

«Аспектуальные пары: префиксальный vs. суффиксальный способ образования»

В.Д. Соловьев, В.В. Бочкарев, В.Р. Байрашева

Казанский федеральный университет

Абстракт. В литературе нет общей точки зрения по вопросу, находятся ли глаголы в аспектуальных парах в словоизменительных или словообразовательных отношениях. При этом в ряде публикаций префиксальный и суффиксальный способы образования аспектуальных пар противопоставляются в этом отношении. В публикациях указывалось на необходимость развития в этом вопросе новых количественных подходов на корпусах текстов. Мы предлагаем два новых подхода, сопоставляющих количественные характеристики аспектуальных пар обоих типов. Один подход основан на корпусе Google Books Ngram и анализирует динамику частот употребления слов в парах. Рассматриваются аспектуальные пары из баз данных, созданных Л. Яндой и О. Ляшевской. Для численной оценки степени близости графиков частот использованы коэффициенты корреляции Пирсона. Вторым подходом вводится числовая характеристика семантической близости глаголов в парах с использованием современных компьютерных методов, таких как векторное представление слов. Семантическая близость глаголов рассчитывается как стандартная косинусная мера между векторами, репрезентирующими сочетаемость рассматриваемых глаголов в корпусах. Рассмотрено несколько компьютерных моделей и корпусов текстов. Оба подхода не выявили существенных численных различий между префиксальным и суффиксальным способом образования пар. Это находится в хорошем согласии с результатами раннего исследования Л. Янды и О. Ляшевской.

Ключевые слова: русский язык, вид, аспектуальные пары, количественный анализ, корпуса, векторное представление слов, словоизменение, словообразование

1. Введение

Несмотря на длительную историю изучения русского вида остается неясным как следует его трактовать с точки зрения грамматического статуса – находятся ли глаголы в аспектуальных парах в словоизменительных или словообразовательных отношениях. Основные точки зрения следующие.

Вид – словоизменительная категория (это классический подход В. В. Виноградова (Виноградов 1947));

Вид – словоклассифицирующая категория (Русская грамматика 1980, Петрухина 2000, Janda 2013, Плунгян 2000);

Вид – словообразовательная категория (Всеволодова 1997, Filip 1999);

Префиксальное образование перфективов от базовых глаголов несовершенного вида – это словообразование, а суффиксальное образование вторичных имперфективов от перфективов – словоизменение (пожалуй, наиболее часто представляемая точка зрения (Karcevski 1927, Маслов 1984, Бондарко 1971)). Таким образом, эти два способа видовой трансформации тесно связаны с грамматическим статусом аспектуальных пар.

В последнее десятилетие эта проблема активно исследовалась в работах Е.В. Горбовой и В.С. Храковского (Горбова 2014, 2015а, 2015б, 2017, Храковский 2015, 2018), однако, к консенсусу они так и не пришли. Проблематика русского вида подробно освещена в монографии (Зализняк 2015).

К сожалению, нет общепризнанного формального (операционного) критерия разграничения словообразования и словоизменения. В различных работах (Dressler 1989, Plank 1991, Перцов 1996, Мельчук 1997) обсуждались разные признаки этих категорий. Авторы работы (Singh 1980) вообще настаивают на отмене этого разграничения. В (Мельчук 1997) предлагается ввести промежуточную категорию квазиграммем. Все же есть общее интуитивное понимание различия между ними, хотя бы в прототипических случаях: при словоизменении само слово не меняется, меняется лишь его форма, а при словообразовании меняется слово, т.е. его значение. Фактически к этому же сводятся и формализованные определения в (Мельчук 1997). Таким образом, мы приходим к вопросу меняется ли семантика слова (кроме чисто видовой) при префиксальной и суффиксальной трансформации.

Проблема в том, что семантика не формализуема или крайне сложно формализуема. Предложенная А. Вежбицкой теория семантических примитивов (Вежбицкая 1999) – в данной задаче мало пригодна, прежде всего, в силу крайней сложности формул представления семантики, а также в силу сохраняющегося субъективизма в построении таких формул. Противоречивая оценка многих пар глаголов, как аспектуальных или нет, в работах (Зализняк 2012, Кузнецова, Янда 2013) показывает, что интуиция даже ведущих лингвистов-аспектологов недостаточна для установления тождества семантики. Это приводит к сделанному (Janda & Lyashevskaya 2011: стр. 201) выводу, что “At the level of intuitive analysis, it has not been possible to resolve this debate” и, следовательно, нужно применять статистические методы к корпусным данным для решения этих проблем.

Исследовательский вопрос, решаемый в данной статье – в какой степени меняется семантика слов при суффиксальном и префиксальном способах образования аспектуальных пар? Проверяемая гипотеза: семантическая близость между глаголами в аспектуальных парах

перфектив – вторичный имперфектив больше, чем между глаголами в парах базовый имперфектив – перфектив.

В данной статье мы предлагаем два подхода к получению объективной численной оценки степени семантической близости слов. Оба подхода применены к аспектуальным парам, получаемым префиксацией и суффиксацией. В итоге мы не нашли заметных количественных различий между этими двумя способами, которые позволили бы их четко дифференцировать как словообразовательные или словоизменительные. Это находится в хорошем согласии с работой (Janda & Lyashevskaya 2011), в которой также не найдено таких различий при применении еще одного количественного метода. В этой работе сравнивались грамматические профайлы – распределения форм глагола в Национальном корпусе русского языка – и получен следующий результат: “we do not find reportable differences between the grammatical profiles of aspectual partners formed with prefixes as opposed to suffixes” (Janda & Lyashevskaya 2011, стр. 213).

В этой статье мы используем предложенную в (Janda 2013) терминологию: перфективы в аспектуальных парах с базовыми имперфективами будем называть естественными, а иные перфективы – специализированными.

2. Данные и методы

2.1. Подход 1: динамика частотности употребления слова

Первый из двух предлагаемых нами подходов основан на использовании корпуса Google Books Ngram (в дальнейшем, сокращенно GBN). Корпус содержит для русского языка более 80 миллиардов слов (в 200 раз больше, чем Национальный корпус русского языка) и охватывает более 2 веков. Сервис визуализации Ngram Viewer (<https://books.google.com/ngrams/>) позволяет строить графики частот встречаемости слов и словосочетаний по годам.

Коллекция GBN содержит только тексты книг, и в этом смысле она не является всеохватывающей. С другой стороны, методика создания этого корпуса была такова (Michel et al. 2011): полное сканирование книг (более 15% от всех изданных в мире книг) более чем из 40 крупнейших библиотек по всему миру, включая библиотеки университетов Оксфорда, Гарварда и др. (<http://www.google.com/intl/en/googlebooks/library/partners.html>). Это обеспечило репрезентативное представление всех аспектов человеческой культуры, в той мере, в которой они представлены в изданных книгах, хранящихся в ведущих библиотеках. Вопросы, связанные со сбалансированностью и репрезентативностью GBN, обсуждаются в (Solovyev 2020, Richey 2020).

Наша основная идея состоит в следующем. Если два слова находятся в словоизменительном отношении и, соответственно, имеют тождественную лексическую

семантику, то частота употребления этих двух слов будет меняться синхронно, в зависимости от степени востребованности слов с этой семантикой в тот или иной период времени. Если же два слова находятся в словообразовательных отношениях, т.е. представляют собой два разных слова, то частоты их употребления могут меняться не синхронно, в зависимости от востребованности каждого из этих слов. Это следует, в частности, из дистрибутивной гипотезы (Sahlgren 2008). Таким образом, для того, чтобы оценить префиксальный и суффиксальный способы образования аспектуальных пар с этой точки зрения, требуется сопоставить динамику частот словоупотребления соответствующих глаголов. В GBN употребление слова репрезентируется временным рядом частот в последовательные годы. Мы используем интервал времени с 1920 г., чтобы избежать влияния старинной орфографии (до реформы 1917 г.), не всегда точно отражаемой в GBN.

Степень синхронности изменения частот употребления определяется близостью форм кривых – временных рядов и может быть подсчитана. Для численной оценки степени близости будем использовать коэффициенты корреляции Пирсона, как часто используемый для оценки близости временных рядов. Данный подход был впервые анонсирован на конференции 15th Slavic Cognitive Linguistics Conference (Соловьев 2017).

Будут использованы два набора данных. Аспектуальные пары с префиксальным способом видообразования взяты из базы данных Exploring Emptiness (Janda et al. 2013), <http://emptyprefixes.uit.no/>. Пары с суффиксальным типом видообразования – из базы данных видовых пар (О. Ляшевская, 2016, <http://ru-eval.ru/go/resources.html>).

При работе с временными рядами частот из GBN следует иметь в виду следующую особенность. Для низкочастотных слов возрастает влияние на форму графиков случайных факторов. Поэтому целесообразно анализировать по отдельности слова с разной частотностью. Будем сравнивать только слова, отличающиеся по частоте не более, чем в 10 раз. Т.е. делим все слова на группы с частотами f : $f > 10^{-4}$, $10^{-4} > f > 10^{-5}$ и т.д. Частота подсчитывается как средняя на интервале 1920-2008. Рассматриваются лишь аспектуальные пары, оба глагола которых принадлежат к одной частотной группе. В таблице 1 приведены усредненные коэффициенты корреляции по Пирсону временных рядов частот глаголов в аспектуальных парах.

Таблица 1. Средний коэффициент корреляции Пирсона частот глаголов в аспектуальных парах

Частота, log10	Префиксальное видообразование	Суффиксальное видообразование
-------------------	----------------------------------	----------------------------------

-4	0,65	0,61
-5	0,66	0,68
-6	0,53	0,58
-7	0,31	0,39
-8	0,19	0,27
-9	-	0,29
Всего пар	673	7800

Table 1. Average Pearson correlation coefficient of verb frequencies in aspectual pairs

Frequency, log10	Prefix aspectual formation	Suffix aspectual formation
-4	0,65	0,61
-5	0,66	0,68
-6	0,53	0,58
-7	0,31	0,39
-8	0,19	0,27
-9	-	0,29
Total pairs	673	7800

Как видно из таблицы, для наиболее высокочастотных глаголов с $f > 10^{-4}$ корреляция несколько выше для префиксального способа видообразования. Для следующих двух частотных групп – для суффиксального способа видообразования, причем тоже не намного выше. Для низкочастотных глаголов в следующих группах коэффициент корреляции имеет небольшие значения в силу фактора, о котором было сказано выше. Таким образом, при таком подсчете существенной разницы не обнаруживается.

Одним из возможных влияющих на результат факторов может быть многозначность глаголов. Это снижает значения коэффициентов корреляции, так как график частот

словоупотреблений строится один для всех значений слов, усредняя частоты для всех его значений. Отберем некоторое число эталонных пар – достаточно высокочастотных глаголов, не являющихся омонимами или полисемичными словами с заведомо различными значениями, и рассчитаем коэффициенты корреляции для них. В таблице 2 приведены рассматриваемые пары слов. Список подготовлен Л. Яндой.

Таблица 2. Пары базовый имперфектив – естественный перфектив и специализированный перфектив – вторичный имперфектив

Базовый имперфектив	Естественный перфектив	Специализированный перфектив	Вторичный имперфектив
делать	сделать	переделать	переделывать
казаться	показаться	оказаться	оказываться
просить	попросить	выпросить	выпрашивать
ставить	поставить	приставить	приставлять
хранить	сохранить	охранить	охранять
прятать	спрятать	припрятать	припрятывать
плевать	наплевать	оплевать	оплевывать
растить	вырастить	отрастить	отращивать
жрать	сожрать	пожрать	пожирать
нюхать	понюхать	вынюхать	вынюхивать
щупать	пощупать	нащупать	нащупывать
играть	сыграть	выиграть	выигрывать
верить	поверить	проверить	проверять
звонить	позвонить	перезвонить	перезванивать
рисовать	нарисовать	срисовать	срисовывать
влечь	повлечь	привлечь	привлекать
жечь	сжечь	зажечь	зажигать
мерзнуть	замерзнуть	промерзнуть	промерзать
копать	выкопать	раскопать	раскапывать
красть	украсть	обокрасть	обкрадывать

Table 2. Pairs of basic imperfective – natural perfective and specialized perfective – secondary imperfective

Basic imperfective	Natural perfective	Specialized perfective	Secondary imperfective
делать	сделать	переделать	переделывать
казаться	показаться	оказаться	оказываться
просить	попросить	выпросить	выпрашивать
ставить	поставить	приставить	приставлять
хранить	сохранить	охранить	охранять
прятать	спрятать	припрятать	припрятывать
плевать	наплевать	оплевать	оплевывать
растить	вырастить	отрастить	отращивать
жрать	сожрать	пожрать	пожирать
нюхать	понюхать	вынюхать	вынюхивать
щупать	пощупать	нащупать	нащупывать
играть	сыграть	выиграть	выигрывать
верить	поверить	проверить	проверять
звонить	позвонить	перезвонить	перезванивать
рисовать	нарисовать	срисовать	срисовывать
влечь	повлечь	привлечь	привлекать
жечь	сжечь	зажечь	зажигать
мерзнуть	замерзнуть	промерзнуть	промерзать
копать	выкопать	раскопать	раскапывать
красть	украсть	обокрасть	обкрадывать

Расчет проводится, так же, как и для полного набора аспектуальных пар. Результаты приведены в таблице 3.

Таблица 3. Средний коэффициент корреляции Пирсона частот глаголов в аспектуальных парах из таблицы 2

Частота, log10	Префиксальное видообразование	Суффиксальное видообразование

-5	0,94	-
-6	0,75	0,72
-7	0,71	0,72
-8	-	0,53

Table 3. Average Pearson correlation coefficient of verb frequencies in aspectual pairs from table 2

Frequency, log10	Prefix aspectual formation	Suffix aspectual formation
-5	0,94	-
-6	0,75	0,72
-7	0,71	0,72
-8	-	0,53

Здесь минус означает отсутствие пар в данной частотной группе. Как и ожидалось, для эталонных пар коэффициент корреляции заметно выше, чем в таблице 1. Аспектуальные пары обоих типов трансформации присутствуют только в двух частотных группах: $10^{-5} > f > 10^{-6}$ и $10^{-6} > f > 10^{-7}$. В обеих группах коэффициенты корреляции почти равны.

Таким образом, при таком подходе различия в степени семантической близости при суффиксальном и префиксальном способах видообразования для эталонных пар также не обнаружены.

2.2. Подход 2: семантическое сходство в векторной семантике

Хорошо известной является дистрибутивная гипотеза, согласно которой лингвистические единицы, встречающиеся в схожих контекстах, имеют близкие значения. Основываясь на ней в компьютерной лингвистике разработаны методы векторной семантики, позволяющие формализовать семантику слова как вектор частот слов, с которыми оно вместе встречается в текстах корпуса. Семантическое сходство между словами легко вычислить, одним из стандартных математических методов, например, как косинусную меру сходства между векторами слов. Предложены различные методы построения векторов: skipgram, fasttext и др. (Mikolov 2013, Wojanowski 2017, Pennington, 2014). Данный подход и его последующие модификации широко используются в различных задачах обработки текстов, таких как классификация текстов, вопросно-ответные системы, суммаризация текстов (Sivakumar 2020, Peters 2018, Yang 2019). Для русского языка существуют системы, в которых

уже построены вектора для сотен тысяч слов на основе сверхбольших корпусов текстов (Grave 2018, Kutuzov 2015). Мы используем эти методы для расчета семантического сходства слов в аспектуальных парах.

В таблице 4 приведена величина сходства (в интервале от 0 до 1; 1 – максимальное сходство) для нашего набора эталонных пар. Для расчетов применялся инструментарий Rusvectors (<https://rusvectors.org/ru/misc/>), использующий модель skipgram, обученную на содержащем более 5 миллиардов слов корпусе Тайга (Shavrina 2017). В нее включены также значения сходства между базовым имперфективом и специализированным перфективом. Сходство приведено в столбцах, расположенных между соответствующими словами. В таблице 5 сходство рассчитывалось по предобученной модели fasttext без дополнительного дообучения, а в таблице 6 по skipgram, обученной на корпусе НКРЯ и Википедия. Различие в корпусах состоит в том, что Тайга состоит почти полностью (95%) из художественных текстов. Прочерк в таблицах ниже означает, что в данной вычислительной модели отсутствуют необходимые слова.

Таблица 4. Величина сходства между словами в векторной семантике, модель skipgram, корпус Тайга

	Естественный перфектив	Расстояние	Базовый имперфектив	Расстояние	Специализированный перфектив	Расстояние	Вторичный имперфектив
1	сделать	0,73	делать	0,20	переделать	0,73	переделывать
2	показаться	0,64	казаться	0,21	оказаться	0,69	оказываться
3	попросить	0,82	просить	0,45	выпросить	0,60	выпрашивать
4	поставить	0,81	ставить	0,30	приставить	0,78	приставлять
5	сохранить	0,52	хранить	0,38	охранить	0,50	охранять
6	спрятать	0,81	прятать	0,68	припрятать	0,82	припрятывать
7	наплевать	0,84	плевать	0,41	оплевать	0,50	оплевывать
8	вырастить	0,73	растить	0,27	отрастить	0,69	отращивать
9	сожрать	0,62	жрать	0,60	пожрать	0,48	пожирать
10	понюхать	0,76	нюхать	0,34	вынюхать	0,59	вынюхивать
11	пощупать	0,70	щупать	0,46	нащупать	0,48	нащупывать
12	сыграть	0,65	играть	0,27	выиграть	0,81	выигрывать
13	поверить	0,76	верить	0,19	проверить	0,81	проверять
14	позвонить	0,77	звонить	0,71	перезвонить	0,75	перезванивать
15	нарисовать	0,80	рисовать	0,42	срисовать	0,55	срисовывать
16	повлечь	0,63	влечь	0,39	привлечь	0,76	привлекать
17	сжечь	0,58	жечь	0,38	зажечь	0,82	зажигать

18	замерзнуть	0,63	мерзнуть	0,52	промерзнуть	0,74	промерзать
19	выкопать	0,70	копать	0,50	раскопать	0,71	раскапывать
20	украсть	0,50	красть	0,46	обокрасть	0,67	обкрадывать
Среднее		0,70		0,41		0,67	

Table 4. The value of the similarity between words in vector semantics, skipgram model, Taiga corpus

	Natural perfective	Distance	Basic imperfective	Distance	Specialized perfective	Distance	Secondary imperfective
1	сделать	0,73	делать	0,20	переделать	0,73	переделывать
2	показаться	0,64	казаться	0,21	оказаться	0,69	оказываться
3	попросить	0,82	просить	0,45	выпросить	0,60	выпрашивать
4	поставить	0,81	ставить	0,30	приставить	0,78	приставляя
5	сохранить	0,52	хранить	0,38	охранить	0,50	охранять
6	спрятать	0,81	прятать	0,68	припрятать	0,82	припрятывать
7	наплевать	0,84	плевать	0,41	оплевать	0,50	оплевывать
8	вырастить	0,73	растить	0,27	отрастить	0,69	отращивать
9	сожрать	0,62	жрать	0,60	пожрать	0,48	пожирать
10	понюхать	0,76	нюхать	0,34	вынюхать	0,59	вынюхивать
11	пощупать	0,70	щупать	0,46	нащупать	0,48	нащупывать
12	сыграть	0,65	играть	0,27	выиграть	0,81	выигрывать
13	поверить	0,76	верить	0,19	проверить	0,81	проверять
14	позвонить	0,77	звонить	0,71	перезвонить	0,75	перезванивать
15	нарисовать	0,80	рисовать	0,42	срисовать	0,55	срисовывать
16	повлечь	0,63	влечь	0,39	привлечь	0,76	привлекать
17	сжечь	0,58	жечь	0,38	зажечь	0,82	зажигать
18	замерзнуть	0,63	мерзнуть	0,52	промерзнуть	0,74	промерзать
19	выкопать	0,70	копать	0,50	раскопать	0,71	раскапывать
20	украсть	0,50	красть	0,46	обокрасть	0,67	обкрадывать
Mean		0,70		0,41		0,67	

Таблица 5. Величина сходства между словами в векторной семантике, модель fasttext

	Естественный перфектив	Расстояние	Базовый имперфектив	Расстояние	Специализированный перфектив	Расстояние	Вторичный имперфектив
1	сделать	0,76	делать	0,51	переделать	0,76	переделывать
2	показаться	0,74	казаться	0,63	оказаться	0,63	оказываться
3	попросить	0,82	просить	0,59	выпросить	0,65	выпрашивать

4	поставить	0,81	ставить	0,41	приставить	0,82	приставлять
5	сохранить	0,59	хранить	0,39	охранить	0,64	охранять
6	спрятать	0,80	прятать	0,65	припрятать	0,64	припрятывать
7	наплевать	0,89	плевать	0,49	оплевать	0,71	оплевывать
8	вырастить	0,74	растить	0,45	отрастить	0,78	отращивать
9	сожрать	0,63	жрать	0,61	пожрать	0,47	пожирать
10	понюхать	0,71	нюхать	0,51	вынюхать	0,71	вынюхивать
11	пощупать	0,67	щупать	0,45	нащупать	0,76	нащупывать
12	сыграть	0,76	играть	0,48	выиграть	0,75	выигрывать
13	поверить	0,72	верить	0,40	проверить	0,75	проверять
14	позвонить	0,76	звонить	0,54	перезвонить	0,63	перезванивать
15	нарисовать	0,75	рисовать	0,55	срисовать	0,80	срисовывать
16	повлечь	0,64	влечь	0,51	привлечь	0,79	привлекать
17	сжечь	0,66	жечь	0,59	зажечь	0,77	зажигать
18	замерзнуть	0,75	мерзнуть	0,60	промерзнуть	0,76	промерзать
19	выкопать	0,61	копать	0,58	раскопать	0,71	раскапывать
20	украсть	0,69	красть	0,50	обокрасть	0,53	обкрадывать
Среднее		0,73		0,52		0,70	

Table 5. The value of the similarity between words in vector semantics, fasttext model

	Natural perfective	Distance	Basic imperfective	Distance	Specialized perfective	Distance	Secondary imperfective
1	сделать	0,76	делать	0,51	переделать	0,76	переделывать
2	показаться	0,74	казаться	0,63	оказаться	0,63	оказываться
3	попросить	0,82	просить	0,59	выпросить	0,65	выпрашивать
4	поставить	0,81	ставить	0,41	приставить	0,82	приставлять
5	сохранить	0,59	хранить	0,39	охранить	0,64	охранять
6	спрятать	0,80	прятать	0,65	припрятать	0,64	припрятывать
7	наплевать	0,89	плевать	0,49	оплевать	0,71	оплевывать
8	вырастить	0,74	растить	0,45	отрастить	0,78	отращивать
9	сожрать	0,63	жрать	0,61	пожрать	0,47	пожирать
10	понюхать	0,71	нюхать	0,51	вынюхать	0,71	вынюхивать
11	пощупать	0,67	щупать	0,45	нащупать	0,76	нащупывать
12	сыграть	0,76	играть	0,48	выиграть	0,75	выигрывать
13	поверить	0,72	верить	0,40	проверить	0,75	проверять
14	позвонить	0,76	звонить	0,54	перезвонить	0,63	перезванивать
15	нарисовать	0,75	рисовать	0,55	срисовать	0,80	срисовывать
16	повлечь	0,64	влечь	0,51	привлечь	0,79	привлекать
17	сжечь	0,66	жечь	0,59	зажечь	0,77	зажигать
18	замерзнуть	0,75	мерзнуть	0,60	промерзнуть	0,76	промерзать

19	выкопать	0,61	копать	0,58	раскопать	0,71	раскапывать
20	украсть	0,69	красть	0,50	обокрасть	0,53	обкрадывать
Mean		0,73		0,52		0,70	

Таблица 6. Величина сходства между словами в векторной семантике, модель skipgram, корпус НКРЯ и Википедия

	Естественный перфектив	Расстояние	Базовый имперфектив	Расстояние	Специализированный перфектив	Расстояние	Вторичный имперфектив
1	сделать	0,77	делать	0,19	переделать	0,73	переделывать
2	показаться	0,69	казаться	0,38	оказаться	0,76	оказываться
3	попросить	0,79	просить	0,53	выпросить	0,62	выпрашивать
4	поставить	0,86	ставить	0,39	приставить	0,69	приставлять
5	сохранить	0,53	хранить	0,39	охранить	0,45	охранять
6	спрятать	0,80	прятать	0,72	припрятать	0,67	припрятывать
7	наплевать	0,65	плевать	0,51	оплевать	0,56	оплевывать
8	вырастить	0,61	растить	0,27	отрастить	0,58	отращивать
9	сожрать	0,51	жрать	0,51	пожрать	0,54	пожирать
10	понюхать	0,69	нюхать	-	вынюхать	-	вынюхивать
11	пощупать	0,65	щупать	0,51	нащупать	0,70	нащупывать
12	сыграть	0,70	играть	0,27	выиграть	0,82	выигрывать
13	поверить	0,78	верить	0,33	проверить	0,82	проверять
14	позвонить	0,75	звонить	0,63	перезвонить	0,60	перезванивать
15	нарисовать	0,80	рисовать	0,53	срисовать	0,52	срисовывать
16	повлечь	0,69	влечь	0,34	привлечь	0,73	привлекать
17	сжечь	0,56	жечь	0,49	зажечь	0,76	зажигать
18	замерзнуть	0,57	мерзнуть	0,51	промерзнуть	0,62	промерзать
19	выкопать	0,68	копать	0,45	раскопать	0,63	раскапывать
20	украсть	0,44	красть	0,39	обокрасть	-	обкрадывать
Среднее		0,69		0,42		0,66	

Table 6. The value of the similarity between words in vector semantics, skipgram model, RNC corpus and Wikipedia

	Natural perfective	Distance	Basic imperfective	Distance	Specialized perfective	Distance	Secondary imperfective
1	сделать	0,77	делать	0,19	переделать	0,73	переделывать
2	показаться	0,69	казаться	0,38	оказаться	0,76	оказываться
3	попросить	0,79	просить	0,53	выпросить	0,62	выпрашивать

4	поставить	0,86	ставить	0,39	приставить	0,69	приставлять
5	сохранить	0,53	хранить	0,39	охранить	0,45	охранять
6	спрятать	0,80	прятать	0,72	припрятать	0,67	припрятывать
7	наплевать	0,65	плевать	0,51	оплевать	0,56	оплевывать
8	вырастить	0,61	растить	0,27	отрастить	0,58	отращивать
9	сжрать	0,51	жрать	0,51	пожрать	0,54	пожирать
10	понюхать	0,69	нюхать	-	вынюхать	-	вынюхивать
11	пощупать	0,65	щупать	0,51	нащупать	0,70	нащупывать
12	сыграть	0,70	играть	0,27	выиграть	0,82	выигрывать
13	поверить	0,78	верить	0,33	проверить	0,82	проверять
14	позвонить	0,75	звонить	0,63	перезвонить	0,60	перезванивать
15	нарисовать	0,80	рисовать	0,53	срисовать	0,52	срисовывать
16	повлечь	0,69	влечь	0,34	привлечь	0,73	привлекать
17	сжечь	0,56	жечь	0,49	зажечь	0,76	зажигать
18	замерзнуть	0,57	мерзнуть	0,51	промерзнуть	0,62	промерзать
19	выкопать	0,68	копать	0,45	раскопать	0,63	раскапывать
20	украсть	0,44	красть	0,39	обокрасть	-	обкрадывать
Mean		0,69		0,42		0,66	

3. Обсуждение результатов

Анализируя данные таблиц 4, 5, 6 следует отметить следующее.

1. Сравнение сходства в парах “базовый имперфектив – естественный перфектив” с парами “базовый имперфектив – специализированный перфектив”.

Как и следовало ожидать, сходство между словами в парах “базовый имперфектив – естественный перфектив” (столбец 3) значительно выше по сравнению с парами “базовый имперфектив – специализированный перфектив” (столбец 5). Различие в семантике слов в парах “базовый имперфектив – естественный перфектив” обусловлено только различием в виде, а в парах “базовый имперфектив – специализированный перфектив” кроме вида семантика глаголов различается еще и семантикой приставки. Это подтверждает правильность предложенного метода вычисления семантической схожести в данной задаче. Отметим, что не только в среднем, но и в каждой строке значение в третьем столбце выше, чем в пятом (исключение в строке 9 со словом *жрать*, где имеет место равенство значений).

2. Сравнение сходства в парах “базовый имперфектив – специализированный перфектив” с парами “специализированный перфектив – вторичный имперфектив”.

Аналогичное соотношение мы видим и при сопоставлении пар “базовый имперфектив – специализированный перфектив” с парами “специализированный перфектив – вторичный имперфектив”. Обратим внимание на интересное исключение в таблицах 4 и 5. В паре *пожрать – пожирать* степень близости слов меньше, чем в паре *жрать – пожрать*. Это

указание на то, что *пожрать* – *пожирать* не является чистой аспектуальной парой, *пожирать* имеет какую-то дополнительную семантику по сравнению с *пожрать*. Проведем лингвистический анализ этих слов по их толкованиям, чтобы выявить возможные расхождения в семантике. Возьмем толкования в словаре Малом академическом словаре (Евгеньева 1981-84). *Пожрать* означает то же, что и *поест* с дополнительным значением ‘с жадностью’ (о животных) или грубое протонародное выражение, если говорится о человеке. Слово *пожирать* кроме этого имеет еще и переносные значения: 1) Быстро, с жадностью и интересом прочитывать (книги, журналы и т. д.). Сочетание ?? *пожрал книги* маргинально. В Google это словосочетание встречается 27 раз, в то время как *пожирал книги* – 6600 раз. 2) Поглощать, требовать для себя что-л. в больших количествах. 3) Уничтожать, сжигая (об огне, пламени). 4) Захватывать целиком (о чувствах). Таким образом, между словами *пожрать* и *пожирать* действительно есть значительные семантические различия, кроме видовых.

То, что суффиксальное образование вторичного имперфектива не всегда является чисто видовой трансформацией ранее отмечалось, например, в работе (Апресян 1995: 112), в которой указано, что глагол *выпивать* имеет дополнительное значение кратности, а чисто видовой парой с *выпить* является именно *пить*.

3. Сравнение средних значений сходства. Средние значения близости в парах “базовый имперфектив – естественный перфектив” и “специализированный перфектив – вторичный имперфектив” очень близки и существенно превышают среднее значение в парах “базовый имперфектив – специализированный перфектив”. При расчетах по модели skipgram с корпусом Тайга в 10 случаях значение близости выше для пар “базовый имперфектив – естественный перфектив”, в 9 случаях – для пар “специализированный перфектив – вторичный имперфектив” (в одном случае равенство).

При расчетах по модели fasttext в 8 случаях значение близости выше для пар “базовый имперфектив - естественный перфектив”, в 10 случаях – для пар “специализированный перфектив - вторичный имперфектив” (в двух случаях равенство).

При расчетах по модели skipgram с корпусами НКРЯ и Википедия в 10 случаях значение близости выше для пар “базовый имперфектив – естественный перфектив”, в 8 случаях – для пар “специализированный перфектив – вторичный имперфектив”.

Таким образом, хотя точные значения величины близости для пар слов несколько различаются для разных вычислительных моделей, общие закономерности сохраняются. Оба способа образования аспектуальных пар – префиксальный и суффиксальный – в одинаковой степени меняют семантику слов.

Полученные численные данные могут быть использованы для получения и других результатов. Рассчитаем коэффициенты корреляции по Пирсону между столбцами 3, 5, 7; результаты приведены в таблице 7.

Таблица 7. Корреляция между разными способами образования пар

	Модель fastText	Модель skipgram+Тайга	Модель skipgram+ НКРЯ+Wiki
Префиксальная аспектуальная пара – суффиксальная аспектуальная пара	0,173	0,018	0,332
Префиксальная аспектуальная пара – способ глагольного действия	0,065	0,097	0,040
Суффиксальная аспектуальная пара – способ глагольного действия	-0,441	-0,150	-0,371

Table 7. Correlation between different ways of pairing

	Model fastText	Model skipgram+Тайга	Model skipgram+ RNC and Wiki
Prefixal aspectual pair – suffixal aspectual pair	0,173	0,018	0,332
Prefixal aspectual pair – Mode of verb action	0,065	0,097	0,040
Suffixal aspectual pair – Mode of verb action	-0,441	-0,150	-0,371

Между префиксальным и суффиксальным способами образования аспектуальных пар корреляция близости пар глаголов слабая (по Чеддоку). Это же касается и сравнения: префиксальные аспектуальные пары – способ глагольного действия. В то же время при сопоставлении данных по суффиксальному способу образования аспектуальных пар с парами, классифицируемыми, как способ глагольного действия, неожиданно обнаруживается значительно более высокий уровень корреляции, причем с отрицательным знаком. Для модели

fastText эта корреляция характеризуется как умеренная, близкая к средней: -0,441. Качественно это означает, что чем меньший вклад вносит приставка в семантику специализированного перфектива по сравнению с базовым имперфективом, тем в большей степени вторичный имперфектив будет семантически отличаться от специализированного перфектива. Насколько нам известно, такой эффект ранее не описывался. Он требует дополнительного изучения.

4. Заключение

Несмотря на длительную историю изучения аспекту в русском языке, не удалось прийти к единому мнению о грамматическом статусе аспектуальных пар – представляют ли они собой словоизменение или словообразование. С появлением больших корпусов текстов появилась надежда на то, что применение статистических, компьютерных методов позволит прояснить данный вопрос. Для постулирования словоизменительного или словообразовательного статуса аспектуальных пар, получаемых суффиксальным или префиксальным способом принципиальным является определение того, в какой степени глаголы в паре являются семантически близкими. В значительном числе работ предполагалось, что суффиксальный способ образования аспектуальных пары является словоизменительным, а префиксальный – словообразовательным. Соответственно, можно было бы ожидать, что суффиксальный способ не меняет семантики глаголов, за исключением вида, а префиксы вносят определенный вклад в семантику, кроме видового.

В статье проверялась эта гипотеза. Предложены два существенно различных подхода к получению точных количественных оценок семантической близости глаголов в аспектуальных парах. Оба подхода реализованы в двух вариантах. Во всех случаях не обнаружено заметной разницы в степени семантической близости глаголов в префиксальных и суффиксальных аспектуальных парах.

Аналогичный результат при применении совершенно иного подхода – сопоставлении грамматических профайлов глаголов – получен в работе (Janda & Lyashevskaya 2011). Таким образом, не удастся найти количественные различия в свойствах суффиксальных или префиксальных аспектуальных пар.

Обратим внимание на то, что второй из предложенных количественных подходов – с вычислением векторной близости между словами – широко используется в компьютерной лингвистике, но до сих пор не применялся в чистой лингвистике. Он может оказаться перспективным в исследованиях, связанных с семантической близостью слов.

5. **Благодарности.** Авторы благодарят Л. Янду и М. Исламова за вклад в подготовку материалов. Исследование поддержано РНФ, грант № 20-18-00206. Конфликт интересов отсутствует.

Вклад авторов. Соловьев В.Д. – основная идея, выполнение расчетов (таблицы 4, 6, 7), количественный анализ результатов, написание текста статьи. Бочкарев В.В. – выполнение расчетов (таблицы 1, 3). Байрашева В.Р. – лингвистический анализ примеров. Янда Л. – подготовка набора слов (таблица 2), Исламов М. – выполнение расчетов (таблица 5).

Список литературы/References

Апресян Ю.Д. Избранные труды, т.2. М.: Школа “Языки русской культуры”. 1995. [Aprėsjan Ju.D. Izbrannye trudy, T.2. M.: Shkola “Jazyki russkoj kul'tury”. 1995].

Виноградов В.В. Русский язык: грамматическое учение о слове. М. 1947. [Vinogradov V.V. Russkij jazyk: grammaticheskoe uchenie o slove. M. 1947].

Вежбицкая А., Семантические универсалии и описание языков. М., 1999. [Vezhbickaja A., Semanticheskie universalii i opisanie jazykov. M., 1999].

Евгеньева А. П. (ред.) Словарь русского языка: В 4-х т / АН СССР, Ин-т рус. яз. М.: Русский язык, 1981—1984. [A. P. Evgen'eva (red.) Slovar' russkogo jazyka: V 4-h t / AN SSSR, In-t rus. jaz. M.: Russkij jazyk, 1981—1984].

Всеволодова М.В. Семный состав глагольного слова. (К вопросу типологии вида) // Типология вида: проблемы, поиски, решения. Тезисы международной научной конференции. Московский государственный университет имени М.В. Ломоносова. Филологический факультет. М.: Изд-во Моск. ун-та. 1997. [Vsevolodova M.V. Semnyj sostav glagol'nogo slova. (K voprosu tipologii vida) // Tipologija vida: problemy, poiski, reshenija. Tezisy mezhdunarodnoj nauchnoj konferencii. Moskovskij gosudarstvennyj universitet imeni M.V. Lomonosova. Filologicheskij fakul'tet. M.: Izd-vo Mosk. un-ta. 1997].

Горбова Е. В. Заметки о видообразовании русского глагола и словоизменительной vs. словоклассифицирующей трактовке вида // Дмитренко С. Ю., Заика Н. М. (ред.). *Studia typologica octogenario Victori Khrakovskij Samuelis filio dedicata. Acta Linguistica Petropolitana.* Вып. X. Ч. 3. СПб.: Наука, 2014. С. 181–211. [Gorbova E. V. Notes on aspectual formation of the Russian verb and on inflectional vs. word-classifying interpretation of aspect. *Studia typologica octogenario Victori Khrakovskij Samuelis filio dedicata. Acta Linguistica Petropolitana.* No. X. Part 3. Dmitrenko S. Yu., Zaika N. M. (eds.). St. Petersburg: Nauka, 2014. Pp. 181–211.]

Горбова Е. В. Видообразование русского глагола: префиксация и/или суффиксация? // Вопросы языкознания. 2015а. № 1. С. 7–37. [Gorbova E. V. Aspectual formation of the Russian verb: Prefication or suffixation? *Voprosy Jazykoznanija*. 2015a. No. 1. Pp. 7–37.]

Горбова Е. В. Словоизменение или словообразование, или является ли русский вид «видом славянского типа»? // Китадзё М. (сост.). Аспектуальная семантическая зона: типология систем и сценарии диахронического развития. Сборник статей V Международной конференции Комиссии по аспектологии Международного комитета славистов. Киото, 13–15 ноября 2015. Киото: Tanaka Print. С. 62–72. 2015б. [Gorbova E. V. Inflection or derivation, or Is the Russian aspect «Slavicstyle aspect»? *Aspektual'naya semanticheskaya zona: tipologiya sistem i stsenarii diakhronicheskogo razvitiya*. Sbornik statei V Mezhdunarodnoi konferentsii Komissii po aspektologii Mezhdunarodnogo komiteta slavistov. Kioto, 13–15 noyabrya 2015. Kitajo M. (comp.). Kyoto: Tanaka Print. Pp. 62–72. 2015б]

Горбова Е. В. Русское видообразование: словоизменение, (слово)классификация или набор квазиграммем? (еще раз о болевых точках русской аспектологии) // Вопросы языкознания. 2017. № 1. С. 24–52. [Gorbova E. V. Aspectual formation of Russian verbs: inflection, derivation, or a set of quasigrammemes? («sore points» of Russian aspectology revisited). *Voprosy Jazykoznanija*. 2017. No. 1. Pp. 24–52.]

Зализняк А.А., Микаэлян И.Л. О некоторых дискуссионных моментах аспектологической концепции Лоры Янды. Вопросы языкознания. 2012, №6. С. 48-65. [Zaliznjak A.A., Mikajeljan I.L. O nekotoryh diskussionnyh momentah aspektologicheskoy koncepcii Lory Jandy. *Voprosy jazykoznanija*. 2012, №6. S. 48-65].

Зализняк Анна А., Шмелев А.Д., Микаэлян И.Л. Русская аспектология. В защиту видовой пары. М.: Языки славянских культур, 2015. [Zaliznjak Anna A., Shmelev A.D., Mikajeljan I.L. *Russkaja aspektologija. V zashhitu vidovoj pary*. M.: Jazyki slavjanskih kul'tur, 2015].

Кузнецова Ю.Л., Янда Л.А. Приставки в свете когнитивной лингвистики и типологии (отклик на статью А.А.Зализняк и И.Л. Микаэлян). Вопросы языкознания. 2013. №4, с. 87-96. [Kuznecova Ju.L., Janda L.A. Pristavki v svete kognitivnoj lingvistiki i tipologii (otklik na stat'ju A.A.Zaliznjak i I.L. Mikajeljan). *Voprosy jazykoznanija*. 2013. №4, s. 87-96].

Маслов Ю.С. Очерки по аспектологии. Л.: ЛГУ. 1984. [Maslov YU.S. *Oчерки po aspektologii*. L.: LGU. 1984].

Мельчук И. А. 1997. Курс общей морфологии. Т.1. М.: Прогресс, Языки русской культуры, Вена: WSA. [Mel'chuk I. A. 1997. *Kurs obshhej morfologii*. T.1. M.: Progress, Jazyki russkoj kul'tury, Vena: WSA].

Перцов Н.В. 1996. Грамматическое и обязательное в языке. Вопросы языкознания. №4. с. 39-61. [Percov N.V. 1996. Grammaticheskoe i objazatel'noe v jazyke. Voprosy jazykoznanija. №4. s. 39-61].

Петрухина Е.В. Аспектуальные категории глагола в русском языке в сопоставлении с чешским, словацким, польским и болгарским языками. М.: МГУ. 2000. [Petruhina E.V. Aspektual'nye kategorii glagola v rusском jazyke v sopostavlenii s cheshskim, slovackim, pol'skim i bolgarskim jazykami. М.: MGU. 2000].

Плунгян В.А. Общая морфология. Введение в проблематику. М.: УРСС, 2000. [Plungjan V.A. Obshhaja morfologija. Vvedenie v problematiku. Mju: URSS, 2000].

Русская грамматика: научные труды. В 2 томах / Н. Ю. Шведова (Гл. ред.). — М.: Институт русского языка им. В.В. Виноградова РАН, 2005. [Russkaja grammatika: nauchnye trudy. V 2 tomah / N. Ju. Shvedova (Gl. red.). — М.: Institut russkogo jazyka im. V.V. Vinogradova RAN, 2005].

Соловьев В., Бочкарев В., Янда Л. Аспектуальная система в русском языке: семантический вклад приставок и суффиксов // 15th Slavic Cognitive Linguistics Conference, Book of Abstracts, p. 75-76. 2017. [Solovyev V., Bochkarev V., Janda L. Aspektual'naja sistema v rusском jazyke: semanticheskij vklad pristavok i suffiksov // 15th Slavic Cognitive Linguistics Conference, Book of Abstracts, p. 75-76. 2017].

Храковский В. С. Категория вида в русском языке: болевые точки // Китадзё М. (сост.). Аспектуальная семантическая зона: типология систем и сценарии диахронического развития. Сборник статей V Международной конференции Комиссии по аспектологии Международного комитета славистов. Киото, 13–15 ноября 2015 г. Киото: Tanaka Print. С. 321–334, 2015. [Xrakovskij V. S. The category of aspect in Russian: Problematic issues. Aspektual'naya semanticheskaya zona: tipologiya sistem i stsenarii diakhronicheskogo razvitiya. Sbornik statei V Mezhdunarodnoi konferentsii Komissii po aspektologii Mezhdunarodnogo komiteta slavistov. Kioto, 13–15 noyabrya 2015 g. Kitajo M. (comp.). Kyoto: Tanaka Print. Pp. 321–334, 2015].

Храковский В. С. И опять о болевых точках категории вида // Вопросы языкознания. 2018. № 1. С. 105–118, 2018. [Xrakovskij V. S. Aspect's "sore points" revisited once more. Voprosy Jazykoznanija. 2018. No. 1. Pp. 105–118, 2018].

Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2017). Enriching word vectors with subword information. *TACL* 5, 135–146.

Dressler W.U. 1989. Prototypical differences between inflection and derivation. *ZPSK*. 42.1, pp. 3-10.

Grave E., Bojanowski P., Gupta P., Joulin A., Mikolov T. Learning Word Vectors for 157 Languages. Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018). <https://www.aclweb.org/anthology/L18-1550>

Janda L. & Lyashevskaya O. (2011): Aspectual Pairs in the Russian National Corpus, *Scando-Slavica*, 57:2, 201-215.

Janda L., et al. Why Russian aspectual prefixes aren't empty. Bloomington: Slavica. 2013.

Filip H. Aspect, Eventuality Types, and Nominal Reference. London. 1999.

Kutuzov A. and Andreev I. Texts in, meaning out: neural language models in semantic similarity task for Russian. Proceedings of the Dialog 2015 Conference, Moscow, Russia (2015).

J.-B. Michel, Y. K. Shen, A. P. Aiden, A. Veres, M. K. Gray, J. P. Pickett, D. Hoiberg, D. Clancy, P. Norvig, J. Orwant, S. Pinker, M. A. Nowak and E. L. Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.

Mikolov, T., Chen, K., Corrado, G. S., and Dean, J. (2013). Efficient estimation of word representations in vector space. ICLR 2013.

Pennington, J., Socher, R., and Manning, C. (2014). Glove: Global vectors for word representation. In Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pages 1532–1543.

Peters M., Neumann M., Iyyer M, Gardner M., Clark C., Lee K., and Zettlemoyer L. 2018. Deep contextualized word representations. arXiv preprint arXiv:1802.05365.

Plank F. 1991. Inflection and derivation (EUROTYP working papers, 10).

Richey S., Taylor J. Google Books Ngrams and political science: Two validity tests for a novel data source, *PS: Political Science & Politics* 53 (2020) 72-77. doi:10.1017/S1049096519001318.

Sahlgren M. The Distributional Hypothesis. From context to meaning. *Distributional models of the lexicon in linguistics and cognitive science* (Special issue of the Italian Journal of Linguistics), *Rivista di Linguistica*. 2008. Vol. 20, no. 1. P. 33—53.

Shavrina, T., Shapovalova, O.: To the methodology of corpus construction for machine learning: “taiga” syntax tree corpus and parser. In: Proceedings of the International Conference «Corpus Linguistics–2017». St. Petersburg. pp. 78–84 (2017).

Singh R., Ford A. Rlexion, derivation et Panini. In: *Studies in the History of Libguistics*, v.20 (ed. K. Koerner). Amsterdam: John Benjamins. Pp. 323-332.

Sivakumar S., Videla L.S., Kumar T.R., Nagaraj J., Itnal Sh., and Haritha D. Review on word2vec word embedding neural net. In *2020 International Conference on Smart Electronics and Communication (ICOSEC)*, pages 282–290. IEEE, 2020.

Solovyev V.D., Bochkarev V.V., Akhtyamova S.S., Google Books Ngram: Problems of representativeness and data reliability, in: A. Elizarov, B. Novikov, S. Stupnikov (Eds.), *Data*

Analytics and Management in Data Intensive Domains. Communications in Computer and Information Science, volume 1223, Springer, Cham, 2019, pp. 147–162. doi:10.1007/978-3-030-51913-1_10.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. arXiv preprint arXiv:1906.08237.

Bionotes:

Valery D. SOLOVYEV is Doctor habil. of Physics and mathematics, Professor, Chief Researcher of the “Text analytics” Research Laboratory at Kazan (Volga Region) Federal University. His research interests include cognitive sciences, computer linguistics, text complexity.

Contact information:

Kazan (Volga Region) Federal University, 18 Kremlevskaya St., Kazan, Russia, 420008

e-mail: maki.solovyev@mail.ru*

tel. work: (843)221-33-22, (843)221-33-09

tel. mobile: +79196910489*

ORCID: 0000-0003-4692-2564

Vladimir V. BOCHKAREV is a research scientist of the “Linguistics and Artificial Intelligence” Research Laboratory at Kazan (Volga Region) Federal University. His research interests include computer linguistics, data analysis, signal processing, statistics, mathematical modelling.

Contact information:

Kazan (Volga Region) Federal University, 18 Kremlevskaya St., Kazan, Russia, 420008

e-mail: vbochkarev@mail.ru

tel. mobile: +79033419696

ORCID: 0000-0001-8792-1491

Venera Rustamovna Bayrasheva – kandidat of Physics and mathematics sciences, docent of the chair of theoretical cybernetics at Kazan (Volga Region) Federal University. His research interests include cognitive sciences, computer linguistics.

Contact information:

Kazan (Volga Region) Federal University, 18 Kremlevskaya St., Kazan, Russia, 420008

e-mail: vbayrasheva@gmail.com

tel. mobile: +79050260285

ORCID: 0000-0002-1728-034X

Сведения об авторах:

Валерий Дмитриевич СОЛОВЬЕВ – доктор физико-математических наук, профессор, главный научный сотрудник НИЛ «Текстовая аналитика» Казанского (Приволжского) федерального университета. В сферу его научных интересов входят: когнитивная наука, компьютерная лингвистика, сложность текстов.

Контактная информация:

Казанский (Приволжский) федеральный университет

Россия, 420008, г. Казань, ул. Кремлевская, 18

e-mail: maki.solovyev@mail.ru*

тел. рабочий: (843)221-33-22, (843)221-33-09

тел. мобильный: +79196910489*

ORCID: 0000-0003-4692-2564

Владимир Владимирович Бочкарев – научный сотрудник НИЛ “Лингвистика и искусственный интеллект” Казанского (Приволжского) федерального университета. В сферу его научных интересов входят: компьютерная лингвистика, анализ данных, обработка сигналов, статистика, математическое моделирование.

Венера Рустамовна Байрашева – кандидат физико-математических наук, доцент кафедры теоретической кибернетики Казанского (Приволжского) федерального университета. В сферу его научных интересов входят: когнитивная наука, компьютерная лингвистика.

Казанский (Приволжский) федеральный университет

Россия, 420008, г. Казань, ул. Кремлевская, 18

e-mail: vbayrasheva@gmail.com

тел. мобильный: +79050260285

ORCID: 0000-0002-1728-034X

Aspectual Pairs: Prefix vs. Suffix Mode of Formation

V.D. Solovyev, V.V. Bochkarev, V.R. Bayrasheva

Kazan Federal University

Abstract. In the literature, there is no common point of view on the question of whether verbs in aspectual pairs are in inflectional or derivational relations. At the same time, in a number of publications, the prefix and suffix methods of forming aspectual pairs are contrasted in this respect. The publications pointed out the need to develop new quantitative approaches in this issue on the text corpus. We propose two new approaches that compare the quantitative characteristics of aspectual pairs of both types.

One approach is based on the Google Books Ngram corpus and analyzes the dynamics of the frequency of the use of words in pairs. The aspectual pairs from the databases created by L. Yanda and O. Lyashevskaya are considered. For a numerical assessment of the degree of proximity of the frequency graphs, the Pearson correlation coefficients were used. The second approach introduces a numerical characteristic of the semantic proximity of verbs in pairs using modern computer methods.

Semantic proximity of verbs is calculated as a standard cosine measure between vectors representing the compatibility of the considered verbs in the corpus. Several computer models and text corpora are considered. Both approaches did not reveal significant numerical differences between prefix and suffix pairing. This is in good agreement with the results of an early study by L. Yanda and O. Lyashevskaya.

Key words: Russian language, aspect, aspectual pairs, inflection, word formation, quantitative analysis, corpora