

*С. П. Тимошенко, Л. Л. Иомдин,
С. А. Гладилин, Е. С. Иншакова
S. P. Timoshenko, L. L. Iomdin,
S. A. Gladilin, E. S. Inshakova*

СИНТАГРУС В СОСТАВЕ НКРЯ: НОВЫЕ ВОЗМОЖНОСТИ¹

SYNTAGRUS WITHIN THE RUSSIAN NATIONAL CORPUS: NEW POSSIBILITIES

Аннотация. Данная статья знакомит читателей с актуальным состоянием корпуса СинТагРус, уделяя особое внимание структурной организации разметок разных типов. На основании этого материала формулируются проблемы, без решения которых невозможно осуществление полноценного поиска по корпусу, и предлагаются некоторые решения.

Ключевые слова. Синтаксис зависимостей, анафорическая разметка, лексико-функциональная разметка, микросинтаксическая разметка, темпоральная разметка, эллипсис, поиск.

Abstract. The paper describes the current state of the SynTagRus corpus, with a particular focus on the structure of its annotation. There are 8 different types of annotation interplaying in the corpus: morphological, syntactic, lexical-functional, anaphoric, microsyntactic, temporal, the annotation of elliptical sentences and the annotation of word meanings. Based on their description, we address the challenges with which the search engine developers are faced. In particular, the search interface should allow multiword queries, where a word can be linked to any other in as many ways as there are annotation types in corpus. Unlike morphological characteristics and other word-based annotations, which are unique, links between words satisfying the query conditions may not be unique, so, the search algorithm must be able to iteratively look for any of them. So, when formulating a query, the user should indicate how the search mechanism should treat the link specified: stop after the first occurrence is found or continue to find all of them.

Keywords. Dependency syntax, anaphora marking, lexical functions, microsyntax markup, temporal markup, ellipsis, search.

1. Вводные замечания

Входящий в состав Национального корпуса русского языка «Синтаксически размеченный корпус русского языка» (или СинТагРус, от **Syntactically Tagged Russian Corpus**), разрабатываемый в лаборатории компьютерной лингвистики ИППИ РАН, пополняется новыми типами разметки — см., в частности, описание в [Иншакова и др. 2019]. На данный момент в корпусе имеется 8 видов разметки: 1) морфологическая; 2) синтаксическая; 3) разметка эллипсиса; 4) лексико-функциональная; 5) микросинтаксическая; 6) анафорическая; 7) темпоральная; 8) разметка значений многозначных слов.

¹ Данная работа выполнена при поддержке гранта 19-07-00842 РФФИ.

Авторы корпуса хотят сделать общим достоянием все результаты своей работы. Для этого требуется выполнить два условия: во-первых, обеспечить корпус достаточным количеством справочных материалов, чтобы разобраться в его содержимом мог любой желающий; во-вторых, обеспечить физическую доступность данных. Что касается физической доступности, то она пока осуществляется Национальным корпусом русского языка (НКРЯ) не в полном объеме: далеко не все, чем корпус в принципе располагает, можно найти с помощью действующего поискового интерфейса. Из перечисленных разметок на сайте НКРЯ доступны для ознакомления только морфологические, синтаксические и лексико-функциональные данные.

Разнообразие разметки порождает ряд специфических проблем поиска: как сделать так, чтобы пользователь мог переходить от одного типа разметки к другому и сочетать их между собой? На первый взгляд, задача кажется чисто технической, однако ее решение во многом зависит от внутренних лингвистических установок того, кто осуществляет поиск.

Данная статья призвана, во-первых, описать актуальное состояние корпуса, уделяя особое внимание формату разметки. Во-вторых, сформулировать требования к организации поиска, которые бы учитывали, с одной стороны, структурные особенности разметки, а с другой — потребности и внутренние установки пользователей. Посвященная поиску часть статьи требует особого внимания корпусного лингвистического сообщества: широкое обсуждение этих вопросов может определяющим образом повлиять на эффективность проведения новых корпусных исследований.

2. Структура разметки СинТагРуса

В самом абстрактном виде любая разметка текста — это набор помет при элементах этого текста. Лингвист базовым элементом текста считает слово, так что первая стадия создания любой разметки — это разбиение на слова, точнее, на словоформы, которые считаются конкретными реализациями слов в тексте. Как правило, словоформа определяется формально как последовательность букв (шире — знаков), расположенная между пробелами. На месте пробела также может выступать пунктуационный знак. Подход, используемый для нахождения границ в СинТагРусе, составляет одну из его особенностей. Границы словоформ не во всех случаях совпадают с традиционным делением.

Своим существованием СинТагРус обязан лингвистическим процессорам серии ЭТАП, которые задумывались прежде всего как средство межязыкового перевода. Для тех случаев, когда выражение из нескольких слов функционирует как цельная единица и не путается (во всяком случае, систематически) со словосочетаниями, в морфологическом и комбинаторном словарях системы ЭТАП имеется соответствующая статья. Например, существует отдельная статья для выражения *что бы то ни было*. Это означает, что данное выражение — с точки зрения словарей — приравнено к слову. Во время работы процессор разбивает текст на словоформы в соответствии со статьями в своих словарях. Соответственно, вхождение в текст выражения *что бы то ни было* будет представлять одну словоформу, а не пять. С одной стороны, деление текста на словоформы не по пробелам, а «по словарю» является более содержательным — грамматически, синтаксически и семантически мотивированным. Можно сказать, что текст действительно поделен на слова — единицы, заданные сочетанием грамматической формы и синтаксической функции. Используя в дальнейшем термин «слово» применительно к СинТагРус'у, мы будем понимать под ним именно такое сочетание: отрезок текста (возможно, с пробелом или некоторыми другими неалфавитными символами), соответствующая ему начальная форма (лемма) и набор грамматических характеристик.

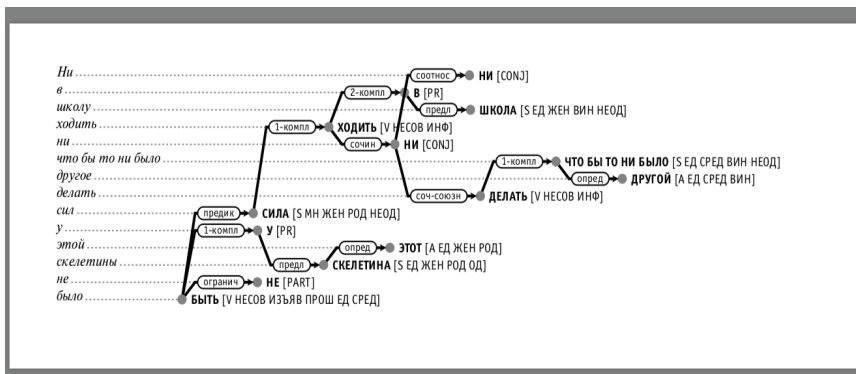


Рис. 1. Синтаксическая структура фразы, содержащей выражение *что бы то ни было*

Опора на словари имеет и другое преимущество. Если рассматривать корпус как базу данных, то наличие словарей делает ее структуру

нормализованной: словари работают как справочная таблица, которая используется для контроля корректности корпусных данных.

С другой стороны, нерегулярность, выражающаяся в том, что в большинстве случаев пробелы отделяют одну словоформу от другой, но иногда оказываются «внутри» словоформы, усложняет работу с корпусом как для пользователей, так и для разработчиков. Поэтому поисковый интерфейс НКРЯ в текущем варианте не позволяет искать единицы типа *что бы то ни было*.

На момент написания этой статьи в состав СинТагРус'а входит 731 текст. Технически это 731 файл в формате, представляющем собой надстройку над xml. Тексты в файлах разбиты на 87 676 фрагментов, каждый из которых заключен в тег <S> — это предложения. Содержимое тега S, в свою очередь, разбито на фрагменты, заключенные в тег W, — это слова. Всего в СинТагРусе 1 239 942 таких слова.

3. Базовая разметка слова

Мы проиллюстрируем базовый формат разметки слова на примере вхождения *что бы то ни было* в предложение, приведенное на рис. 1. В исходном файле оно выглядит так:

```
<W DOM=>8> FEAT=>S ЕД СРЕД ВИН НЕОД> НУПОТ=>1-компл.11> ID=>6>  
KSNAME=>ЧТО$БЫ$ТО$НИ$БЫЛО> LEMMA=>ЧТО БЫ ТО НИ БЫЛО> LINK=>1-  
компл>>что бы то ни было</W>
```

Текстовый фрагмент никак не выделен, синим цветом выделен собственно тег, красным — его атрибуты, а фиолетовым — их значения. Атрибут DOM содержит информацию о синтаксическом хозяине слова. В качестве значения используется порядковый номер слова в предложении. Атрибут FEAT содержит набор грамматических характеристик. Атрибут НУПОТ может показаться дублером атрибута LINK, хранящего имя синтаксического отношения между словом и его хозяином. На самом деле НУПОТ предназначен для хранения имени правила ЭТАПа, обеспечившего построение этого синтаксического отношения между словом и его хозяином. ID — идентификатор, порядковый номер слова в предложении. KSNAME — это ссылка на статью комбинаторного словаря (далее — КС) системы ЭТАП, а LEMMA — начальная форма слова.

Разметка эллипсиса и разметка значений многозначных слов хранятся в виде атрибутов слова. Анафорическая и микросинтаксическая разметки хранятся в виде атрибутов предложения. Лексико-функцио-

нальная и темпоральная разметки хранятся в виде специальных тегов, которые так же, как и тег слов W, «вложены» в тег предложения S.

4. Разметка эллипсиса

В СинТагРус'е принят следующий подход: эллипсис понимается как пропуск на поверхностном уровне одного или нескольких синтаксических элементов. Соответственно, пропущенные слова восстанавливаются в структуре и имеют все полагающиеся им атрибуты, а также помету FANTOM (значение дополнительного атрибута при теге W), позволяющую понять, что данному слову соответствует фрагмент текста нулевой длины (однако у него есть линейная позиция).

На рис. 2 представлена структура предложения *К людям врожденную любовь слон испытывает, особенно к красивым женщинам, но еще большую — к маленьким детям*. В нем восстановлено два пропуска, каждый из которых включает два слова — *испытывать* и *любовь*. В левом столбце видны пустые строки. Таково наглядное представление эллипсиса в действующей версии СинТагРус'а на сайте НКРЯ, однако использовать соответствующую помету для поиска там нельзя.

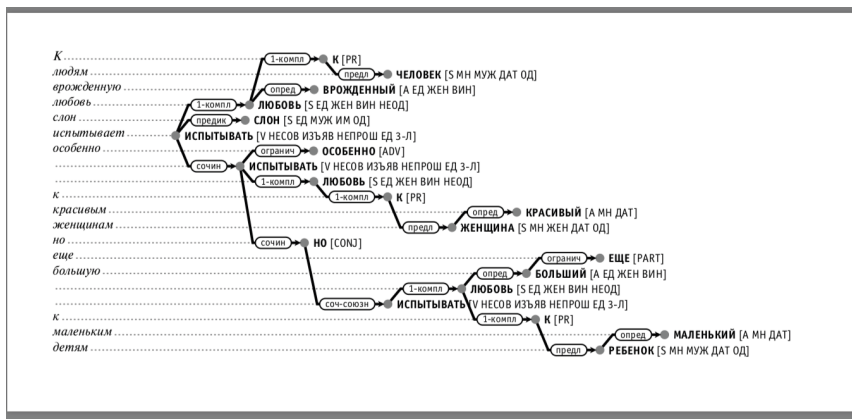


Рис. 2. Синтаксическая структура предложения с эллипсисом

К эллипсису относятся как случаи, когда можно дословно восстановить пропущенное, так и случаи, когда очевиден только смысл пропуска, но не его конкретное лексическое наполнение. Нулевая связка не считается эллипсисом и в структуре не восстанавливается. На дан-

ный момент в СинТагРус'е количество восстановленных элидированных слов составляет 2693.

5. Разметка значений многозначных слов

Как уже говорилось, атрибут слова KSNNAME хранит имя статьи комбинаторного словаря (КС) систем ЭТАП. Различных лемм в СинТагРус'е на данный момент насчитывается 50 350, а статей КС цитируется 41 812. Тот факт, что лемм больше, чем статей КС, отчасти объясняется тем, что одна и та же лемма может соответствовать двум и более статьям КС. Это происходит в первую очередь потому, что в корпусе встречается заметное число слов, отсутствующих в комбинаторном словаре (это могут быть собственные имена, названия и просто редкие слова или редкие значения многозначных слов). Для всех таких единиц в качестве леммы используется начальная форма слова, а в качестве имени статьи КС выступает имя некоторой дежурной статьи КС (например, ФИКТ-ФАМИЛИЯ, ФИКТ-ЛИЧИМЯ, ФИКТ-МЕСТО и т. п.): понятно, что такие единицы встречаются многократно.

Принципы выделения значений в словарях системы ЭТАП обусловлены, во-первых, традицией академических словарей, во-вторых, исторической задачей обеспечения автоматического перевода, а в-третьих, ориентацией на автоматическую разметку. Отдельные значения выделяются тогда, когда это продиктовано нуждами перевода или семантического анализа и когда их можно достаточно надежно детектировать автоматически на основе контекста, в частности, по реализованным валентностям или лексическим функциям. Поэтому значения в КС по сравнению с обычными толковыми словарями укрупнены: слово ТОЧКА представлено в КС двумя статьями — ТОЧКА 1 объединяет все центральные значения («малый пространственный или временной объект», «знак препинания»), а ТОЧКА 2 обозначает действие по глаголу *точить*. А многозначное с точки зрения конвенционального словаря ЯДРО и вовсе представлено единственной статьей.

Из 41 812 статей КС, цитируемых в СинТагРус'е, 4414 имеют индекс. Если сгруппировать их по леммам, то получится 2787 наборов омонимов или многозначных слов. Эти наборы уступают в подробности и точности сведениям из толковых словарей, но, несомненно, могут быть использованы для прицельного поиска. Кроме того, интересно исследовать, насколько такие укрупненные и ориентированные

на контекст значения подходят для решения задач автоматического разрешения неоднозначности.

6. Микросинтаксическая разметка

Под микросинтаксическими единицами (далее — МЕ) понимается широкий класс конструкций, которые находятся на стыке грамматики и лексики и характеризуются высокой степенью идиоматичности (см., например, [Иомдин 2006]). Словарь МЕ на сегодня насчитывает 2132 единицы. Общее количество их вхождений в СинТагРус составляет 18 939. Эти вхождения распределены по 14 836 предложениям: в некоторых предложениях содержится больше одной МЕ. Таким образом, МЕ содержатся примерно в 17 % предложений. Как уже отмечалось выше, информация о МЕ в xml-файл записывается в виде атрибута предложения. Начало и конец единицы задаются с помощью идентификаторов ее первого и последнего слова. Кроме того, указывается имя единицы: по нему ее можно найти в соответствующем словаре.

Следует иметь в виду, что некоторые МЕ совпадают со словами-оборотами типа *что бы то ни было*. В частности, само это слово тоже представляет собой МЕ. Однако эта информация записывается не в теге W, а в теге предложения S в таком виде:

```
<S DATE=>19 03 2021 14:15:34> ID=>181> MICROSYNТ=>(что бы то ни было,{6:что бы то ни было...6:что бы то ни было})&#xA;>
```

В данном случае МЕ однословна (в понимании СинТагРус'a), но подавляющее большинство МЕ не таковы. Неоднословные с точки зрения СинТагРус'a МЕ имеют внутреннюю синтаксическую структуру, и при поиске по синтаксическим условиям лингвисту может потребоваться исключить из выборки примеры, если они находятся внутри МЕ, или, наоборот, исследовать именно МЕ с точки зрения их внутренней синтаксической организации.

7. Анафорическая разметка

Анафорическая разметка устанавливает связи между местоимениями (3-го лица, *себя, свой, друг друга, тот, который, кто, чей*) и синтаксическими вершинами их антецедентов. Черновой вариант разметки выполняется процессором серии ЭТАП, снабженным специальными правилами. Затем разметка редактируется лингвистом. На сегодняшний день анафорическая разметка охватывает не весь корпус

и недоступна для поиска на сайте НКРЯ. Подробнее об этой разметке см. [Иншакова 2019].

8. Лексико-функциональная разметка

Лексико-функциональная разметка отмечает словосочетания, которые могут быть описаны в терминах лексических функций модели «Смысл-Текст» (см., в частности, [Mel'chuk 1996; Апресян 2011]). Она охватывает весь корпус. Лексико-функциональную разметку можно мыслить как разметку связей (хотя и не синтаксических в изложенном выше понимании): лексико-функциональное отношение связывает слово-аргумент с его значением. Таких связей на сегодняшний день в СинТагРус'е установлено 34 106. Количество предложений с лексико-функциональной разметкой составляет 24 516 (приблизительно 28 % всех предложений), в одном предложении может содержаться больше одной лексико-функциональной связи. Есть примеры, когда слово является аргументом сразу двух различных лексических функций, и примеры, когда значение одной лексической функции является аргументом другой. Рассмотрим предложение

*Но я никак не мог **найти ответ на вопрос**: почему я вынужден здесь **придерживаться строгого распорядка** дня, когда, пребывая у бабушки, я мог бы **вставать, ложиться и развлекаться по собственному усмотрению**?*

В этом предложении реализовано 4 лексико-функциональных связи: слово *распорядок* является аргументом лексических функций REAL1-M и MAGN, значениями которых являются слова *придерживаться* и *строгий* соответственно. А слово *ответ* является аргументом лексической функции OPER1 со значением *найти* и одновременно значением лексической функции S0-REAL3-M для аргумента *вопрос*. Эта информация записана в xml-файле следующим образом:

```
<LF LFARG=»16» LFFUNC=»_REAL1-M» LFVAL=»14»/>  
<LF LFARG="16" LFFUNC="_MAGN" LFVAL="15"/>  
<LF LFARG="9" LFFUNC="_S0_REAL3-M" LFPREP="8" LFVAL="7"/>  
<LF LFARG="7" LFFUNC="_OPER1" LFVAL="6"/>
```

За каждую лексико-функциональную связь отвечает отдельный элемент LF, имеющий атрибуты LFARG (аргумент), LFFUNC (имя функции), LFVAL (значение), LFPREP (вспомогательный предлог). Значениями атрибутов являются идентификаторы слов в пределах предложения, как и во всех остальных типах разметки.

9. Темпоральная разметка

Темпоральная разметка основывается на синтаксической. Каждое темпоральное выражение — это выделенный в синтаксической структуре подграф. Ему приписан набор специальных признаков, описывающих его семантику. Например, в предложении *17 октября пал Таганрог* темпоральным выражением является *17 октября*. Оно относится к событию, обозначенному глаголом *пасть*, и имеет признаки «локализация» и «календарное». Первый признак означает, что это выражение локализует событие во времени, а второй — что оно передает календарную информацию, нужную для логических выводов о времени. Темпоральными выражениями считаются любые выражения, описывающие протекание событий во времени, в том числе те, которые упорядочивают события друг относительно друга — *после войны*. Сведения о темпоральных выражениях, как и сведения о лексико-функциональных связях, хранятся в специальных элементах, обозначаемых как ТЕ.

```
<TE HEAD="1" WORDS="1, 2" DESCRIPTION="локализация|календарное" RELATED_TO="3"/>
```

Атрибут HEAD указывает синтаксическую вершину временного выражения, WORDS содержит перечисление всех входящих в него слов, DESCRIPTION — набор признаков. Значениями атрибутов, как и в других разметках, являются номера-идентификаторы в пределах текущего предложения.

Это самый новый тип разметки СинТагРуса, первые файлы были размечены в ручном режиме только в прошлом году. Очевидно, что даже в таком простом виде эта разметка многое сообщает о событиях. Планируется обогащение схемы разметки для дальнейших исследований семантики событий в тексте.

10. Организация поиска

Описанные выше разметки можно разделить на два больших класса: на те, которые характеризуют слова, и на те, которые характеризуют связи слов. К первому классу будут относиться морфологическая разметка, разметка значений слов, микросинтаксическая разметка и разметка эллипсиса. Микросинтаксическая разметка попадает в этот класс потому, что существенным лингвистическим фактом является включенность или невключенность слова в состав микросин-

таксической единицы. Разметка эллипсиса относится к этому классу потому, что эллипсис понимается как пропуск слов, соответственно, про каждое слово известно, является оно восстановленным пропуском или нет. Остальные типы разметки — синтаксическая, анафорическая, лексико-функциональная, темпоральная — задают связи между словами. Соответственно, для полноценной поисковой работы требуется интерфейс, позволяющий строить запросы из нескольких слов, произвольным образом связанных между собой. Кроме того, данные небольшого опроса, проведенного нами среди пользователей синтаксически размеченных корпусов, показывают, что информация о линейном расположении слов является важным поисковым критерием. Она востребована тремя четвертями пользователей. Поиск по линейным условиям — это прототипический поиск контекста: в любом корпусе без исключения есть возможность искать не только слово само по себе, но и сочетание слов, и в этом случае задаются не только характеристики второго слова, но и расстояние от первого, в пределах которого оно должно встретиться. Можно сказать, что линейное расстояние в этом случае связывает два слова, так что задание линейного контекста и задание ограничений по всем разным типам разметки, предполагающим связь между словами, при поиске функционально эквивалентны.

Разумеется, запросы составляются с целью найти некий объект, но по достижении запросом определенной сложности перестает быть очевидно, какой именно объект находится в фокусе поиска. Когда ищется определенная словоформа или слово, или отдельное синтаксическое отношение, или лексическая функция определенного типа, определить объект не составляет труда. Однако как только в запросе оказывается несколько частей или условий, появляется неоднозначность. Допустим, ищется слово, участвующее в роли хозяина в двух разных синтаксических отношениях, одном повторимом и одном неповторимом. Это может быть глагол, управляющий одним из комплетивных отношений (они относятся к неповторимым, то есть таким, по которым можно иметь не больше одного зависимого) и обстоятельством отношением (это повторимое отношение). Условию удовлетворяет следующее предложение:

У наших современников есть возможность встретиться с инопланетянами 24 декабря 2013 года, когда они могут прилететь, то есть после последнего их посещения пройдет как раз 3600 лет.

Глагол *встретиться* управляет предложением с по первому ком- плексивному отношению, а также является хозяином двух обстоятель- ственных связей — с числительным *24* и с союзом *когда*. Фокусом запроса лингвиста в данном случае может быть либо слово *встре- титься*, которое в указанном предложении встречается 1 раз, либо вся лексико-синтаксическая конфигурация в целом, которая в пред- ложении встречается дважды. Нужен эксплицитный способ, позволя- ющий обозначить фокус лингвистического интереса. Для этого следу- ет добавить к каждому поисковому условию, задающему связь между словами, возможность использовать ограничитель «есть хотя бы 1». Это различие актуально не только для синтаксически аннотирован- ных корпусов.

11. Заключение

Чтобы сделать доступным все богатство разметки СинТагРуса, не- обходим интерфейс, который поддерживает поиск сочетаний слов, связанных разными типами связей. Имеющийся в НКРЯ интерфейс, ориентированный на синтаксис, необходимо существенно перестро- ить. Должен присутствовать также поиск по линейному контексту. Способ задания линейных условий должен быть аналогичен тому, который применяется для задания прочих связей. При этом должна быть учтена специфика выделения словоформ в СинТагРусе. Кроме того, относительно каждого условия, касающегося какой-либо связи, должна быть возможность указать, как его следует понимать — «хотя бы 1 связь» или «строго 1 связь». Микросинтаксическая разметка и разметка эллипсиса должны быть представлены в виде параметров: является ли слово частью микросинтаксической единицы и является ли оно восстановленным пропуском.

Литература

1. Апресян Ю. Д. (2011), К новой версии теории лексических функций (ЛФ). Междуна- родная конференция, посвященная 50-летию Петербургской типологиче- ской школы: Материалы и тезисы докладов, СПб, с. 21–26.
2. Иншакова Е. С., Иомдин Л. Л., Митюшин Л. Г., Сизов В. Г., Фролова Т. И., Цин- ман Л. Л. (2019), СинТагРус сегодня. Труды Института русского языка им. В. В. Виноградова. М., т. 21, с. 14–41. doi: 10.31912/pvrtli-2019.21.1
3. Иншакова Е. С. (2019), Система разрешения анафоры для русского языка на базе лингвистического процессора ЭТАП-4. Компьютерная лингвистика и интеллек- туальные технологии: По материалам ежегодной международной конференции

«Диалог» (г. Москва, 29 мая — 1 июня 2019 г.). М.: РГГУ, вып. 18(25), с. 239–251. ISSN 2221-7932.

4. *Иомдин Л. Л.* (2006), Многозначные синтаксические фраземы: между лексикой и синтаксисом. Компьютерная лингвистика и интеллектуальные технологии («Диалог-2006»). Труды международной конференции. Бекасово, 31 мая — 4 июня 2006 г. М.: Изд-во РГГУ, с. 202–206.
5. *Mel'chuk I.* (1996), Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In: L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia, pp. 37–102.

References

1. *Apresjan Ju. D.* (2011), К новой версии теории лексических функций (LF) [The theory of lexical functions renewed]. In: *Mezhdunarodnaja konferencija, posvsjaschennaja 50-letiju Peterburgskoj tipologičeskoj shkoly: Materialy i tezisy dokladov* [The international conference dedicated to the 50th anniversary of the typological school in St. Petersburg: Materials and abstracts], St. Petersburg, pp. 21–26.
2. *Inshakova E. S., Iomdin L. L., Mitjushin L. G., Sizov V. G., Frolova T. I., Cinman L. L.* (2019), *SinTagRus segodnja* [The SynTagRus Today]. In: *Trudy Instituta russkogo jazyka im. V. V. Vinogradova* [Proceedings of Vinogradov Russian Language Institute]. Vol. 21, Moscow, pp. 14–41. DOI: 10.31912/pvri-2019.21.1
3. *Inshakova E. S.* (2019). Sistema razreshenija anafory dlja russkogo jazyka na baze lingvističeskogo protссора ETAP-4 [An anaphora resolution system for Russian based on ETAP-4 linguistic processor]. In: *Komp'yuternaja lingvistika i intellektual'nye tekhnologii. Po materialam ezhegodnoi mezhdunarodnoi konferentsii "Dialog"* [Computational Linguistics and Intellectual Technologies. Proceedings of International Conference "Dialog"], pp. 239–251.
4. *Iomdin L. L.* (2006), Многозначные синтаксические фраземы: между лексикой и синтаксисом [Ambiguous syntactic phrasemes: between lexicon and syntax]. In: *Komp'yuternaja lingvistika i intellektual'nye tekhnologii ("Dialog-2006". Tруды международной конференции. Бекасово, 31 мая — 4 июня 2006)* [Computational Linguistics and Intellectual Technologies. Proceedings of International Conference "Dialog-2006", Beкасово, May 31 - June 4, 2006]. Moscow: RGGU, pp. 202–206.
5. *Mel'chuk I.* (1996), Lexical Functions: A Tool for the Description of Lexical Relations in the Lexicon. In: L. Wanner (ed.), *Lexical Functions in Lexicography and Natural Language Processing*, Amsterdam/Philadelphia, pp. 37–102.

Тимошенко Светлана Петровна

Институт проблем передачи информации им. А. А. Харкевича РАН (Россия)

Timoshenko Svetlana

Kharkevich Institute for Information Transmission Problems of
the Russian Academy of Sciences (Russia)

E-mail: timoshenko@iitp.ru

Иомдин Леонид Леонидович

Институт проблем передачи информации им. А. А. Харкевича РАН (Россия)

Iomdin Leonid

Kharkevich Institute for Information Transmission Problems of
the Russian Academy of Sciences (Russia)

E-mail: iomdin@gmail.com

Гладилин Сергей Александрович

Институт проблем передачи информации им. А. А. Харкевича РАН (Россия)

Gladilin Sergey

Kharkevich Institute for Information Transmission Problems of
the Russian Academy of Sciences (Russia)

E-mail: gladilin@iitp.ru

Иншакова Евгения Сергеевна

Институт проблем передачи информации им. А. А. Харкевича РАН (Россия)

Inshakova Evgenia

Kharkevich Institute for Information Transmission Problems of
the Russian Academy of Sciences (Russia)

E-mail: e.s.inshakova@gmail.com