

Компьютерная лингвистика и интеллектуальные технологии:
по материалам международной конференции «Диалог 2016»

Москва, 1–4 июня 2016

СРАВНЕНИЕ КОРПУСНОГО И ЭКСПЕРИМЕНТАЛЬНОГО МЕТОДА НА ПРИМЕРЕ ИССЛЕДОВАНИЯ СИНТАКСИЧЕСКИХ СВОЙСТВ ЭНКЛИТИКИ ЖЕ

Валова Е. А. (dunya_v@yahoo.com)

НИУ ВШЭ, Москва, Россия

Слюсарь Н. А. (slioussar@gmail.com)

НИУ ВШЭ, Москва; СПбГУ, Санкт-Петербург, Россия

Ключевые слова: русский язык, фразовые энклитики, закон
Ваккернагеля, корпусный метод, экспериментальный метод

COMPARING CORPUS AND EXPERIMENTAL APPROACHES: A STUDY OF SYNTACTIC PROPERTIES OF THE RUSSIAN ENCLITIC ЖЕ

Valova E. A. (dunya_v@yahoo.com)

HSE, Moscow, Russia

Slioussar N. A. (slioussar@gmail.com)

HSE, Moscow;

St. Petersburg State University, St. Petersburg, Russia

Corpus and experimental approaches in linguistics are often seen as incompatible, and there are very few studies of grammatical phenomena that rely on both of them, without one or the other being subsidiary. In this paper,

we would like to show that they are complimentary and can be fruitfully combined on the example of Russian phrasal enclitic *že*. We analyze various factors influencing its position in the sentence, in particular, whether it obeys Wackernagel's law, which applied to phrasal enclitics in Old Russian. Data from the National Russian corpus show that *že* appears in the strict Wackernagel's position in the absolute majority of cases in the main subcorpus and the newspaper subcorpus, while the subcorpus of spoken Russian exhibits more variation. Corpus data allow tracing diachronic tendencies and identifying several factors (primarily, the semantics of *že*). Experimental data let us estimate the role of these and other factors on a carefully balanced set of examples. Apart from syntactic and semantic factors, the age and educational level of participants was demonstrated to influence the results.

Keywords: Russian, phrasal enclitics, Wackernagel's law, corpus approach, experimental approach

1. Введение

В середине XX в. в лингвистике началось несколько важных дискуссий, которые в том или ином виде продолжают до сих пор. Одна из них касалась того, какие источники данных должны использоваться в исследовании. Ее инициировал Н. Хомский, который обрушился с критикой на корпусный подход и предложил использовать вместо него интроспективные суждения о грамматической правильности тех или иных примеров (*grammaticality judgments*) ([Chomsky 1957, 1962, 1965] и далее). Эта идея сыграла ключевую роль в развитии экспериментального подхода: многие авторы, опасаясь полагаться исключительно на свою интуицию, стали опрашивать большие группы носителей языка.¹

С тех пор и корпусные, и экспериментальные методы в лингвистике претерпели множество изменений и значительно усовершенствовались. Тем не менее, антагонизм между их последователями во многом сохранился. Продолжают появляться работы, предлагающие новые аргументы против одного из этих подходов, а также отвечающие на высказанную ранее критику (см. например, [McEnery, Wilson 2001; Riemer 2009; Sampson 1992; Schütze 1996; Tremblay 2005]).

Мы полагаем, что, во всяком случае, на данном этапе развития лингвистики эта дискуссия неплодотворна, потому что никто не заставляет исследователя ограничиваться одним подходом: их можно рассматривать не как взаимоисключающие, а как взаимодополняющие. Однако для того, чтобы убедиться в этом, нужны конкретные примеры. Между тем, подавляющее большинство исследований либо полагается только на один из этих источников данных, либо, используя оба, рассматривает один из них сугубо как вспомогательный (см., например, обзор в [Guilquin, Gries 2009]). В данной работе мы хотим

¹ Один из рецензентов отметил, что терминологически правильнее говорить о противопоставлении корпусного метода и метода опроса, но мы будем придерживаться устоявшейся терминологии.

внести свою лепту в решение этой проблемы, продемонстрировав, как корпусные и экспериментальные данные могут дополнить друг друга, на примере исследования синтаксических свойств фразовой энклитики *же*.

Клитиками называют слова, не имеющие собственного ударения и подчиняющиеся акцентуации предшествующего или последующего слова (например, энклитика *бы* в *сказал бы*, проклитика *на* в *на полке*). Некоторые энклитики являются фразовыми, то есть относятся не к конкретному слову, а к предложению в целом (например, *Я же говорил об этом! Говорил ли я об этом?*). В древнерусском языке, как и в праиндоевропейском, расположение таких энклитик подчинялось закону Ваккернагеля [Wackernagel 1892]: они примыкали к первому фонетическому (полноударному) слову фразы. Однако затем его действие ослабло (ср. [Зализняк 2008]).

Мы провели корпусное и экспериментальное исследование, чтобы оценить, насколько оно ослабло, на примере энклитики *же*, которая лучше всего сохранила свои ваккернагелевские свойства [Зализняк 2008]. Подробному описанию различных результатов этих исследований будет посвящена другая работа². Здесь же мы хотели бы, осветив вкратце основные моменты, сфокусироваться на вопросах, связанных с совместным использованием корпусных и экспериментальных данных.

Ослабление закона Ваккернагеля во многих языках выражается в том, что он начинает действовать в так называемой нестрогой формулировке, согласно которой фразовые энклитики могут располагаться как после первого полноударного слова, так и после первой синтаксической составляющей (ср., например, [Spencer, Luís 2012]). Подобную вариативность можно наблюдать и для фразового *же*, и в данной работе она была исследована на материале именных групп. Возможные варианты размещения энклитики мы будем называть соответственно первой позицией, или строгой позицией Ваккернагеля, как в примере (1), и второй, или нестрогой ваккернагелевской позицией, как в примере (2).

- (1) *Но глупый плачет от бедства, а умный ищет средства. Татьяна же Акимовна была женщина умная.* [Н.Г. Помяловский. Махилев (1855)]
- (2) *ФИФА унифицировала свой регламент по переходам еще в прошлом году. У национальных федераций же есть время на его адаптацию.* [Мутко: В российском футболе появится статус свободного агента // Советский спорт, 2006.10.11]

2. Корпусное исследование

Корпусное исследование было проведено на материале Национального корпуса русского языка (НКРЯ, www.ruscorpora.ru). Так как примеров с фразовым *же*

² Кроме того, предварительные результаты корпусного исследования были представлены в [Валова 2014а, б].

в НКРЯ очень много, мы остановились на предложениях, где носителями энклитики являются два типа именных групп: комплекс имен собственных, как в примере (1) выше, и сочетание прилагательного и существительного, как в примере (2). Набор факторов, которые могут действовать в первой группе примеров, ограничен (в частности, не идет речи о большей или меньшей семантической связности двух слов внутри именной группы), их число невелико. Мы ожидали, что в них особенно ярко проявится тенденция к ослаблению закона Ваккернагеля. Вторая группа примеров намного более разнородна и обширна.

После отсева многочисленных примеров с локальным *же*³, в нашу итоговую выборку вошло около пятнадцати тысяч предложений. Для примеров из основного подкорпуса НКРЯ мы проследили, как менялось число предложений с *же* во второй позиции с течением времени. С середины 19 в. наблюдается рост их доли в обеих рассматриваемых группах (до этого подобные примеры единичны), применение логистической регрессии показывает, что эта тенденция статистически значима ($p < 0,01$). Однако, как показывают обобщенные данные в Таблице 1, если в случае с именами собственными картина изменяется заметно, то в случае с прилагательными и существительными число таких примеров остается настолько незначительным, что они совершенно незаметны на общем фоне.

Таблица 1. Частотность примеров с *же* в первой и второй позициях: основной корпус

Временной интервал	Имена собств.		Прил. и сущ.		Общее кол-во словоупотреблений
	поз. 1	поз. 2	поз. 1	поз. 2	
1851–1900	155	4 (2,5%)	1690	5 (0,3%)	41 107 495
1901–1950	131	17 (11,5%)	1964	17 (0,9%)	60 237 302
1951–2000	103	16 (13,5%)	1632	22 (1,3%)	73 970 923
2001–н.вр.	37	19 (33,9%)	1551	35 (2,2%)	283 906 971

Более внушительную долю примеров с *же* во второй позиции можно найти, сравнивая данные основного, газетного и устного подкорпусов НКРЯ. Если взять все вошедшие в нашу выборку предложения с прилагательными и существительными, мы увидим, что в основном подкорпусе таких примеров 85 (1,1%), в газетном — 166 (3,2%), а в устном — 25 (25,5%) (в устном и газетном подкорпусе все примеры относятся к самому позднему периоду, в основном самые ранние — к началу 18 в.). Все три подкорпуса значимо отличаются друг от друга ($p < 0,01$ согласно критерию хи-квадрат с поправкой Бонферрони).

Кроме того, корпусное исследование позволило выявить ряд факторов, которые влияют на позицию энклитики. Прежде всего, это ее значение. Рассмотрению различных значений *же* посвящена обширная литература (например,

³ В этих примерах энклитика относится не к предложению в целом, а к предшествующему слову (например, *такой же ответ, на следующий же день*).

[Киселев 1976; Paducheva 1987; Бонно, Кодзасов 1998]). Опираясь на нее, в нашей работе мы выделяли следующие значения: противительное, усилительное, пояснительное и присоединительное.

Во всех группах примеров во всех подкорпусах более высокий процент предложений с нестрогой ваккернагелевской позицией наблюдается для пояснительного и усилительного *же*, самый низкий — для противительного. Однако при этом примеров с противительным *же* на порядок больше, чем всех прочих. Примеров с присоединительным *же* слишком мало, чтобы делать какие-либо выводы. Также мы отметили, что в предложениях с отрицанием *же* в абсолютном большинстве случаев занимает строгую ваккернагелевскую позицию, а в предложениях, где прилагательное и существительное образуют устойчивое словосочетание (коллокацию), *же* чаще оказывается во второй позиции. Были описаны и некоторые другие тенденции. В целом же мы пришли к выводу, что, хотя случаи нарушения строгого закона Ваккернагеля в рассмотренных материалах не единичны, ослабление его действия далеко не так значительно, как можно было бы ожидать.

3. Экспериментальное исследование

Мы провели два эксперимента с использованием разных методик. В первом респонденты должны были разместить *же* в предложениях, используя одну из двух предложенных позиций. Во втором респондентам было предложено оценить предложения по шкале от 1 до 5, где 1 — «предложение звучит плохо, я бы так никогда не сказал», а 5 — «предложение звучит хорошо, я сам мог бы так сказать»⁴.

Основной минус первого метода заключается в том, что, если испытуемые считают приемлемыми (или, наоборот, неприемлемыми) оба варианта, мы не получим об этом информации — только информацию о том, какой из двух вариантов кажется (несколько) лучше. Зато он дает сведения о предпочтениях в максимально простом и удобном для анализа виде. Основным минусом второго метода можно считать то, что нельзя предъявить одному испытуемому оба варианта предложения (с *же* в первой и во второй позициях), так как повторять стимулы в эксперименте не принято. Поэтому во втором эксперименте было два экспериментальных листа (половина испытуемых видела то или иное предложение с *же* в первой позиции, половина — с *же* во второй). Анкетирование проводилось с использованием форм Google (www.google.com/intl/ru_ru/forms/about/).

В первом эксперименте приняло участие 214 человек в возрасте от 14 до 80 лет, из них 53 мужчины и 161 женщина. Второму прошли 181 человек в возрасте от 16 до 80 лет, из них 42 мужчины и 139 женщин. Уровень образования респондентов варьировал от незаконченного среднего до наличия ученой степени.

Чтобы можно было сопоставить результаты корпусного и экспериментального исследования, носителем энклитики в стимульных предложениях

⁴ В настоящее время в экспериментах обычно используются не бинарные оценки («правильно/неправильно»), а шкалы, которые позволяют отслеживать более тонкие различия.

была сделана именная группа с прилагательным и существительным. Ниже перечислены различные группы стимулов (их сокращенные названия, описание и примеры)⁵. Возможные позиции энклитики, которые рассматривались в данной работе, обозначены скобками, для термина *именная группа* используется сокращение *ИГ*.

- Прот. *же* (не ИГ): *же* в противительном значении, противопоставлены клаузы.

Я три года деньги коплю, деревянный () домик () всё дорожает и дорожает.

- Прот. *же* (ИГ): *же* в противительном значении, противопоставлены ИГ.

Ты как в воду глядел. Кирпичные многоэтажки никого не впечатлили, деревянный () домик () вызвал всеобщий энтузиазм.

- Прот. *же* (прил.): *же* в противительном значении, противопоставлены прилагательные внутри ИГ.

Ты как в воду глядел. Кирпичные многоэтажки никого не впечатлили, деревянный () домик () вызвал всеобщий энтузиазм.

- Поясн. *же*: *же* в пояснительном значении.

Сфотографируй это! Деревянный () домик () такой красивый.

- Поясн. *же* (прот. ИГ): *же* в пояснительном значении, при этом в предложении есть противопоставление ИГ.

Всё вышло, как ты хотел. Деревянный () домик () останется, а бетонные громадины снесут.

- Поясн. *же* (отр.): *же* в пояснительном значении, в предложении есть отрицание.

Чему ты радуешься? Не деревянный () домик () тут будет!

- Усил. *же* (констр.): *же* в усилительном значении, конструкция с вынесением ремы.

⁵ Кроме того, была создана группа примеров, в которых именной группой — носителем клитики является устойчивое выражение, или коллокация, а также аналогичные сочетания, демонстрирующие меньшую связанность, однако мы не будем рассматривать их в данной работе.

Красивый () домик () купили наши соседи!

- Присоед. *же*: *же* в присоединительном значении.

Чтобы что-то построить, надо было снести один деревянный домик или купить еще земли. Деревянный () домик () охранялся как памятник, и снести его было сложно.

Таким образом, у нас было восемь групп стимулов, в каждой из которых было по три примера (то есть всего 24 предложения). В этих предложениях *же* появлялось в разных значениях. Кроме того, внутри группы с противительным *же* рассматривались примеры с разным фокусом контраста (в которых противопоставлены именные группы — носители клитики, входящие в них прилагательные или клаузы в целом), а в группе примеров с пояснительным *же* были примеры с отрицанием и с противопоставлением именных групп (мы хотели оценить, какую роль играет само наличие в предложении противопоставления, а какую — значение *же*). Пример с усилительным *же* содержал особую конструкцию с вынесением ремы в начало предложения. Также в эксперимент были включены 64 филлера: предложения с частицами *бы* и *не*, предполагающие выполнение тех же заданий, что и стимулы, но на другом материале (например, в предложение «Я всю жизнь ждала человека, который () понял () меня» надо было вставить частицу *бы* или оценить его с частицей в разных позициях).

Результаты экспериментов обобщены в Таблице 2. Первая колонка отражает процент выбора нестройгой ваккернагелевской позиции в первом эксперименте. Средняя оценка первой и второй позиций, а также их разница приведены на основании второго эксперимента.

Таблица 2. Результаты двух экспериментов

Тип стимула	% выбора поз. 2	Оценка поз. 1	Оценка поз. 2	Разница
Прот. <i>же</i> (не ИГ)	58,9%	3,6	4,0	-0,4
Прот. <i>же</i> (ИГ)	33,2%	4,4	3,9	0,5
Прот. <i>же</i> (прил.)	6,3%	4,7	3,3	1,4
Поясн. <i>же</i>	85,3%	2,4	3,9	-1,5
Поясн. <i>же</i> (прот. ИГ)	51,9%	2,8	3,1	-0,4
Поясн. <i>же</i> (отр.)	14,3%	4,4	3,5	0,9
Усил. <i>же</i> (констр.)	3,7%	4,5	2,6	1,9
Присоед. <i>же</i>	57,2%	3,9	4,2	-0,3

Сперва рассмотрим результаты первого эксперимента. Используя метод логистической регрессии, мы оценили вероятность ответа «1» или «2» в зависимости от различных характеристик предложений. Три типа предложений с противительным *же* значительно отличаются друг от друга ($p < 0,01$), что объясняется различным фокусом контраста. Значимо отличаются друг от друга

и три типа предложений с пояснительным *же* ($p < 0,01$): процент выбора второй позиции снижается при появлении противопоставления и особенно при наличии в предложении отрицания. Предложения с противительным и пояснительным *же*, включающие противопоставление именных групп, значимо отличаются друг от друга ($p < 0,01$).

В целом мы полагаем, что роль значения энклитики правильнее оценивать на материале корпусных данных, поэтому в эксперименте стремились не сравнить друг с другом все возможные значения, а скорее удостовериться, что этот фактор играет роль, и посмотреть, как он будет взаимодействовать с другими. Однако в корпусе было слишком мало предложений с присоединительным *же*, поэтому мы сравнили примеры с пояснительным, присоединительным и противительным *же* на материале экспериментальных данных (из рассмотрения были исключены предложения с отрицанием и противопоставлением прилагательных, так как в них действуют мощные независимые факторы). В результате было выявлено то же распределение, которое наблюдается в корпусе, и различия между группами оказались статистически значимыми ($p < 0,01$).

На примере усилительного *же* мы планировали оценить не фактор значения (таких примеров достаточно в корпусе), а роль особой конструкции, для которой характерен определенный интонационный контур и определенная интерпретация (именная группа является ремой, вынесенной в начало предложения). Результаты показали, что в таких предложениях процент выбора нестрогой вакернагелевской позиции был самым низким (в частности, эта группа значимо отличается от группы, содержащей отрицание, $p < 0,01$).

О результатах второго эксперимента мы скажем очень кратко. Прежде всего, заметим, что результаты двух экспериментов согласуются друг с другом: если какой-то вариант чаще выбирают в первой части, у него всегда выше средний балл во второй. Учитывая это, мы не будем повторять для второй части приведенные выше сравнения, а остановимся на других моментах. Прежде всего, бросается в глаза, что, даже при наличии явных предпочтений, очень часто оба варианта предложения получают достаточно высокие оценки. Это верно для всех примеров с противительным *же* (несколько в меньшей степени при противопоставлении прилагательных), для присоединительного *же* и, что было для нас неожиданным, даже для предложений с пояснительным *же*, в которых есть отрицание. Как нам кажется, такую картину можно считать дополнительным свидетельством ослабления закона Вакернагеля в его строгой формулировке.

Относительно низкие оценки получили предложения с пояснительным *же* в первой позиции и с усилительным *же* в рамках особой конструкции во второй — судя по всему, эти два варианта действительно режут респондентам слух. Кроме того, невысокие оценки получили оба варианта предложений с пояснительным *же*, включавшие противопоставление именных групп. Видимо, они в целом не понравились испытуемым (например, потому, что в них лучше смотрелось бы *ведь*, а не *же*), поэтому данные по этим предложениям надо рассматривать с осторожностью.

Кроме того, на материале первого эксперимента мы изучили, влияют ли на ответы различные социолингвистические характеристики респондентов.

Мы выяснили, что уровень образования и особенно возраст являются значимыми факторами ($p = 0,04$ и $p < 0,01$ соответственно), а пол не является ($p = 0,80$). Более молодые и/или менее образованные респонденты чаще выбирали для энклитики вторую позицию.

4. Заключение

В данной работе мы представили результаты корпусного и экспериментального исследования синтаксических свойств энклитики *же*. Прежде всего, отметим, что многие тенденции проявились и в корпусных, и в экспериментальных данных. Например, в обоих исследованиях реже всего оказывается в нестрогой ваккернагелевской позиции противительное *же*, а чаще всего — пояснительное, в обоих проявилась роль наличия в предложении отрицания. Это можно считать лишним доказательством того, что эти два метода не противоречат друг другу.

В определенных случаях эти методы дополняют друг друга, что представляется особенно ценным. Так, корпусное исследование лучше подходит для того, чтобы судить о роли значения энклитики для ее расположения в предложении — в этом случае важно опираться на большую выборку разнообразных примеров, порожденных в естественных условиях. А используя экспериментальный метод, легче на однородном материале оценить влияние различных формальных факторов: наличия отрицания, противопоставления, особой конструкции, изменение фокуса контраста. Корпус дает представление о вариации в текстах разных эпох и жанров, эксперимент — о вариации, связанной с социолингвистическими факторами (возраст, пол, уровень образования респондентов).

Наконец, нельзя не заметить и существенные различия между результатами корпусного и экспериментального исследования. В основном подкорпусе Национального корпуса русского языка примеры с прилагательными и существительными, где *же* находится во второй позиции, практически незаметны на общем фоне, в то время как в эксперименте таких ответов чуть меньше половины. Прежде всего, отметим, что расхождения оказываются далеко не такими разительными, если сравнить результаты эксперимента с данными устного подкорпуса. Таким образом, отчасти мы имеем дело не с различиями между методами, а с различиями между жанрами.

Определенные расхождения, безусловно, связаны и с различиями между методами. Однако, как нам кажется, неверно будет говорить, что один из методов подводит нас ближе к истинной картине, чем другой: скорее они позволяют посмотреть на нее с разных сторон. Например, мы показали в эксперименте, что есть конструкции, прежде всего, примеры с пояснительным *же*, в которых подавляющее большинство опрошенных носителей языка уверенно выбирает для энклитики нестрогую ваккернагелевскую позицию. Однако нам нужен корпус для того, чтобы понять, что *же* в пояснительном значении используется на порядок реже, чем в противительном. Соответственно, роль таких конструкций может быть велика (всякая глобальная перестройка системы начинается

с ограниченного класса примеров и затем захватывает всё новые и новые), но, во всяком случае, на данном этапе развития русского языка, они мало заметны в текстах в силу общей низкой частотности. В связи с этим мы полагаем, что наиболее полное представление о том или ином языковом явлении можно получить только путем использования обоих методов.

Часть представленных в данной статье исследований была выполнена при поддержке гранта РФФ № 16-18-02071.

Литература

1. Бонно К., Кодзасов С. В. (1998), Семантическое варьирование дискурсивных слов и его влияние на линеаризацию и интонирование (на примере частиц «же» и «ведь»), Дискурсивные слова русского языка: опыт контекстно-семантического описания, Киселева К., Пайар Д. (ред.), Метатекст, Москва, с. 382–446.
2. Валова Е. А. (2014), Синтаксические свойства энклитической частицы *же* в диахроническом аспекте: корпусное исследование. // Научно-техническая информация. Серия 2. Информационные процессы и системы, № 10, с 31–36.
3. Валова Е. А. (2014). Синтаксические свойства русской энклитической частицы *же*. // Вестник Православного Свято-Тихоновского гуманитарного университета. Серия 3. Филология. №III:4 (39), с. 16–33.
4. Киселев И. А. (1976), Частицы в современных восточнославянских языках, БГУ, Минск.
5. Зализняк А. А. (2008), Древнерусские энклитики, Языки славянских культур, Москва.
6. Chomsky N. (1957), Syntactic Structures, Mouton and Co., The Hague.
7. Chomsky N. (1962), Explanatory Models in Linguistics, Logic, Methodology, and Philosophy of Science, Stanford University Press, Stanford, Calif., pp. 528–550.
8. Chomsky N. (1965), Aspects of the Theory of Syntax, The M. I. T. Press, Cambridge, Mass.
9. Gilquin G., Gries S. Th. (2009), Corpora and experimental methods: a state-of-the-art review, Corpus Linguistics and Linguistic Theory, Vol. 5(1), pp. 1–26.
10. McEnery T., Wilson A. (2001), Corpus Linguistics: An Introduction. University Press, Edinburgh.
11. Paducheva E. V. (1987), Particle *že*: semantics, syntax, prosody [La particule *že*: sémantique, syntaxe et prosodie], Narrative particles in modern Russian [Les particules énonciatives en Russe contemporain], Vol. 3, pp. 11–44.
12. Riemer N. (2009), Grammaticality as evidence and as prediction in a Galilean linguistics, Language Sciences, Vol. 31, pp. 612–633.
13. Sampson G. (1992), Probabilistic parsing, Directions in Corpus Linguistics, Mouton de Gruyter, Berlin, pp. 425–447
14. Schütze C. T. (1996), The empirical base of linguistics: Grammaticality judgments and linguistics methodology, The University of Chicago, Chicago.

15. *Spencer A., Luis Ana R.* (2012), *Clitics: an introduction*, Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge.
16. *Tremblay A.* (2005), Theoretical and methodological perspectives on the use of grammaticality judgment tasks in linguistic theory, *Second Language Studies*, Vol. 24 (1), pp. 129–167.
17. *Wackernagel J.* (1892), On a law about Indo-European word order [Über ein Gesetz der indogermanischen Wortstellung], *Indo-European Studies [Indogermanische Forschungen]*, Vol. 1, pp. 333–436.

References

1. *Bonno K., Kodzasov S. V.* (1998), Semantic variation of discourse words and its effect on the linearization and intonation (on the example of particles ‘že’ and ‘ved’’) [Semanticheskoe var’irovanie diskursivnykh slov i ego vliyanie na linearizaciju i intonirovanie (na primere chastic «zhe» i «ved’»)], The discourse words of the Russian language: the experience of context-semantic description [Diskursivnye slova russkogo jazyka: opyt kontekstno-semanticheskogo opisanija], *Metatekst*, Moscow, pp. 382–446.
2. *Chomsky N.* (1957), *Syntactic Structures*, Mouton and Co., The Hague.
3. *Chomsky N.* (1962), *Explanatory Models in Linguistics, Logic, Methodology, and Philosophy of Science*, Stanford University Press, Stanford, Calif., pp. 528–550.
4. *Chomsky N.* (1965), *Aspects of the Theory of Syntax*, The M. I. T. Press, Cambridge, Mass.
5. *Gilquin G., Gries S. Th.* (2009), Corpora and experimental methods: a state-of-the-art review, *Corpus Linguistics and Linguistic Theory*, Vol. 5(1), pp. 1–26.
6. *Kiselev I. A.* (1976), *Particles in modern East Slavic languages [Chasticy v sovremennykh vostochnoslavjanskikh jazykakh]*, BGU, Minsk.
7. *McEnery T., Wilson A.* (2001), *Corpus Linguistics: An Introduction*. University Press, Edinburgh.
8. *Paducheva E. V.* (1987), Particle že: semantics, syntax, prosody [La particule že: sémantique, syntaxe et prosodie], *Narrative particles in modern Russian [Les particules énonciatives en Russe contemporain]*, Vol. 3, pp. 11–44.
9. *Riemer N.* (2009), Grammaticality as evidence and as prediction in a Galilean linguistics, *Language Sciences*, Vol.31, pp. 612–633.
10. *Sampson G.* (1992), Probabilistic parsing, *Directions in Corpus Linguistics*, Mouton de Gruyter, Berlin, pp. 425–447
11. *Schütze C. T.* (1996), *The empirical base of linguistics: Grammaticality judgments and linguistics methodology*, The University of Chicago, Chicago.
12. *Spencer A., Luis Ana R.* (2012), *Clitics: an introduction*, Cambridge Textbooks in Linguistics, Cambridge University Press, Cambridge.
13. *Tremblay A.* (2005), Theoretical and methodological perspectives on the use of grammaticality judgment tasks in linguistic theory, *Second Language Studies*, Vol. 24 (1), pp. 129–167.

14. Valova E. A. (2014), Syntactic properties of Russian enclitic particle 'zhe' in the diachronic aspect: a corpus-based study [Sintaksicheskie svojstva enkliticheskoj chasticy zhe v diahronicheskom aspekte: korpusnoe issledovanie], Scientific and technical information. Series 2. Information processes and systems [Nauchno-tehnicheskaja informacija. Serija 2. Informacionnye processy i sistemy.], Vol. 10, pp. 31–36.
15. Valova E. A. (2014), Syntactic properties of Russian enclitic particle 'že' [Sintaksicheskie svojstva russoj jenkliticheskoj chasticy zhe], Bulletin of an orthodox St. Tikhon humanitarian University. Series 3. Philology. [Vestnik Pravoslavnogo Sv-jato-Tihonovskogo gumanitarnogo universiteta. Serija 3. Filologija.], Vol. III:4 (39), pp. 16–33.
16. Wackernagel J. (1892), On a law about Indo-European word order [Über ein Gesetz der indogermanischen Wortstellung], Indo-European Studies [Indogermanische Forschungen], Vol. 1, pp. 333–436.
17. Zaliznyak A. A. (2008), Old Russian enclitics [Drevnerusskie enklitiki], Jazyki slavjanskih kul'tur, Moscow.