionary of constructions equivalent to the word]. Moscow, «Astrel'-AST» Publ., 2003. 416 p. (In Russ.)

Cheremisina M. I., Kolosova T. A. *Ocherki po teorii slozhnogo predlozheniya* [An outline of the complex sentence theory]. Moscow, «URSS» Publ., 2010. (1st ed. Novosibirsk, «Nauka» Publ., 1987). 226 p. (In Russ.)