

Компьютерная лингвистика и интеллектуальные технологии:  
по материалам международной конференции «Диалог 2018»

Москва, 30 мая — 2 июня 2018 г.

## НАДКОРПУСНАЯ БАЗА ДАННЫХ КАК ИНСТРУМЕНТ ИЗУЧЕНИЯ ФОРМАЛЬНОЙ ВАРИАТИВНОСТИ КОННЕКТОРОВ<sup>1</sup>

**Инькова О. Ю.** (Olga.Inkova@unige.ch)

ИПИ ФИЦ ИУ РАН, Москва, Россия;  
Женевский университет, Женева, Швейцария

**Ключевые слова:** базы данных коннекторов, русский язык, корпусная лингвистика, структура коннекторов, формальная вариативность, статистические данные

## SUPRACORPORA DATABASE AS AN INSTRUMENT OF THE STUDY OF THE FORMAL VARIABILITY OF CONNECTIVES

**Inkova O. Yu.** (Olga.Inkova@unige.ch)

Institute of Informatics Problems, FRC CSC RAS, Moscow, Russia;  
University of Geneva, Geneva, Switzerland

The article intends to describe the formal variation of the connectors of the Russian language on the basis of a cognitive-semantic approach. Every *discourse variant DV* of a connector *K*, *i.e.* the specific form assumed by *K* in a discourse section, is singled out, and registered in the supracorpora database of connectors (*SCDB*), in which a system of intersecting clusters has been developed, allowing to assign in the course of the annotation the same *DV* to different structural clusters. In the next phase,

---

<sup>1</sup> Исследование выполнено в ИПИ ФИЦ ИУ РАН при поддержке РФФИ (грант № 16-06-00070).

on the base of further semantic analysis, the *DVs* with a common element are combined into a structural-semantic complex around a *basic form*: the minimal linguistic unit that enables the speaker to express a certain logical-semantic relation, and the listener to identify it. In conclusion, criteria for describing the formal variation of the connectors are proposed, as well as examples of the “profiles” of the basic forms. They reflect the potential of linguistic means that the speaker has at his disposal to express one or another logical-semantic relations or one of their combinations.

**Key words:** supracorpora database of connectives, Russian, corpus linguistics, formal variability of connectives, structure of connectives, statistical data

## 1. Введение

Хорошо известно, что состав связующих средств русского языка продолжает пополняться, как за счет новых образований из «блуждающих частиц» (Николаева 1985, 2008), так и за счет возникновения новых сочетаний из уже существующих языковых единиц, в частности союзов с частицами, модальными словами, наречиями и др. В связи с этим возникает вопрос: правомерно ли рассматривать каждое из них как отдельную языковую единицу? Представляется очевидным, что простой союз иначе соотносится с возникшими на его основе сочетаниями, чем с другим простым союзом. Иначе говоря, место в системе русского языка, с одной стороны, простых союзов *и — да — а — но — или*, а с другой, образований типа *да — да и — да еще и, да притом еще, да притом еще и, да еще притом, да вдобавок еще* и т.д., по-видимому, различно (см. постановку проблемы в [Прияткина 1977](#), [Черемисина, Колосова 2009: 123–139](#)).

Словари служебных слов и «сочетаний, эквивалентных слову» исходят из упрощенной трактовки проблемы, фиксируя в отдельной словарной статье — причем на довольно субъективной основе — некоторые из этих комбинаций и оставляя за бортом другие, имеющие, на наш взгляд, такой же статус. Так, из перечисленных выше сочетаний с *да* фиксируются, как правило (см. сводные таблицы в [Богданов, Рыжова 1997](#)), *да, да и, да еще* и *да еще и*. При этом [[Морковкин 1997](#)], [[Бурцева 2010](#)], [[Рогожникова 2003](#)] рассматривают *да еще и* как вариант *да еще*, а [[Ефремова 2004](#)] посвящает каждому из них словарную статью. [[Рогожникова 2003](#)] фиксирует также форму *да вдобавок еще*. Еще сложнее обстоит дело с двухместными коннекторами: например, если для *хотя* ни один из четырех словарей не фиксирует возможность образовывать двухместный коннектор, то для *едва* [[Ефремова 2004](#)] и [[Рогожникова 2003](#)] дают форму *едва... как*, а [[Морковкин 1997](#)] — *едва... а*.

Схожая позиция характеризует корпуса русского языка. Хельсинский аннотированный корпус русских текстов (ХАНКО) опирается в этом вопросе на список служебных многокомпонентных единиц, которые считаются «фразами» [[Мустайоки, Копотев 2004](#)]. Он создан разработчиками корпуса на основе словарей и используется для морфологической разметки. В списке союзов 323 единицы.

В НКРЯ создан «Корпусный словарь неоднословных лексических единиц», которые квалифицируются как «устойчивые лексические обороты» (<http://ruscorpora.ru/obgrams.html>). Необходимость создания такого словаря объясняется тем, что морфологическая разметка дается в НКРЯ для орфографического слова, т. е. отделяемого пробелом. В основу словаря положены данные частотных коллокаций НКРЯ с дополнениями из словаря [Рогожникова 2003] и МАС. Список оборотов в функции союза и союзного слова включает 159 единиц (т. е. меньше половины списка ХАНКО), для каждой из которой дано ориентировочное количество употреблений. В число «устойчивых» попадают обороты *добро бы еще* и *диви б* с единичными употреблениями; фиксируются как самостоятельные лексические единицы *добро бы* и *добро б, если бы* и *если б*, тогда как из перечисленных выше комбинаций с *да* фиксируются только *да и* и *да еще*, хотя поиск в НКРЯ по *для притом еще* выдает 57 вхождений, по *да вдобавок еще* — 50. Двухместных коннекторов в списке нет, хотя некоторые логико-семантические отношения (ЛСО) выражаются только ими (аналогия: *как... так*, сопоставительное *если... то* и др.). С другой стороны, в списке союзов и союзных слов оказываются такие единицы, как *что касается* и *что до*: их оценка как показателей связи требует теоретического обоснования.

Оба корпуса (НКРЯ и ХАНКО) считают многокомпонентные или неоднословные единицы устойчивыми сочетаниями, фразеологизмами, эквивалентами слова, иначе говоря, единицами, принадлежащими языку, отсюда акцент на их устойчивость и частотность. Такой подход приводит к определенной бессистемности и субъективности в описании состава служебных единиц русского языка, поскольку критерий частотности применяется непоследовательно, а критерий устойчивости не определен с достаточной четкостью<sup>2</sup>.

## 2. Когнитивно-семантический подход к проблеме формального варьирования коннекторов

Для описания структуры коннекторов, состоящих более чем из одного 'слова', мы предлагаем использовать когнитивно-семантический подход, позволяющий дать адекватное отражение способности носителя языка выражать или устанавливать при восприятии речи то или иное ЛСО между ситуациями. Это означает, прежде всего, разграничение понятийного и языкового уровня анализа (соответственно, тип ЛСО и языковые средства, которыми он может выражаться), поскольку выбор коннектора и его состава для выражения ЛСО зависит не только от знаний языка, но и от коммуникативного намерения говорящего, в передаче которого задействованы также общие принципы общения

<sup>2</sup> В связи с созданием корпусов, посвященных дискурсивным, или риторическим, в терминологии данного направления исследований, отношениям (ср., например, *The Penn Discourse Treebank*), возрос интерес и к их показателям, в частности, к тем случаям, когда в одном высказывании присутствует не один, а несколько маркеров риторических отношений (см., например, *Grote & al. 1995, Webber & Joshi 1998, Oates 2000, Fraser 2013, Webber 2016*). Однако внимание в этих работах уделяется преимущественно самим отношениям, и вопрос о формальной вариативности коннекторов не ставится.

и знания о мире. Чем типичней сценарий, лежащий в основе описываемых событий, тем меньше необходимость выражать между ними отношения соответствующим показателем (ср. *Наша Таня громко плачет: уронила в речку мячик*); с другой стороны, чем важнее с коммуникативной точки зрения данное ЛСО для говорящего, тем тяжеловесней будет его показатель: ср. ниже три примера с *как только*. Если говорящий решает эксплицировать ЛСО, то дальше в силу вступают семантические законы, определяющие возможности сочетаний языковых единиц и состав показателя ЛСО.

При таком подходе 'неоднословные' коннекторы, обладающие свойством вариативности формы, рассматриваются не как целостные грамматические единицы, входящие в знание языка и воспроизводимые в готовой форме, а как свободные сочетания, принадлежащие сфере речи. Они, в частности, обладают важнейшим для автономности составляющих их словоформ свойством отделмости, допуская вставку между ними других слов [Инькова 2016, Кобозева 2016]. Ср.: *Нет я усядусь на скамеечке. Да и что я стану там делать и Звонить зря было нечего, да ему и не к фигуре.*

Основной единицей анализа при когнитивно-семантическом подходе становится *речевая реализация* (РР), т. е. та форма, в которой коннектор встречается в данном конкретном высказывании. Именно она регистрируется в надкорпусной базе данных коннекторов (НБД). Ср. три РР для *как только*, выражающие ЛСО контактного предшествования: *Как только* дело дошло до «честного слова», я махаю руками и сажусь за стол (Чехов); *Как только* кончается отношение служебное, *так* кончается всякое другое (Л. Толстой); *Он, как только* проснулся, *тотчас же* вознамерился встать, умыться (Гончаров).

### 3. НБД как инструмент анализа структуры коннекторов

Применяемая в НБД система аннотирования (подробнее см. Inkova, Rorkova 2017) позволяет решать теоретически сложный вопрос о том, являются ли приведенные выше три РР с общим элементом *как только* вариантами одного коннектора (и какого именно) или разными коннекторами не на этапе аннотирования, а после семантического анализа соответствующей группы РР. На первом этапе аннотирования каждая регистрируемая РР приписывается к двум типам кластеров. В четырех кластерах первого типа учитывается количественный состав РР, а также наличие показателя ЛСО в одном или в каждом из соединяемых фрагментов текста:

- одноэлементные: состоящие из одного элемента, т. е. РР, неразложимые на более мелкие языковые единицы, например, *и, или, но, а, хотя, тогда, иначе, словом*;
- многоэлементные: РР, состоящие из нескольких элементов, например, *и вообще, к тому же, но при всем при том, еще и|вдобавок*<sup>3</sup>;

<sup>3</sup> Одиночная вертикальная черта в названии РР означает, что элементы РР разделены другими словами.

- двухкомпонентные: РР, компоненты которых вводят два текстовых фрагмента, например, *даже если бы||то*<sup>4</sup>, *как только||тут же, хотя|и||но| между тем*;
- многокомпонентные: РР, компоненты которых вводят более двух текстовых фрагментов (например, *не только||но даже и||и даже, хотя||хотя|| однако все же*).

Заметим, что из 926 РР, зарегистрированных в НБД на 01.02.2018, на долю одноэлементных приходится чуть более 5% (47), многоэлементных — почти 50% (457), двухкомпонентных — около 40% (357) и многокомпонентных — чуть более 5% (47).

Перечисленные кластеры первого типа, отражая важный классификационный признак коннекторов (их состав), не учитывают тот факт, что одна и та же языковая единица может входить в различные сочетания, см. выше примеры для *как только*. Поэтому описание состава многоэлементных, двух- и многокомпонентных РР и сочетаемости составляющих их компонентов или элементов решается в НБД через систему «перекрестных», или «пересекающихся», кластеров второго типа, которые позволяют разметчикам одновременно помещать аннотируемую РР в разные кластеры и которые пополняются в процессе аннотирования. См. [Таблицы 1 и 2](#).

**Таблица 1.** Примеры РР с указанием кластеров первого и второго типов, в которые они входят (состояние на 1 февраля 2018 г.)

Речевые реализации (РР)				
		<i>а впрочем</i>	<i>да еще вдобавок</i>	<i>не только    а даже и</i>
Кластеры РР	«Многоэлементные»	«Многоэлементные»	«Двухкомпонентные»	
	«а»	«вдобавок»	«даже»	
	«впрочем»	«да»	«а»	
		«еще»	«и»	
			«не»	
			«не только»	
			«только»	

<sup>4</sup> Двойная вертикальная черта в названии двух- и многокомпонентных РР коннекторов означает, что компоненты не только разделены другими словами, но и вводят разные текстовые фрагменты.

**Таблица 2.** Примеры кластеров второго типа и входящих в них РР (состояние на 1 февраля 2018 г.)

	Кластеры РР		
	«вдобавок»	«значит»	«притом»
РР, входящие в указанные кластеры	вдобавок	а значит	да и притом
	да вдобавок	если  значит	да притом
	да еще вдобавок	если только  значит	да притом еще
	еще вдобавок	значит	и притом
	еще и вдобавок	коли  то значит	и притом
	и вдобавок		

На основе последующего семантического анализа зарегистрированные в НБД РР объединяются в структурно-семантический комплекс вокруг базовой формы (БФ), т. е. той минимальной РР, которая позволяет говорящему выразить некоторое ЛСО, а слушающему — его идентифицировать. Например, для выражения отношения контактного предшествования говорящий может использовать РР *лишь*, а также целый ряд РР, включающих этот элемент: *лишь||как*; *едва лишь*; *едва лишь||как*; *только лишь*; *только лишь||как*; *лишь только*; *лишь только||как*; *чуть лишь*; *чуть лишь||как*; *едва лишь только*; *едва лишь только||как* и ряд других [Кобозева 2017]. Форма *лишь* будет считаться БФ, поскольку она может выражать это ЛСО сама по себе. Аналогичным образом, выделяются кластеры, объединенные вокруг БФ *едва*, *только* и *чуть*, пересекающиеся друг с другом благодаря наличию РР, включающих две БФ.

БФ удобно использовать и для номинации кластера РР, который можно рассматривать как абстрактную единицу системы коннекторов данного языка. Поэтому в НБД наряду, например, с кластером *только* создан кластер *не только*: последний является также именем БФ, выражающей отношение неединственности [Инькова 2016].

Статистический анализ употреблений РР позволил заметить, что БФ обычно далеко не самая частотная РР в своем кластере. Наиболее частотную по корпусным данным РР для выражения того или иного ЛСО предлагается называть *основной формой коннектора*. Так, в кластере *лишь* основной формой будет РР *лишь только*. В кластере *не только* — РР *не только||но* и [Амеличева 2017].

Описание формальной вариативности неоднословных коннекторов предлагается делать исходя из описанных выше теоретических принципов и с учетом приводимых ниже параметров, характеризующих БФ коннектора.

1. Возможность для БФ образовывать многоэлементные РР, выражающие то же ЛСО.

*Доп. критерий:*

- контактное vs. дистантное расположение элементов.

2. Возможность для БФ образовывать двухкомпонентные РР.

*Доп. критерии:*

- симметричность vs. асимметричность компонентов,

- вариативность компонентов (возможность образовывать многоэлементные компоненты),
- контактное vs. дистантное расположение элементов.

### 3. Возможность для БФ образовывать многокомпонентные РР.

*Доп. критерии:*

- симметричность vs. асимметричность компонентов,
- вариативность компонентов (возможность образовывать многоэлементные компоненты).

### 4. Возможность для БФ сочетаться с показателями других ЛСО.

*Доп. критерии:*

- ЛСО той же семантической группы,
- ЛСО другой семантической группы.

Классы РР, полученные на основе этих параметров благодаря информации из НБД, будут «перекрестными», но они позволяют очертить семантический и формальный профиль БФ показателя того или иного ЛСО. Несколько таких профилей приведены ниже в Таблице 3. Важно подчеркнуть, что речь идет не о лексических единицах, существующих в языке а priori, а именно о потенциале языковых средств, которым располагает говорящий для выражения ЛСО<sup>5</sup>, что отражено в использовании слова *возможность* при формулировании параметров.

---

<sup>5</sup> Для названия ЛСО в данной работе намеренно используется нейтральная терминология русской грамматической традиции, дополняемая при отсутствии необходимого термина, рубриками из классификации, разрабатываемой автором статьи. Этот терминологический компромисс обусловлен тем, что существующие на сегодняшний день классификации, используемые для разметки корпусов и в основе которых лежит, как правило, Теория риторической структуры (RST), не могут быть использованы для анализа внутренней формы коннекторов и их сочетаемости. С одной стороны, в них не хватает рубрик, чтобы различить, например, ЛСО, выражаемые *а, зато и наоборот*; ср. предложение «мирового стандарта» отношений в (Prasad & Bunt 2016), включающий 20 отношений: *Cause, Condition, Negative Condition, Purpose, Manner, Concession, Contrast, Exception, Similarity, Substitution, Conjunction, Disjunction, Exemplification, Elaboration, Restatement, Synchrony, Asynchrony, Expansion, Functional dependence, Feedback dependence*, или, наоборот, чрезвычайно детальную классификацию, включающую 62 риторических отношения, но тем не менее слабо разработанную в зоне ЛСО, в [Кибрик, Подлесская 2009]. С другой стороны, все они содержат «риторические отношения», многие из которых не являются логико-семантическими; ср. в «мировом стандарте», например, последние два или в классификации А. Кибрика и В. Подлесской *Begin, Emotional reaction, Headline, Problem* и некоторые другие. Разрабатываемая автором классификация предназначена именно для анализа ЛСО, выражаемых коннекторами. Она основана на семантических операциях, лежащих в основе отношения, и последовательно различает уровни, на которых ЛСО может быть установлено (пропозиция, высказывание, метатекст); см. обзор существующих классификаций и основные теоретические положения, лежащие в основе предлагаемой классификации, в [Inkova 2017].

Таблица 3. Примеры профилей БФ коннекторов

Основной параметр	Дополнительный параметр	например	также	зато	если... то (сопоставительный)	не только	или
1. Возможность образовывать однокомпонентные многоэлементные РР, выражающие то же ЛСО	контактное расположение	-	+	-	-	-	+
	дистантное расположение	-	+	-	-	-	-
2. Возможность образовывать двухкомпонентные РР, выражающие то же ЛСО	симметричность vs. асимметричность компонентов	-	-	+ симметричность	+ асимметричность	+ асимметричность	+ симметричность + асимметричность
	вариативность компонентов	-	-	-	-	-	+ второй компонент
3. Возможность образовывать многокомпонентные РР, выражающие то же ЛСО	контактное vs. дистантное расположение элементов	-	-	-	-	-	+ второй компонент
	симметричность vs. асимметричность компонентов	-	-	+ симметричность	+ асимметричность	+ асимметричность	+ симметричность + асимметричность
4. Возможность сочетаться с показателями других ЛСО	вариативность компонентов	-	-	-	-	-	+ второй компонент
	ЛСО той же семантической группы	+	+	+	-	-	+
	ЛСО другой семантической группы	+	+	+	+	+	+



### *Например*

1. Данная БФ для выражения ЛСО спецификации не может изменять свой состав для выражения того же отношения. В этом она схожа с БФ *зато* и отличается от БФ *также*, характеризующейся формальной вариативностью.
- 2–3. *Например* не может образовывать ни двух-, ни многокомпонентные РР для выражения того же ЛСО.
4. Данная БФ сочетается с показателями других ЛСО, относящихся как к той же семантической группе, так и к другим. Ср. (1) в сочетании с другим показателем спецификации *в частности*, и (14) ниже с показателем ЛСО сопоставления:
  - (1) Под флагами повсеместно висели листовки с «обещаниями» фюрера своим рабочим: два десятка пунктов, *в частности, например*, обещание сделать 1 мая Общенациональным праздником немецких трудящихся и оплаченным выходным днем. [Елена Съянова. Лей — обольститель немецкого рабочего класса // «Знание — сила», 2005]

### *Также*

1. Данная БФ может образовывать многоэлементные РР для выражения того же ЛСО (аддитивность), как с контактным (3), так и с дистантным (4) расположением элементов.
  - (2) (...) около него красовались лубочные картинки, представляющие взятие Кистрина и Очакова, *также* выбор невесты и погребение кота [А. С. Пушкин. Капитанская дочка (1836)]
  - (3) Выяснилось, что в бюро иностранцев ни о каком Воланде, *а равно также* и Фаланде, маге, ровно ничего не слыхали. [М. А. Булгаков. Мастер и Маргарита, (1929–1940)]
  - (4) Мы не будем здесь останавливаться на деталях обряда увенчания — развенчания (хотя они очень интересны) и на различных вариациях его по эпохам и разным празднествам карнавального типа. Не будем *также* анализировать *и* различные побочные обряды карнавала. [М. М. Бахтин. Проблемы поэтики Достоевского (1963)]
- 2–3. Данная БФ не может образовывать двух- и многокомпонентные РР.
4. *Также* может сочетаться с показателями того же ЛСО, например, с градационными показателями аддитивных пропозициональных ЛСО, как *сверх того* в (5), и с показателями ЛСО других семантических групп; ср. в (6) с показателем отношения неединственности *не только*||*но*:
  - (5) Тургенев мне все более и более нравится: у него ум — необыкновенный ум, *сверх того также* чутье, способность к анализу, умение хорошо подступиться ко всякой вещи. [А. В. Дружинин. Дневник (1845)]
  - (6) Самоубийство есть *не только* ложное и греховное отношение к жизни, *но также* ложное и греховное отношение к смерти. [Н. А. Бердяев. О самоубийстве (1931)]

### **Зато**

1. Как и БФ *например*, БФ *зато* не может изменять свой состав для выражения того же ЛСО (возместительное противопоставление).
- 2–3. В отличие от *например* и *также*, БФ *зато* для выражения того же ЛСО может образовывать двух- (7) и многокомпонентные (8) РР, но только с симметричными компонентами (повторяющееся два раза или более *зато*):
  - (7) У него не было идолов, *зато* он сохранил силу души, крепость тела, *зато* он был целомудренно-горд. [И. А. Гончаров. Обломов (1859)]
  - (8) Ни жеманства, ни кокетства, никакой лжи, никакой мишуры, ни умысла! *Зато* ее и ценил почти один Штольц, *зато* не одну мазурку просидела она одна, не скрывая скуки; *зато*, глядя на нее, самые любезные из молодых людей были неразговорчивы, не зная, что и как сказать ей... [И. А. Гончаров. Обломов (1859)]
4. БФ *зато* может сочетаться с показателями других ЛСО, той же семантической группы и других групп. В (9) *зато* образует многоэлементную РР с дистантным расположением элементов с показателем ЛСО той же семантической группы *наоборот* (контрастное противопоставление):
  - (9) В последние два года средства от дорожных штрафов несколько сократились, *зато* аппетиты чиновников, *наоборот*, возросли. [Сергей Ждакаев. Ильич для камуфляжа. Кто кормился из внебюджетного фонда (2001) // «Известия», 2001.07.18]

В (10) *зато* сочетается с показателем уступительного ЛСО *хоть и* в составе двухкомпонентной РР:

- (10) Дай же, я думаю, *хоть и* упущу на время одно, *зато* другое схвачу за хвост, — своего-то, своего-то, по крайности, не упущу. [Ф. М. Достоевский. Преступление и наказание (1866)]

### **Если... то (сопоставительный)**

- 1–2. Одиночный союз *если* не может выражать ЛСО сопоставления, поэтому БФ для этого ЛСО будет считаться двухкомпонентная *если... то*. Ср. (11) и (12)
  - (11) *Если* в Александровском округе климат морской, *то* в Тымовском он континентальный [А. П. Чехов. Остров Сахалин. Пример (БАС)]
  - (12) *Если* в Александровском округе климат морской, *то* в Тымовском он континентальный

*Если... то* сопоставления не может менять свой состав для выражения того же ЛСО.

3. Можно предположить, хотя такие примеры пока не зафиксированы в НБД, что данная БФ для выражения того же ЛСО может входить в состав многокомпонентных РР с повторяющимся первым компонентом *если* и вторым компонентом *то*:

- (13) *Если* черно-белый котенок всего боялся и жалобно мяукал, *если* рыжий все время от нас прятался, *то* полосатый сразу почувствовал себя как дома.
4. БФ *если...* *то* не может сочетаться с другими показателями сопоставительных отношений, но может сочетаться с показателями других ЛСО, которые будут входить, как правило, в состав второго компонента: ср. (14) с показателем ЛСО спецификации:
- (14) *Если* в Москве Лужков решительно отказал в этом МГТС, *то, например*, в Нижневартовске «повременка» будет внедрена до конца 2003 г. [Михаил Классон. Тарифные аппетиты (2003) // «Время МН», 2003.05.26]

### ***Не только***

1. БФ *не только* выражает ЛСО неединственности:
- (15) Он до того углубился в себя и уединился от всех, что боялся даже всякой встречи, *не только* встречи с хозяйкой. [Ф. М. Достоевский. Преступление и наказание (1866)]
- Она может расширять свой минимальный состав, образуя многоэлементные РР:
- (16) Важно раскрыть функцию идей в полифоническом мире Достоевского, *а не только* их монологическую субстанцию. [М. М. Бахтин. Проблемы поэтики Достоевского (1963)]
2. БФ *не только* может выражать то же ЛСО в составе двухкомпонентных РР, но, в отличие от *если... то*, характеризуется формальным варьированием во втором компоненте. Ср. (17), и (18):
- (17) В деревне, без службы Иван Ильич в первый раз почувствовал *не только* скуку, *но* тоску невыносимую. [Л. Н. Толстой. Смерть Ивана Ильича (1886)]
- (18) *Не только* драньем вихров, *но даже и* помелом было бы полезно обойтись с иными дураками. [Ф. М. Достоевский. Преступление и наказание (1866)]
3. БФ *не только* может входить в состав многокомпонентных асимметричных РР с формальным варьированием в компонентах, не включающих БФ.
- (19) *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 (1929–1940)]

4. БФ *не только* может сочетаться с показателями ЛСО, относящихся к другим семантическим группам: в (20) — с показателем ЛСО контрастного противопоставления *напротив*:

(20) <...> и это объясняется *не только* его положением журналиста, требующим трактовки всего в разрезе современности; *напротив*, мы думаем, что пристрастие Достоевского к журналистике и его любовь к газете <...> объясняются именно основной особенностью его художественного видения. [М. М. Бахтин. Проблемы поэтики Достоевского (1963)]

*Или*

1. БФ *или* может выражать ЛСО альтернативы в своем минимальном составе, допуская незначительное формальное варьирование в сочетании с *же* (21):

(21) Как будто дело происходит в гастрономе *или же* на рынке.  
[Сергей Довлатов. Иностранка (1985)]

- 2–3. Она может выражать то же ЛСО в составе двух- и многокомпонентных РР, как симметричных (повторяющееся *или*), так и асимметричных (22). В составе таких РР БФ *или* может варьироваться (как правило, заключительный компонент ряда).

(22) Требовалось выяснить, *были ли* похищены эти женщины шайкой убийц и поджигателей *или же* бежали вместе с преступной компанией добровольно? [М. А. Булгаков. Мастер и Маргарита (1929–1940)]

4. БФ *или* может сочетаться с показателями других ЛСО, относящихся как к той же семантической группе, например, разделительное *а то* в (23), так и к другим семантическим группам; ср. (24) с показателем ЛСО коррекции *вернее*:

(23) Возьмите две капельки амбре, одну капельку вервены и получите дух настоящий... настоящий, — она пожевала губами, ища слова, — земной и небесный. *А то* возьмите основной дух Трефль инкарнат, пряный, точно с корицей, да в него на три капли одну белого ириса... <...> *Или* возьмите нежную Икзору, — не слушая, продолжает фантазировать мадам Лазенская, — а к ней подлейте одну каплю тяжелого Фужеру... [Н. А. Тэффи. За стеной (1910)]

(24) Основное, что определяло его лицо, это было, пожалуй, выражение добродушия, которое нарушали, впрочем, глаза, *или, вернее*, не глаза, а манера пришедшего глядеть на собеседника. [М. А. Булгаков. Мастер и Маргарита (1929–1940)]

## 4. Выводы

Предлагаемый когнитивно-семантический подход к проблеме формального варьирования коннекторов дает, на наш взгляд, более адекватное представление о механизмах выражения того или иного ЛСО и о том потенциале языковых средств, которым располагает для этого говорящий. Регистрируемые в НБД РР являются при таком подходе единицами речи, каждый раз *ex novo* создаваемыми говорящим с учетом его коммуникативного намерения и на основе семантических законов, определяющих возможности сочетаемости входящих в их состав элементов. Разработанная в НБД система аннотирования с использованием перекрестных кластеров позволяет решать теоретически сложный вопрос о статусе регистрируемой РР не в момент ее аннотирования, а на основе ее дальнейшего семантического анализа. Особый интерес представляет соотношение состава РР русского языка и частотности их употребления. Если на долю одноэлементных РР приходится, как мы видели всего 5% (47) от общего числа зарегистрированных РР, то доля их употребления в текстах составляет 38% (они были зарегистрированы в 2684 аннотациях из 7016). Наиболее частотными по данным НБД являются многоэлементные РР (42.5% или 2983 аннотации), наименее частотными — многокомпонентные (0.8%). Употребление двухкомпонентных РР составляет 16.9%. НБД является, таким образом, эффективным инструментом, позволяющим судить как о формальном составе связующих средств русского языка, так и о частоте их употребления в тексте, и исчислить их возможные варианты.

## Литература

1. *Amelicheva V.* (2017). Formal'noe i semanticheskoe var'irovanie russkogo konektora *ne tol'ko... no i* i ego frantsuzskie ekvivalenty [Formal and semantic variations of the Russian connector *ne tol'ko... no i* and its French equivalents]. *Contrastive linguistics*. 2017. Vol. XLII, No. 4. Pp. 9–20.
2. *Bogdanov S. I., Ryzhova Yu. V.* (1997). *Russkaya sluzhebnyaya leksika* [Russian structural words], St. Petersburg, St. Petersburg State Univ.
3. *Burtseva V. V. ed.* (2010). *Slovar' narechii i sluzhebnykh slov russkogo yazyka* [Dictionary of the adverbs and structural words of Russian language], Moskva, Drofa.
4. *Cheremisina M. I., Kolosova T. A.* (2010). *Oчерки po teorii slozhnogo predlozheniya* [An outline of the complex sentence theory], Moscow, URSS. (1st ed. Novosibirsk, Nauka, 1987).
5. *Efremova T. F.* (2004). *Tolkovyi slovar' sluzhebnykh chastei rechi russkogo yazyka* [Explanatory dictionary of the Russian auxiliary parts of speech], Moscow, Astrel'-Ast.
6. *Fraser B.* (2013). Combinations of Contrastive Discourse Markers in English. *International Review of Pragmatics* 5. 2013. Pp. 318–340.

7. Grote B., Lenke N., Stede M. (1995). Mar(k)ing Concessions in English and German. Proceedings of the 5th European Workshop on Natural Language Generation, Leiden, May, Leiden University. Pp. 11–32.
8. HANCO — The Helsinki Annotated Corpus, available at: [http://www.ling.helsinki.fi/projects/hanco/index\\_e.html](http://www.ling.helsinki.fi/projects/hanco/index_e.html).
9. Inkova O. (2016). K probleme opisaniya mnogokomponentnykh konnektorov russkogo yazyka [On the problem of the description of multiword connectives of Russian language: *ne tol'ko... no i* (no only... but also)]. *Voprosy jazykoznanija*. 2016. No. 2. Pp. 37–60.
10. Inkova O. (2017). Le relazioni logico-semantiche tra gli enunciati: una proposta di classificazione [The logical-semantic relations: a classification proposal]. M. di Filippo, F. Esvan (eds). *Studi di linguistica slava*. Napoli, Il Torcoliere. Pp. 105–124.
11. Inkova O., Popkova N. (2017). Statistical data as information source for linguistic analysis of Russian connectors. *Informatics and applications*. 2017. Vol. 11, No. 3. Pp. 123–131.
12. Kibrik A. A., Podlesskaya V. I. (2009). Rasskazy o snovideniyakh. Korpusnoe issledovanie ustnogo russkogo diskursa [Night dream stories: A corpus study of Russian spoken discourse]. Moscow, Yazyki Slavyanskikh Kul'tur.
13. Kobozeva I. M. (2016). Kognitivno-semanticeskii podkhod k opisaniyu sredstv svyazi predlozhenij (na primere konnektorov so znacheniem neposredstvennogo sledovaniya) [Cognitive-semantic approach to the description of means of connection of sentences (by the example of connectors with the meaning of direct following)]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova*. 2016. No. 10. Pp. 120–133.
14. Kobozeva I. (2017). Konnektory kontaktного predshestvovaniya v russkom i frantsuzskom jazykakh v zerkale nadkorpusnoj bazy dannykh [The connectors expressing contact precedence in Russian and French as reflected in the supracorpora database]. *Contrastive linguistics*. 2017. Vol. XLII, No. 4. Pp. 48–62.
15. MAS — Slovar' russkogo yazyka [Dictionary of the Russian language]: In 4 vol. A. P. Evgen'eva (ed.), Moscow, Russkii Yazyk, 1981.
16. Morkovkin V. V. ed. (1997). Slovar' strukturnykh slov russkogo yazyka [Dictionary of structural words of the Russian language], Moscow, Lazur'.
17. Mustaioki A., Kopotev M. (2004). K voprosu o statuse ekvivalentov slova tipa *potomu chto, v zavisimosti ot, k sozhaleniyu* [On the status of word-equivalents of the type *potomu chto, v zavisimosti ot, k sozhaleniyu*]. *Voprosy jazykoznanija*. 2004. No. 3. Pp. 88–107.
18. Nikolaeva T. M. (1985). Funktsii chastits v vyskazyvanii [Functions of particles in the utterance]. Moskva, Nauka.
19. Nikolaeva T. M. (2008). Neparadigmaticheskaya lingvistika (istoriya «bluzhdayushchikh chastits») [Non-paradigmatic linguistics (history of the “wandering particles”)]. Moscow, Yazyki Slavyanskikh Kul'tur.
20. NKRYa — Natsional'nyi korpus russkogo yazyka [Russian National Corpus], available at: [www.ruscorpora.ru](http://www.ruscorpora.ru).

21. Oates S. (2000). Multiple Discourse Marker Occurrence: Creating Hierarchies for Natural Language Generation, Proceedings of ANLP-NAACL 2000, April 29—May 4, 2000. Seattle, Washington, USA. Pp. 41–45.
22. Prasad R., Bunt H. (2016). Semantic relations in discourse: the current state of ISO 24617-8. H. Bunt (ed.), Proceedings of 12th Joint ACL-ISO Workshop on Interoperable semantic annotation, 28 May 2016, Slovenia, Portorož. Pp. 80–92. Available at: <http://www.lrec-conf.org/proceedings/lrec2016/workshops/LREC2016Workshop-ISA12proceedings.pdf>.
23. Priyatkina A. F. (1977). Ob otlichii soyuza ot drugikh svyazuyushchikh sredstv [On the difference between the conjunction and other connecting words]. Russkii yazyk v shkole. 1977. No. 4. Pp. 102–106.
24. Rogozhnikova R. P. (2003). Tolkovyi slovar' sochetanii, ekvivalentnykh slovu [Explanatory dictionary of constructions equivalent to the word], Moscow, Astrel'-AST.
25. The Penn Discourse Treebank, available at: <https://www.seas.upenn.edu/~pdtb/>.
26. Webber B. L., Joshi A. K. (1998). Anchoring a Lexicalised Tree-Adjoining Grammar for Discourse. M. Stede, L. Warner, E. Hovy (eds), Discourse Relations and Discourse Markers. Proceedings from the Workshop. COLING-ALC'98, Montreal. Pp. 86–92.
27. Webber B. L. (2016). Concurrent Discourse Relations, Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference “Dialogue 2016”, Moscow, June 1–4, 2016. available at: <http://www.dialog-21.ru/media/3488/webber.pdf>.