



Литвинова Т. А. Фактор гендера в ассоциативных связях слов : данные словаря и дистрибутивно-семантической модели / Т. А. Литвинова, Е. С. Котлярова, В. А. Заварзина // Научный диалог. — 2022. — Т. 11. — № 5. — С. 136—156. — DOI: 10.24224/2227-1295-2022-11-5-136-156.

Litvinova, T. A., Kotlyarova, E. S., Zavarzina, V. A. (2022). Gender Factor in Associative Links of Words: Dictionary and Distributive-Semantic Model Data. *Nauchnyi dialog*, 11(5): 136-156. DOI: 10.24224/2227-1295-2022-11-5-136-156. (In Russ.).



Журнал включен в Перечень ВАК

DOI: 10.24224/2227-1295-2022-11-5-136-156

Фактор гендера в ассоциативных связях слов: данные словаря и дистрибутивно- семантической модели

Литвинова Татьяна Александровна
orcid.org/0000-0002-6019-3700
кандидат филологических наук,
научный сотрудник Регионального
центра русского языка,
ведущий научный сотрудник
научно-исследовательской лаборатории
компьютерной семасиологии
centr_rus_yaz@mail.ru

Котлярова Елена Сергеевна
orcid.org/0000-0002-4162-5002
лаборант научно-исследовательской
лаборатории компьютерной семасиологии
kray-voronezh@mail.ru

Заварзина Виктория Александровна
orcid.org/0000-0001-6484-4606
аспирант, лаборант научно-
исследовательской лаборатории
компьютерной семасиологии
zva0604@yandex.ru

Воронежский государственный
педагогический университет
(Воронеж, Россия)

Благодарности:

Исследование выполнено при поддержке гранта
Российского научного фонда № 21-78-10148
«Моделирование значения слова
в индивидуальном языковом сознании
на основе дистрибутивной семантики»

Gender Factor in Associative Links of Words: Dictionary and Distributive-Semantic Model Data

Tatyana A. Litvinova
orcid.org/0000-0002-6019-3700
PhD in Philology, Researcher,
Regional Russian Language Center;
Leading Researcher, Research
Laboratory of Computer Semasiology
centr_rus_yaz@mail.ru

Elena S. Kotlyarova
orcid.org/0000-0002-4162-5002
Laboratory Assistant, Research
Laboratory of Computer Semasiology
kray-voronezh@mail.ru

Victoria A. Zavarzina
orcid.org/0000-0001-6484-4606
Post-graduate Student, Assistant,
Research Laboratory
of Computer Semasiology
zva0604@yandex.ru

Voronezh State
Pedagogical University
(Voronezh, Russia)

Acknowledgments:

The study is supported by Russian Science
Foundation, project number 21-78-10148
“Modeling the meaning of a word
in individual linguistic consciousness based
on distributive semantics”



ОРИГИНАЛЬНЫЕ СТАТЬИ

Аннотация:

Целью работы является сравнительный анализ ассоциативных связей полнозначных слов из верхней зоны частотного списка, составленного на основе исследовательского корпуса текстов блогов на русском языке, в психолингвистическом эксперименте и дистрибутивно-семантической модели Global Vectors (GloVe), обученной на данном корпусе. Актуальность работы обусловлена необходимостью комплексного исследования психологически актуального значения слова. Новизна исследования заключается в том, что подобный анализ проводится с учетом фактора гендера респондента / автора текста. Использование комплекса методов интеллектуального анализа данных (кластеризация, классификация) и визуализации его результатов позволило установить влияние гендера на состав семантических ассоциатов анализируемых слов (то есть слов с близкими векторами в дистрибутивно-семантической модели) и отсутствие такого эффекта у их ассоциатов, зафиксированных в ассоциативном словаре. Как показало исследование, дистрибутивно-семантические модели и словарные ассоциативные нормы отражают разные аспекты психологически актуального содержания слова и должны использоваться как комплементарные источники при моделировании психологически актуального значения слова с учетом индивидуальных характеристик говорящего, при этом при проведении подобного анализа целесообразно использовать методы интеллектуального анализа данных.

Ключевые слова:

ассоциативный эксперимент; дистрибутивная семантика; дистрибутивно-семантическая модель; корпус текстов; интеллектуальный анализ данных.

ORIGINAL ARTICLES

Abstract:

The aim of the work is to make a comparative analysis of the associative links of full-meaning words from the upper zone of the frequency list, compiled on the basis of a research corpus of blog texts in Russian, in a psycholinguistic experiment and the distributive-semantic model Global Vectors (GloVe), trained on this corpus. The relevance of the work is due to the need for a comprehensive study of the psychologically relevant meaning of the word. The novelty of the study lies in the fact that such an analysis is carried out taking into account the gender factor of the respondent / author of the text. The use of a set of methods for data mining (clustering, classification) and visualization of its results made it possible to establish the influence of gender on the composition of the semantic associates of the analyzed words (that is, words with close vectors in the distributive-semantic model) and the absence of such an effect in their associates recorded in the associative dictionary. As the study showed, distributive-semantic models and dictionary associative norms reflect different aspects of the psychologically relevant content of the word and should be used as complementary sources when modeling the psychologically relevant meaning of the word, taking into account the individual characteristics of the speaker, while conducting such an analysis it is advisable to use data mining methods.

Key words:

association experiment; distributive semantics; distributive-semantic model; corpus of texts; data mining.



Фактор гендера в ассоциативных связях слов: данные словаря и дистрибутивно-семантической модели

© Литвинова Т. А., Котлярова Е. С., Заварзина В. А., 2022

1. Введение = Introduction

Важным направлением современных семантических и психолингвистических исследований является изучение психологически актуального для носителей языка содержания слова [Пищальникова, 2019б; Васильева, 2020]. В подобных работах активно используются такие методы исследований, как ассоциативный эксперимент (направленный и свободный), семантическое шкалирование и др., при этом ассоциативный эксперимент является «одним из основных методов сбора эмпирических данных при моделировании ассоциативного поля, которое ... используется для решения самых разных теоретических и прагматических задач» [Ассоциативный эксперимент ..., 2019, с. 5].

Подобная методология обладает большой объясняющей силой при условии учета влияния на результаты экспериментов совокупности собственно лингвистических и экстралингвистических факторов, в том числе связанных с характеристиками респондентов. В последние годы создаются специальные ассоциативные словари (например, ассоциативный словарь школьников [Гольдин, 2017], гендерный психолингвистический словарь [Гендерный..., 2020]), возрастной психолингвистический словарь [Возрастной..., 2020]) и базы данных, которые содержат информацию о респондентах (прежде всего пол, а также возраст, профессия) (РАТ, СИБАС). Указанные источники содержат уникальный психолингвистический материал, который до настоящего времени не в полной мере исследован и осмыслен.

Сбор и интерпретация данных ассоциативных экспериментов представляют собой сложную исследовательскую задачу (см. подробнее об этом: [Пищальникова, 2019а]), требующую значительных временных и трудовых ресурсов. В связи с этим в исследованиях последних лет (преимущественно англоязычных) для изучения феномена психологически актуального содержания слова стали применяться дистрибутивно-семантические модели (далее также — ДСМ), основанные на идее о том, что слова, которые используются в похожих контекстах, близки по значению [Kumar et al., 2021], при этом в дистрибутивной семантике контекст, как правило, понимается как окно заданной длины, в которое входит искомое слово (например, 2 слова до и после искомого слова).



Следует отметить, что ДСМ разрабатываются и используются преимущественно для решения прикладных задач информационного поиска (определение тематики документа, кластеризация документов и т. д.) специалистами в области компьютерной лингвистики, IT-специалистами. Однако стремительное развитие данной области, разработка готовых (предобученных) моделей, доступных для скачивания, привела к появлению исследований, в которых подобные модели используются для решения теоретических проблем семантики, см. подробнее об этом: [Boleda, 2020]. Преимуществом ДСМ в сравнении с традиционными психолингвистическими методами изучения значения слова является то обстоятельство, что с помощью первых можно исследовать семантическую структуру лексики на того или иного языка в полном объеме, тогда как, к примеру, самая большая база данных ассоциаций на английском языке “Small World of Words” [De Deyne, 2019] содержит 12 000 стимулов, что намного меньше общего числа слов в языке.

Взрывной рост интереса к названным моделям отмечается начиная с выхода модели word2vec [Mikolov et al., 2013], авторы которой предложили принципиально новый (основанный на использовании нейронных сетей) способ получения векторов малой размерности (см. подробнее о развитии ДСМ в работе [Kumar et al., 2021]). Такие низкоразмерные векторные представления каждого слова по корпусу (набору) текстов носят название *word embeddings*, не имеющее общепринятого перевода на русский язык. Предполагается, что знание о смысле слова распределено между всеми компонентами вектора, при этом каждая компонента отвечает за некоторый признак, входящий в значение слова [Потапенко, 2018].

В последние годы семантические модели, использующие машинное обучение (их называют предсказательными), активно развиваются; появляются новые модели, основанные на методах глубокого обучения [Kumar et al., 2021].

В настоящее время появляются экспериментальные доказательства того, что представление значений слов в дистрибутивно-семантических моделях ближе, чем тезаурусное представление, к тому, как значение слова хранится и обрабатывается в мозге человека [Sassenhagen et al., 2020]. Высказываются мысли о том, что модели векторного представления значений слов являются новой методологией исследования человеческого разума [Там же].

Заметим, однако, что большинство собственно лингвистических исследований с применением аппарата ДСМ выполняется на материале английского языка; как справедливо отмечает М. К. Тимофеева, лингвистические аспекты дистрибутивной семантики применительно к русскому языку остаются малоисследованными [Тимофеева, 2018а, 2018б].



Дистрибутивно-семантические модели, однако, не позволяют напрямую идентифицировать семантические характеристики слова, в связи с чем основным методом анализа значения слова при использовании ДСМ является анализ его ближайших соседей, то есть слов, чьи вектора наиболее близки к вектору данного слова, а значит, «находятся с входным словом в определенном семантическом отношении» [Тимофеева, 2018б, с. 77]. В нашей работе мы будем условно называть ближайших соседей слова в ДСМ *семантическими ассоциатами* (вслед за создателями сервиса RusVectoRēs, содержащего набор семантических моделей для русского языка; данный термин применительно к ближайшим соседям слова в ДСМ используется, например, в работах М. К. Тимофеевой [Тимофеева, 2018а, 2018б], в которых было показано, что наиболее частотными типами семантических отношений между искомым словом и его семантическими ассоциатами в проанализированных ею ДСМ русского языка являются отношения гипонимии и согипонимии).

Важно отметить, что ДСМ, как правило, обучаются на корпусах текстов, содержащих речевую продукцию большого числа носителей языка, при этом индивидуальным характеристикам авторов текстов не уделяется внимание. Названное обстоятельство ведет к получению обобщенных семантических представлений, не отражающих специфику психологически актуального для отдельного индивида значения слова.

В то же время ДСМ могут быть чрезвычайно полезны для изучения аспектов значения слова, актуальных для мужчин и женщин, лиц разных возрастных и социальных групп и т. д. В исследованиях, направленных на сравнение ассоциатов, зафиксированных в психолингвистическом эксперименте и отраженных в базах данных (словарях) (далее будем называть их словарными ассоциатами), и ассоциатов ДСМ, полученных на основе корпусов текстов (семантических ассоциатов), не уделяется внимания характеристикам респондентов / авторов текстов, чем обусловлена новизна работы.

Целью исследования является сравнительный анализ словарных и семантических ассоциатов слов из верхней зоны списка самых частотных полнозначных слов корпуса блогов с учетом фактора гендера респондента / автора блога с использованием комплекса методов data mining, в том числе многомерного анализа и визуализации данных.

2. Материалы, методы, обзор = Material, Methods, Review

В российской лингвистике накоплено множество данных об особенностях ассоциативных реакций мужчин и женщин, лиц разных возрастных групп и т. д. [Горошко, 2001; Значение как феномен актуального..., 2015].



Однако, несмотря на то что «ассоциативные нормы благодаря своей статистической “благонадежности” легко поддаются математической обработке, являясь уникальным материалом для выдвижения и проверки статистических гипотез» [Леонтьев, 1997, с. 2], выводы об особенностях ассоциативных реакций лиц тех или иных групп делаются, как правило, на основании простого сравнения пропорций (например, 14 % женщин и 20 % мужчин дают реакции N на слово-стимул A), без какого-либо статистического анализа и проверки гипотез. Мы полностью разделяем мнение Е. И. Горошко о том, что дополнение обычно проводимого в подобных работах качественного анализа статистическими методами привело бы к существенному повышению надежности выводов исследований, выполняемых на основе данных ассоциативного эксперимента [Горошко, 2001]. Кроме того, во многих работах анализируются ассоциативные реакции небольшого числа респондентов; практически не учитывается взаимодействие факторов, которые оказывают влияние на характер ассоциативных реакций (например, пола, возраста, профессии информантов).

Вопрос о вызванных теми или иными связанными с характеристиками авторов текстов различиях в семантических ассоциатах слов также является сложным и малоисследованным; не разработана методология подобных исследований. Существующие единичные решения требуют большого объема данных и сложны в реализации. Как показали наши экспериментальные исследования, одной из наиболее удачных из предложенных на сегодняшний день является методика, изложенная в работе 2022 года [Rodriguez et al., 2022]. Преимуществом предложенного авторами подхода является то, что он дает интерпретируемые результаты. Кроме того, названный подход позволяет выявлять статистически значимые отличия в семантических ассоциатах слов, обусловленные теми или иными анализируемыми факторами (например, характеристиками авторов текстов), путем анализа групповых эмбедингов. Кроме того, предложенный в работе [Rodriguez et al., 2022] подход позволяет получать вектора даже для слов, отсутствующих в предобученной модели (путем усреднения векторов соседних слов), и работать с корпусами малого объема.

В трудах, направленных на качественное сравнение словарных и семантических ассоциатов слов, как правило, отмечается невысокий процент их совпадения, что рядом исследователей объясняется различиями в структуре и типологических характеристиках ассоциативных полей и ДСМ [Nematzadeh et al., 2017; Rodriguez et al., 2020].

Высказывается мнение о том, что словарные ассоциаты слов, являясь воплощением внутренних языковых моделей (Internal language models, I-language), отражают язык как совокупность знаний в мозгу говорящего; ДСМ



рассматривают язык как внешний объект (External language models, E-language), состоящий из всех высказываний говорящих на языке [Taylor 2012].

Новейшие исследования, выполненные с привлечением визуальных данных [Kumar et al., 2021], показали, что словарные и семантические ассоциаты слова отражают его значение на разных уровнях.

В целом, как свидетельствует проведенный нами анализ литературы, проблема сопоставления словарных и семантических ассоциатов слов является активно изучаемой в настоящее время (см. обзоры [Kumar, 2021; Kumar et al., 2021]), и многое в данной области еще предстоит открыть и многие гипотезы проверить.

Следует отметить, что подавляющее большинство подобных исследований проводится на материале английского языка. Существуют единичные работы, направленные на сравнение ассоциатов слов, отраженных в ассоциативных словарях и извлеченных с помощью ДСМ, применительно к русскоязычному материалу. Так, в работе [Антипенко и др., 2020] проводилось сравнение словарных ассоциатов, выбранных из ассоциативных словарей русского языка, и семантических ассоциатов, полученных с использованием ДСМ word2vec, обученной на текстах соцсетей. Было показано, что для большинства исследуемых слов процент совпадения между словарными и семантическими ассоциатами не превышает 10—15 %. Авторы указанной работы объясняют названный факт прежде всего динамикой языкового сознания русскоговорящих (словарные ассоциативные нормы отражают результаты экспериментов, проведенных в более ранние периоды времени, чем время написания текстов, на которых обучали word2vec).

В указанных выше исследованиях не принимаются во внимание характеристики информантов / авторов.

В качестве материала настоящего исследования использовался созданный нами корпус текстов блогов на русском языке объемом 3 072 866 токенов (после удаления стоп-слов), средняя длина текста одного автора (все посты автора объединялись в один) составила 1300 токенов ($sd = 865$ слов). Авторств-женщин в анализируемом корпусе — 966, мужчин — 1398. Алгоритм сбора корпуса описан нами в работе [Litvinova et al., 2018]. Для целей настоящего исследования корпус был подвергнут ручному анализу: удалялись профили, в которых сведения о поле автора не совпадали с полом автора, эксплицитно выраженным через соответствующие грамматические формы; профили авторов, чьи тексты целиком состояли из заимствованных текстов (новостей и т. д.). Очищенный таким образом корпус доступен в созданной нами базе данных RusIdiolect, специально предназначенной для идиолектных исследований [Litvinova, 2021]. Средний возраст женщин в корпусе — 32,7 года ($sd = 9,9$), мужчин — 39,5 года ($sd = 11,3$).



Насколько нам известно, только в двух общедоступных ассоциативных базах данных приведена информация о характеристиках респондентов: это РАТ и СИБАС, однако время проведения ассоциативных экспериментов, результаты которых легли в основу СИБАС (2008—2013), ближе ко времени создания блогов (2014—2018), в связи с чем именно СИБАС (основной корпус) был выбран нами в качестве материала для сравнения, несмотря на возможную региональную специфику ассоциативных полей. Как было показано в работе [Антипенко и др., 2020], а также в многочисленных работах, в которых сравнивались словарные ассоциаты слов в различных ассоциативных базах, в том числе РАС и СИБАС (см., например, [Кафтанов, 2019]), состав многих ассоциативных полей изменяется со временем, в связи с чем в данном эксперименте нами была выбрана СИБАС в качестве материала для сравнения семантических ассоциатов исследуемых слов как наиболее близкая база из существующих по времени сбора материала ко времени создания текстов исследовательского корпуса и содержащая данные о гендере информантов.

Для каждого анализируемого слова (список слов и критерии их отбора приведены ниже) из СИБАС нами выбирались ассоциаты с частотностью не менее 4 (как для респондентов-женщин, так и для респондентов-мужчин).

Хорошо известно, что обработка текста перед проведением дальнейших экспериментов является важным этапом работы, которому необходимо уделять пристальное внимание [Rodriguez et al., 2022].

В наших экспериментах мы применяли минимальную предобработку: удалили токены, не являющиеся словами (символы, знаки препинания), привели текст к строчному виду. Были исключены токены, встречающиеся в корпусе менее 5 раз. Служебные слова удалены с использованием пакета *quanteda* [Quanteda..., 2018], лемматизация проводилась с использованием морфоанализатора *Treetagger* [Schmid, 1994].

Отбор слов для анализа осуществлялся по следующей схеме: были выбраны самые частотные леммы (первые 40 лемм) в блогах женщин и блогах мужчин (вычислялись относительные частоты), после чего из двух полученных списков были выбраны совпадающие леммы. Итоговый список слов, чьи словарные и семантические ассоциаты сравниваются в нашей работе, после удаления наречий, порядковых числительных, а также слов, ассоциаты которых отсутствуют в СИБАС, выглядел следующим образом: *год, свой, человек, очень, время, самый, день, наш, новый, говорить, сказать, место, дело, работа, сделать, друг, делать, мир, слово, первый, видать, хороший, думать, дом*.

Для выявления семантических ассоциатов исследуемых слов нами была выбрана модель *GloVe* [Pennington et al., 2014]. Выбор названной модели об-

условливался следующими обстоятельствами. Во-первых, названная модель была апробирована в работе авторов метода выявления группо-специфичных семантических ассоциатов [Rodriguez et al., 2022]. Во-вторых, как было показано в работе [Лыченко и др., 2019], несмотря на то что данная модель редко используется при проведении исследований на русскоязычном материале, она требует пристального внимания и изучения на материале русскоязычных корпусов текстов, так как показывает очень хорошие результаты при решении задачи анализа тональности текстов на русском языке. В-третьих, GloVe требуется намного меньше времени на обучение, чем word2vec.

В открытом доступе находится предобученная многоязычная модель *GloVe*, в том числе для русского языка, однако, поскольку нашей задачей являлось сравнение семантических ассоциатов выбранных слов в текстах женщин и мужчин, мы не использовали предобученную модель, а обучали *GloVe* на собранном нами корпусе текстов блогов, снабженном информацией о гендере их авторов, с использованием параметров, рекомендуемых в работе [Rodriguez et al., 2022] (размерность — 300, длина окна — 6), и пакетов *quanteda* [Quanteda..., 2018] и *text2vec* [Selivanov et al., 2020].

Для отобранных для анализа слов вычислялось соотношение мер косинусового расстояния для каждого из 10 его ближайших соседей в ДСМ, построенной на текстах блогов авторов-женщин и авторов-мужчин, с использованием функции *Nearest neighbors cosine similarity ratio (nns_ratio)* пакета ‘*conText*’ [Rodriguez, Spirling, 2022]. Полученные отношения отражают уровень различий во встречаемости семантического ассоциата того или иного слова в текстах женщин и мужчин (указанная функция позволяет также оценивать статистическую значимость найденных отличий).

Пример визуализации результатов анализа с использованием названной функции представлен далее (рис. 1).

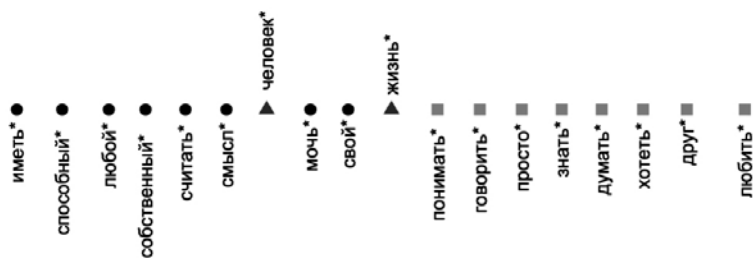


Рис. 1. Пример визуализации результатов анализа, направленного на выявление различий в составе семантических ассоциатов слова *человек* в векторном пространстве, построенном на текстах мужчин и женщин (кругом обозначены семантические ассоциаты слова *человек*, характерные для текстов мужчин, квадратом — для текстов женщин, треугольником — общие ассоциаты, $p < 0,01$)



Словарные ассоциаты слова *человек* в СИБАС у респондентов-мужчин (в порядке убывания частотности ответов) таковы: *разумный 21; существо 12; обезьяна 10; волк; животное; личность 8; я 7; паук 6; добрый; и закон 5*. Реакции респондентов женщин на указанный стимул выглядят следующим образом: *разумный 14; личность; существо 10; обезьяна; паук 9; жизнь; хороший 8; друг; животное 7; добрый; люди 6; тело 5*.

На основе полученных данных вычислялся индекс гендерной дифференциальности (ИГД) ассоциативного значения для каждого анализируемого слова в словаре и ДСМ как отношение числа несовпадающих ассоциатов в реакциях респондентов-женщин и респондентов-мужчин (ближайших соседей слова в ДСМ, построенной на текстах блогов женщин и мужчин) к общему числу разных ассоциатов анализируемого слова.

Далее мы объединили все ассоциаты каждого слова в один файл (словарные и корпусные ассоциаты входили в разные файлы), получив таким образом 92 текстовых файла (46 файлов, содержащих словарные ассоциаты изучаемых слов, — отдельно респондентов-женщин и респондентов-мужчин; 46 файлов, содержащих семантические ассоциаты тех же исследуемых слов, — отдельно для текстов женщин и мужчин). Полученные таким образом датасеты были проанализированы с использованием пакета *quanteda*: была создана матрица встречаемости признаков (ассоциатов), после этого произведен анализ полученного набора данных методом главных компонент (PCA), и рассчитаны связи между фактором (гендер респондента / автора текста) и компонентами (анализ проводился с использованием пакета *factoMineR* [Lê et al., 2008]).

3. Результаты и обсуждение = Results and Discussion

Как свидетельствуют данные, представленные на рис. 2, ИГД ассоциативных значений изучаемых слов в ДСМ в среднем выше, чем в словаре, что подтверждается также статистическим тестом ($W = 50$, $p = 0,0000025$).

Корреляционный анализ значений индексов гендерной дифференциальности ассоциативного значения исследуемых слов в ДСМ и словаре не выявил корреляций между ними ($S = 1338,6$, $p = 0,114$), однако, как показывает рис. 2, можно отметить тенденцию, согласно которой ряд слов с высоким индексом гендерной дифференциальности ассоциативного значения в ДСМ имеют также высокий индекс гендерной дифференциальности ассоциативного значения в СИБАС (дело, место, время).

3.1. Сравнительный анализ словарных и семантических ассоциатов анализируемых слов с учетом гендера информанта / автора текста

Как показал анализ, словарные ассоциаты исследуемых слов тесно связаны с реалиями периода сбора материала, например, среди реакций

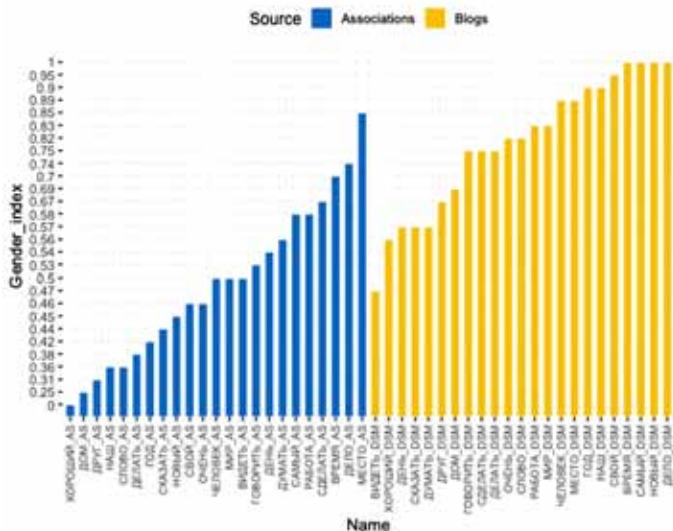


Рис. 2. Индекс гендерной дифференциальности ассоциативного значения исследуемых слов в ДСМ и СИБАС

на стимул *год* как у женщин, так и у мужчин содержатся такие ассоциаты, как *2008, 2010* (годы проведения эксперимента), *свиньи* и т. д. Семантические ассоциаты связаны с заглавным словом тематически: *май, сентябрь, февраль, декабрь* (семантические ассоциаты, характерные для текстов авторов-мужчин), *неделя, сегодня, день, месяц* (семантические ассоциаты, характерные для текстов авторов-женщин). Также наши наблюдения показывают, что для словарных ассоциатов характерны синтагматические связи (время — *бежит, идет*), для семантических — парадигматические (время — *неделя, день, момент, год*). Полученные данные находятся в соответствии с наблюдениями, сделанными в работе [Антипенко и др., 2020], а также в работах М. К. Тимофеевой [Тимофеева, 2018а, 2018б].

Среднее значение совпадений словарных и корпусных ассоциатов слов составляет 3 %.

Анализ семантических ассоциатов анализируемых слов в текстах мужчин и женщин методом главных компонент, проведенный с использованием пакета *factoMineR*, показал, что первая компонента достоверно связана с фактором «пол» ($R^2 = 0,091$, $p = 0,042$) (рис. 3), как и вторая компонента ($R^2 = 0,095$, $p = 0,0368$). Иными словами, гендерный эффект является системным и оказывает влияние на состав семантических ассоциатов всех рассматриваемых слов в совокупности.

Для словарных ассоциатов связи с гендерным фактором выявлено не было.

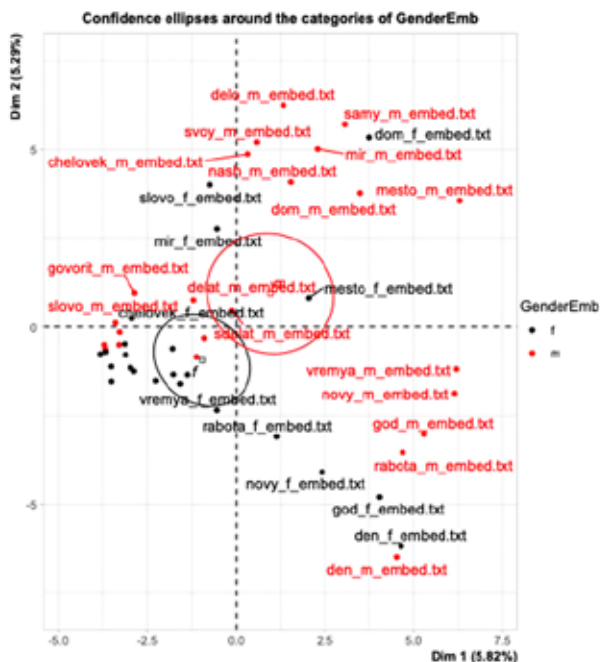


Рис. 3. Анализ семантических ассоциатов анализируемых слов в текстах мужчин и женщин с использованием метода главных компонент

Для текстов женщин более характерны (в порядке убывания значения v -test) семантические ассоциаты *хотеться*, *любить*, *хотеть*, *наверное*, *общее*, *хороший*, для мужчин — *свой*, *ночь*.

Далее мы провели эксперимент по классификации ассоциативных значений анализируемых слов в зависимости от гендера респондента / автора текста с использованием алгоритма PLS-DA с 5-фолдной кросс-проверкой. Было установлено, что на основе набора семантических ассоциатов возможно классифицировать значение анализируемых слов как характерное для женщины / мужчин со средней ошибкой CER = 0,33913. Пермутационный тест показал, что различия между классами являются значимыми ($p = 0,006$). Для словарных ассоциатов различий между классами не выявлено (CER = 0,56957, $p = 0,924$). Иными словами, классификатор показал, что ассоциативное значение анализируемых слов, отраженное в ДСМ, имеет гендерную специфику.

Также мы выполнили кластерный анализ ассоциатов анализируемых слов (использовались разные меры расстояния; на рис. 4—5 представлены результаты кластерного анализа с использованием меры расстояния Жаккарда, специально предназначенной для бинарных данных, по методу Варда). Подобное представление данных позволяет выявить кластеры слов, чьи ассоциаты наиболее близки друг другу, для дальнейшего содержательного анализа.

На рис. 4 выделяются «гендерные» кластеры, отражающие различия в ассоциативных значениях слов в терминах ДСМ в текстах женщин и мужчин.

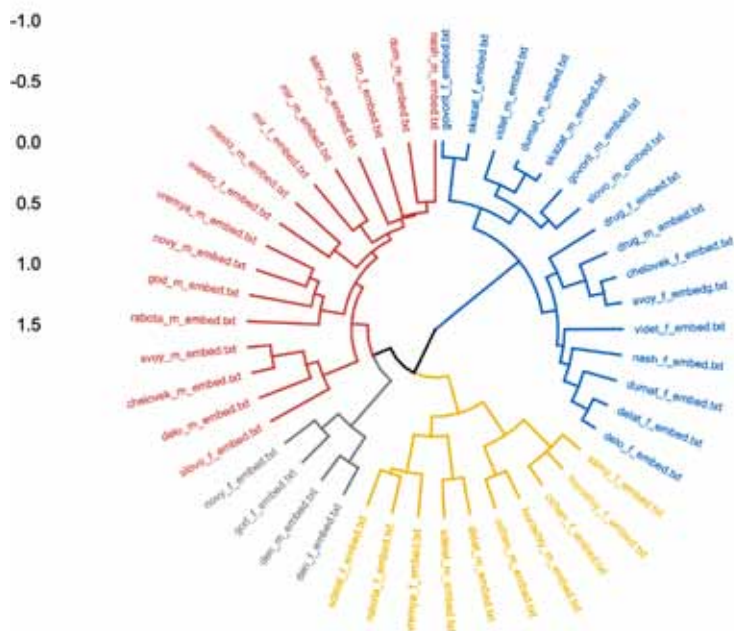


Рис. 4. Кластерный анализ семантических ассоциатов исследуемых слов

Как показывает рис. 5, словарные ассоциаты исследуемых слов, зафиксированные в реакциях женщин и мужчин, объединяются в мелкие кластеры. Более крупные кластеры образуют слова, близкие по значению; их ассоциаты схожи, независимо от гендера информантов (например, серый кластер, содержащий слова-стимулы *слово*, *говорить*, *сказать*; синий подкластер, состоящий из слов-стимулов *наши*, *свой*; желтый кластер — *делать*, *сделать*).

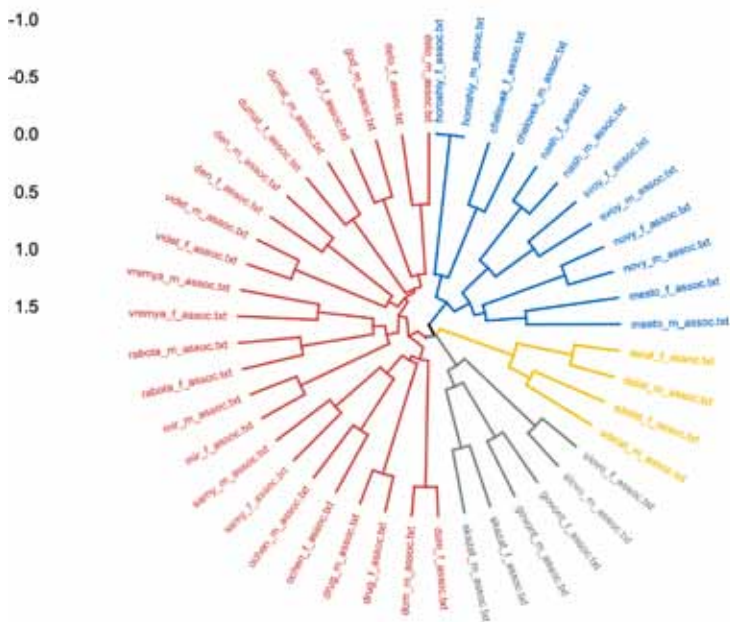


Рис. 5. Кластерный анализ словарных ассоциатов исследуемых слов

Отметим, что в данном случае мы использовали бинарные значения (наличие / отсутствие слова среди ассоциатов искомого слова); учет силы ассоциативной связи (выраженной как отношение числа респондентов, давших реакцию на заданный стимул, к общему числу респондентов в случае словарных ассоциатов либо значение отношения мер косинусового расстояния в случае семантических ассоциатов) может дать еще более точную картину, отражающую близость ассоциативных значений слов.

4. Заключение = Conclusions

Таким образом, проведенное нами сравнительное исследование ассоциативных связей исследуемых слов по данным ассоциативного словаря и дистрибутивно-семантической модели в гендерном аспекте позволило сделать следующие выводы.

Задача выявления гендерно-специфичных семантических ассоциатов слова (ближайших соседей слова в дистрибутивно-семантической модели) является сложной исследовательской задачей, методология решения которой еще только начинает разрабатываться. При выборе методов ее ре-



нения следует учитывать возможность интерпретируемости результатов и применимости методов к текстам малого объема. Подход, сочетающий в себе регрессионный анализ и представление векторов редких слов на основе усреднения векторов соседних слов, предложенный изначально для решения задач социальных наук (в частности, исследования тех или иных концептов в речи представителей различных политических партий), может быть успешно использован для сравнительного исследования семантических ассоциатов слов в гендерном аспекте.

Списки семантических и словарных ассоциатов анализируемых слов имеют низкий процент совпадений (в среднем 3 %). Низкий процент совпадений названных типов ассоциатов отмечался ранее в многочисленных исследованиях, выполненных на материале английского языка, а также в работе, выполненной на материале текстов соцсетей на русском языке. Названное обстоятельство свидетельствует о том, что словарные и семантические ассоциаты отражают разные аспекты психологически актуального содержания слова.

Словарные ассоциаты тесно связаны с реалиями периода их сбора, что необходимо учитывать при проведении анализа.

Индекс гендерной дифференциальности ассоциативного значения слов в ДСМ выше, чем в словаре, что подтверждается статистическим тестом.

С помощью различных методов многомерного анализа нами был выявлен системный эффект гендера, оказываемый им на ассоциативное значение наиболее частотных слов корпуса блогов на русском языке, извлеченное при помощи дистрибутивно-семантической модели. Для словарных ассоциатов данного эффекта не наблюдалось.

Кластерный анализ ассоциатов анализируемых слов позволяет выявить группы слов с близким ассоциативным значением, что позволяет перейти от анализа значений отдельных слов к выявлению взаимосвязей между ними.

Безусловно, наше исследование, являясь пилотным, имеет ряд ограничений. Мы исследовали только один жанр текстов, один тип ДСМ и одну базу данных ассоциаций (причем региональную). Однако проведенное пилотное исследование показало, что фактор гендера оказывает системный эффект на ассоциативное значение самых частотных полнзначных слов анализируемого корпуса. Все дистрибутивно-семантические модели отражают значения слов, представленные в корпусах текстов, на которых они обучались. Следовательно, при отборе текстов для обучения ДСМ необходимо уделять пристальное внимание характеристикам авторов текстов. В настоящее время названному вопросу не уделяется внимания, что приводит к тому, что предобученные ДСМ отражают взгляды доминирующей социальной группы [Shah et al., 2020].



Дистрибутивно-семантические модели и словарные ассоциативные нормы, по всей видимости, отражают разные аспекты психологически актуального содержания слова и должны использоваться как комплиментарные методы при моделировании значения слова с учетом индивидуальных характеристик говорящего. Преимущество первого типа данных заключается в том, что они могут быть собраны и проанализированы в достаточно короткие сроки, однако при этом необходимо уделять внимание разметке корпуса по факторам идиолектного варьирования, относящимся как к зоне «Автор», так и к зоне «Условия осуществления идиолектной деятельности» [Litvinova, 2021].

Российскими исследователями собраны уникальные базы ассоциаций с данными об информантах, но их исследовательский потенциал еще не раскрыт в полной мере. В частности, перспективным направлением дальнейших исследований является анализ ассоциаций с использованием программных пакетов, специально предназначенных для построения и анализа семантических сетей. Комбинирование подобных инструментов и разработанных отечественными психолингвистами [Пищальникова, 2019а; Пищальникова, 2019б] методов исследования отношения «стимул — реакция» (в противоположность наиболее распространенному в текущей исследовательской практике анализу конкретного содержания ассоциата и частотности отдельных ассоциатов) позволит получить принципиально новые результаты моделирования содержания слова.

Комплексный анализ дистрибутивно-семантических моделей и словарных ассоциаций в совокупности с метаданными о респондентах (авторов текстов) с применением новых методов data mining и визуализацией результатов позволит получить новые данные о психологически актуальном значении слова для говорящих с учетом их индивидуальных характеристик.

В качестве ближайших перспектив исследования следует отметить проведение анализа применительно к словам, отобранным не по частотности, а в зависимости от их принадлежности к тем или иным лексико-семантическим группам, а также семантическую разметку словарных и семантических ассоциатов.

Источники и принятые сокращения

1. РАТ — *Русский ассоциативный тезаурус* [Электронный ресурс]. — Режим доступа : <http://thesaurus.ru/dict/> (дата обращения 21.11.2021).
2. СИБАС — (*Сибирский ассоциативный словарь*) Русская региональная ассоциативная база данных (Сибирь и Дальний Восток) [Электронный ресурс]. — 2008—2020. — Режим доступа : <http://adict.ru.nsu.ru> (дата обращения 24.11.2021).
3. GloVe — *Global Vectors* [Electronic resource]. — Access mode : https://www.cs.cmu.edu/~afm/projects/multilingual_embeddings.html (accessed 24.11.2021).



4. RusVectōrēs — *Веб-сервис* дистрибутивно-семантических моделей для русского языка. — Режим доступа : <http://rusvectors.org/ru/> (дата обращения 24.11.2021).

5. Selivanov, D., Bickel, M., & Wang, Q. (2020). text2vec: Modern text mining framework for R. Version 0.6.1. <https://cran.rstudio.com/web/packages/text2vec/>.

Литература

1. *Антипенко А. А.* Сравнительный анализ ассоциаций в корпусах социальных сетей на основе дистрибутивно-семантических моделей для русского языка / А. А. Антипенко, О. А. Митрофанова // International Journal of open information technologies. — 2020. — № 1. — С. 27—33.

2. *Ассоциативный эксперимент* : Теоретические и прикладные перспективы психолингвистики : монография / В. А. Пищальникова ; под ред. В. А. Пищальниковой. — Москва : Р-Валент, 2019. — 200 с. — ISBN 978-5-93439-575-0

3. *Васильева С. П.* Опыт выявления специфики обыденного языкового сознания русских по данным ассоциативных словарей XX—XXI веков / С. П. Васильева // Научный диалог. — 2020. — № 1. — С. 27—44. — DOI: 10.24224/2227-1295-2020-1-27-44.

4. *Возрастной дифференциальный психолингвистический словарь русского языка* / сост.: И. А. Стернин, А. В. Рудакова. — Воронеж : Центр коммуникативных исследований ВГУ, 2020. — 158 с. — ISBN 978-5-6045255-5-5.

5. *Гендерный дифференциальный психолингвистический словарь* / науч. ред. : И. А. Стернин, А. В. Рудакова. — Воронеж : Ритм, 2020. — 198 с. — ISBN 978-5-6045255-7-9.

6. *Гольдин В. Е.* Проект «Ассоциативный словарь школьников : стабильность и динамика» / В. Е. Гольдин, А. П. Сдобнова // Вопросы психолингвистики. — 2017. — № 4 (34). — С. 174—187.

7. *Горошко Е. И.* Интегративная модель свободного ассоциативного эксперимента / Е. И. Горошко. — Харьков ; Москва : РА — Каравелла, 2001. — 320 с.

8. *Значение как феномен актуального языкового сознания носителя языка* : материалы межрегиональной научной конференции / науч. ред. И. А. Стернин. — Воронеж : Истоки, 2015. — 69 с.

9. *Кафтанов Р. А.* Динамика образа войны в русском языковом сознании (психолингвистический аспект) / Р. А. Кафтанов // Вестник НГУ. Серия : Лингвистика и межкультурная коммуникация. — 2019. — Т. 17. — № 1. — С. 149—160. — DOI: <https://doi.org/10.25205/1818-7935-2019-17-1-149-160>.

10. *Леонтьев А. А.* Основы психолингвистики / А. А. Леонтьев. — Москва : Смысл, 1997. — 287 с. — ISBN 5-89357-191-6.

11. *Лыченко Н. М.* Сравнение эффективности методов векторного представления слов для определения тональности текстов / Н. М. Лыченко, А. В. Сорокова // МСйМ. — 2019. — № 4 (52). — С. 97—110. — DOI: 10.24147/2222-8772.2019.4.97-110.

12. *Пищальникова В. А.* Интерпретация ассоциативных данных как проблема методологии психолингвистики / В. А. Пищальникова // Russian Journal of Linguistics. — 2019a. — Т. 23. — № 3. — С. 749—761. — DOI: 10.22363/2312-9182-2019-23-3-749-761.

13. *Пищальникова В. А.* Эксперимент как составная часть методологии сопоставительных исследований / В. А. Пищальникова // Вестник Волгоградского государственного университета. Серия 2, Языкознание. — 2019b. — Т. 8. — № 3. — С. 52—63. — DOI: <https://doi.org/10.15688/jvolsu2.2019.3.4>.



14. *Потапенко А. А.* Семантические векторные представления текста на основе вероятностного тематического моделирования : диссертация ... кандидата физико-математических наук : 05.13.17 / А. А. Потапенко. — Москва, 2018. — 147 с.
15. *Тимофеева М. К.* Возможности использования сервиса RusVectōrēs для выявления семантических ассоциатов глаголов русского языка / М. К. Тимофеева // Научный диалог. — 2018а. — № 9. — С. 117—131. — DOI: 10.24224/2227-1295-2018-9-117-131.
16. *Тимофеева М. К.* Типология семантических отношений, выявляемых посредством инструмента RusVectōrēs / М. К. Тимофеева // Научный диалог. — 2018б. — № 8. — С. 74—87. — DOI: 10.24224/2227-1295-2018-8-74-87.
17. *Boleda G.* Distributional Semantics and Linguistic Theory / G. Boleda // Annual Review of Linguistics. — 2020. — Vol. 6 (1). — Pp. 213—234. — DOI: 10.1146/annurev-linguistics-011619-030303.
18. *De Deyne S.* The “Small World of Words” English word association norms for over 12,000 cue words / S. De Deyne // Behav Res Methods. — 2019. — Vol. 51 (3). — Pp. 987—1006. — DOI: <https://doi.org/10.3758/s13428-018-1115-7/>.
19. *Kumar A. A.* A Critical Review of Network-Based and Distributional Approaches to Semantic Memory Structure and Processes / A. A. Kumar, M. Steyvers, D. A. Balota // Topics in cognitive science. — 2021, Jun 6. — DOI: 10.1111/tops.12548.
20. *Kumar A. A.* Semantic memory : A review of methods, models, and current challenges / A. A. Kumar // Psychon Bull Rev. — 2021. — Vol. 28 (1). — Pp. 40—80. — DOI: <https://doi.org/10.3758/s13423-020-01792-x>.
21. *Lê S.* FactoMineR : A Package for Multivariate Analysis / S. Lê, J. Josse, F. Husson // Journal of Statistical Software. — 2008. — Vol. 25 (1). — Pp. 1—18. — DOI: 10.18637/jss.
22. *Litvinova T.* Profiling the Age of Russian Bloggers / T. Litvinova, A. Soeov, P. Panicheva // Communications in Computer and Information Science. — 2018. — Vol. 930. — Pp. 167—177. — DOI: 10.1007/978-3-030-01204-5_16.
23. *Litvinova T.* RusIdiolect : A New Resource for Authorship Studies / T. Litvinova // Lecture Notes in Networks and Systems. — 2021. — Vol. 186. — Pp. 14—23. — DOI: https://doi.org/10.1007/978-3-030-66093-2_2.
24. *Mikolov T.* Efficient Estimation of Word Representations in Vector Space / T. Mikolov, K. Chen, G. Corrado, J. Dean // Proceedings of Workshop at ICLR, arXiv. — 2013. — Pp. 1301—3781.
25. *Nematzadeh A.* Evaluating Vector-Space Models of Word Representation, or, The Unreasonable Effectiveness of Counting Words Near Other Words / A. Nematzadeh, S. Meylan, T. Griffiths // Proceedings of the 39th Annual Meeting of the Cognitive Science Society. — London, UK : [b. i], 2017. — Pp. 859—854.
26. *Pennington J.* GloVe : Global Vectors for Word Representation / J. Pennington, R. Socher, C. Manning // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). — 2014. — Vol. 14. — Pp. 1532—1543.
27. *Quanteda :* An R package for the quantitative analysis of textual data / K. Benoit, K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, A. Matsuo // Journal of Open Source Software. — 2018. — Vol. 3 (30). — 774 p. — DOI: 10.21105/joss.00774.
28. *Rodriguez M. A.* Word associations and the distance properties of context-aware word embeddings / M. A. Rodriguez, P. Merlo // CONLL. — 2020. — November 19—20. — Pp. 376—385. — DOI: 10.18653/v1/2020.conll-1.30.
29. *Rodriguez P. L.* Word Embeddings : What Works, What Doesn’t, and How to Tell the Difference for Applied Research / P. L. Rodriguez, A. Spirling // The Journal of Politics. — 2022. — Vol. 84 (1). — Pp. 101—115. — DOI: <https://doi.org/10.1086/715162>.



30. *Sahlgren M.* The Distributional Hypothesis. From context to meaning / M. Sahlgren // *Distributional models of the lexicon in linguistics and cognitive science* (Special issue of the Italian Journal of Linguistics), *Rivista di Linguistica*. — 2008. — Vol. 20. — № 1. — Pp. 33—53.
31. *Sassenhagen J.* Traces of Meaning Itself : Encoding Distributional Word Vectors in Brain Activity / J. Sassenhagen, C. Fiebach // *Neurobiology of Language*. — 2020. — Vol. 1 (1). — Pp. 54—76. — DOI: https://doi.org/10.1162/nol_a_00003
32. *Schmid H.* Probabilistic Part-of-Speech Tagging Using Decision Trees / H. Schmid // *Proceedings of International Conference on New Methods in Language Processing*. — Manchester : UK, 1994.
33. *Taylor J. R.* The mental corpus : How language is represented in the mind / J. R. Taylor. — Oxford : University Press, 2012. — 384 p.

Material resources

- RusVectōrēs — *A web service of distributive semantic models for the Russian language*. Available at: <http://rusvectors.org/ru/> (accessed 24.11.2021). (In Russ.).
- GloVe — *Global Vectors*. Available at: https://www.cs.cmu.edu/~afm/projects/multilingual_embeddings.html (accessed 24.11.2021).
- RAT — *Russian associative thesaurus*. Available at: <http://thesaurus.ru/dict/> (accessed 21.11.2021). (In Russ.).
- Selivanov et al. — *Selivanov, D., Bickel, M., Wang, Q.* (2020). text2vec: Modern text mining framework for R. Version 0.6.1 / D. Selivanov, M. Bickel, Q. Wang. Available at: <https://cran.rstudio.com/web/packages/text2vec/>.
- SIBAS — *(Siberian Associative Dictionary) Russian Regional Associative database (Siberia and the Far East)*. (2008—2020). Available at: <http://adictru.nsu.ru> (accessed 24.11.2021). (In Russ.).

References

- Age differential psycholinguistic dictionary of the Russian language*. (2020). Voronezh: VSU Center for Communication Studies. 158 p. ISBN 978-5-6045255-5-5.
- Antipenko, A. A., Mitrofanova, O. A. (2020). Comparative analysis of associations in social network corpora based on distributive semantic models for the Russian language. *International Journal of open information technologies, 1*: 27—33. (In Russ.).
- Associative experiment: Theoretical and applied perspectives of psycholinguistics: monograph*. (2019). Moscow: R-Valent. 200 p. ISBN 978-5-93439-575-0. (In Russ.).
- Boleda G. (2020). Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics, 6* (1): 213—234. DOI: 10.1146/annurev-linguistics-011619-030303.
- De Deyne, S. (2019). The “Small World of Words” English word association norms for over 12,000 cue words. *Behav Res Methods, 51* (3): 987—1006. DOI: <https://doi.org/10.3758/s13428-018-1115-7/>.
- Gender differential psycholinguistic dictionary*. (2020). Voronezh: Rhythm. 198 p. ISBN 978-5-6045255-7-9. (In Russ.).
- Goldin, V. E., Sdobnova, A. P. (2017). Project “Associative dictionary of schoolchildren: stability and dynamics”. *Questions of psycholinguistics, 4* (34): 174—187. (In Russ.).
- Goroshko, E. I. (2001). *Integrative model of a free associative experiment*. Kharkiv; Moscow: RA — Karavella. 320 p. (In Russ.).
- Quanteda: An R package for the quantitative analysis of textual data. *Journal of Open Source Software, 3* (30): P. 774. DOI: 10.21105/joss.00774.



- Kaftanov, R. A. (2019). Dynamics of the image of war in the Russian linguistic consciousness (psycholinguistic aspect). *Bulletin of the NSU. Series: Linguistics and Intercultural Communication*, 17 (1): 149—160. DOI: <https://doi.org/10.25205/1818-7935-2019-17-1-149-160>. (In Russ.).
- Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychon Bull Rev*, 28 (1): 40—80. DOI: <https://doi.org/10.3758/s13423-020-01792-x>.
- Kumar, A. A., Steyvers, M., Balota, D. A. (2021). A Critical Review of Network-Based and Distributional Approaches to Semantic Memory Structure and Processes. *Topics in cognitive science*. Jun 6. DOI: 10.1111/tops.12548.
- Leontiev, A. A. (1997). *Fundamentals of psycholinguistics*. Moscow: Smysl. 287 p. ISBN 5-89357-191-6. (In Russ.).
- Lê, S., Josse, J., Husson, F. (2008). FactoMineR: A Package for Multivariate Analysis. *Journal of Statistical Software*, 25 (1): 1—18. DOI: 10.18637/jss.
- Litvinova, T., Sboev, A., Panicheva, P. (2018). Profiling the Age of Russian Bloggers. *Communications in Computer and Information Science*, 930: 167—177. DOI: 10.1007/978-3-030-01204-5_16.
- Litvinova, T. (2021). Rusldiolect: A New Resource for Authorship Studies. *Lecture Notes in Networks and Systems*, 186: 14—23. DOI: https://doi.org/10.1007/978-3-030-66093-2_2.
- Lychenko, N. M., Sorokovaya, A. V. (2019). Comparison of the effectiveness of methods of vector representation of words for determining the tonality of texts. *MSiM*, 4 (52): 97—110. DOI: 10.24147/2222-8772.2019.4.97-110. (In Russ.).
- Mikolov, T., Chen, K., Corrado, G., Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. In: *Proceedings of Workshop at ICLR, arXiv*. 1301—3781.
- Nematzadeh, A., Meylan, S., Griffiths, T. (2017). Evaluating Vector-Space Models of Word Representation, or, The Unreasonable Effectiveness of Counting Words Near Other Words. In: *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. London, UK: [b. i]. 859—854.
- Pennington, J., Socher, R., Manning, C. (2014). GloVe: Global Vectors for Word Representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 14: 1532—1543.
- Pishchalnikova, V. A. (2019a). Interpretation of associative data as a problem of psycholinguistics methodology. *Russian Journal of Linguistics*, 23 (3): 749—761. DOI: 10.22363/2312-9182-2019-23-3-749-761. (In Russ.).
- Pishchalnikova, V. A. (2019b). Experiment as an integral part of the methodology of comparative studies. *Bulletin of Volgograd State University. Series 2, Linguistics*, 8 (3): 52—63. DOI: <https://doi.org/10.15688/jvolsu2.2019.3.4>. (In Russ.).
- Potapenko, A. A. (2018). *Semantic vector representations of text based on probabilistic thematic modeling*. PhD Diss. Moscow. 147 p. (In Russ.).
- Rodriguez, M. A., Merlo, P. (2020). Word associations and the distance properties of context-aware word embeddings. *CONLL. November 19—20*. 376—385. DOI: 10.18653/v1/2020.conll-1.30.
- Rodriguez, P. L., Spirling, A. (2022). Word Embeddings: What Works, What Doesn't, and How to Tell the Difference for Applied Research. *The Journal of Politics*, 84 (1): 101—115. DOI: <https://doi.org/10.1086/715162>.
- Sahlgren, M. (2008). 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*, 20 (1): 33—53.



- Sassenhagen, J., Fiebach, C. (2020). Traces of Meaning Itself: Encoding Distributional Word Vectors in Brain Activity. *Neurobiology of Language*, 1 (1): 54—76. DOI: https://doi.org/10.1162/nol_a_00003.
- Schmid, H. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In: *Proceedings of International Conference on New Methods in Language Processing*. Manchester: UK.
- Sternin, I. A. (ed.). (2015). *Meaning as a phenomenon of actual linguistic consciousness of a native speaker: materials of an interregional scientific conference*. Voronezh: Istoki. 69 p. (In Russ.).
- Taylor, J. R. (2012). *The mental corpus: How language is represented in the mind*. Oxford: University Press. 384 p.
- Timofeeva, M. K. (2018). Possibility of Extracting Semantic Associates of Russian Verbs by the Instrument RusVectōrēs. *Nauchnyy dialog*, 9: 117—131. DOI: 10.24224/2227-1295-2018-9-117-131. (In Russ.)
- Timofeeva, M. K. (2018). Typology of Semantic Relations Extracted by the Instrument RusVectōrēs. *Nauchnyi dialog*, 8: 74—87. <https://doi.org/10.24224/2227-1295-2018-8-74-87>. (In Russ.).
- Vasilyeva, S. P. (2020). Experience of Identifying the Specifics of Everyday Linguistic Consciousness of Russians according to Associative Dictionaries of the XX—XXI Centuries. *Nauchnyi dialog*, 1: 27—44. <https://doi.org/10.24224/2227-1295-2020-1-27-44>. (In Russ.).

Статья поступила в редакцию 03.02.2022,
одобрена после рецензирования 06.05.2022,
подготовлена к публикации 01.06.2022.