

Научная статья

УДК 81'33

DOI: 10.47438/2309-7078_2024_1_215

ОТКРЫТАЯ МНОГОКОМПОНЕНТНАЯ БАЗА ВЕРБАЛЬНЫХ АССОЦИАЦИЙ, АННОТИРОВАННАЯ ПО ТИПАМ ОТНОШЕНИЙ «СТИМУЛ – АССОЦИАТ» И ВРЕМЕНИ РЕАКЦИИ: СВЯЗЬ ХАРАКТЕРИСТИК И МЕТАДААННЫХ

Виктория Александровна Заварзина¹, Елена Сергеевна Котлярова²,
Светлана Геннадьевна Любова³

*Воронежский государственный педагогический университет^{1, 2, 3}
Воронеж, Россия*

¹Ассистент кафедры теории, истории и методики преподавания русского языка и литературы, лаборант научно-исследовательской лаборатории компьютерной семасиологии, ORCID ID: 0000-0001-6484-4606, e-mail: zva0604@yandex.ru

²Аспирант, старший лаборант научно-исследовательской лаборатории компьютерной семасиологии, ORCID ID: 0000-0002-4162-5002, e-mail: kray-voronezh@mail.ru

³Кандидат филологических наук, старший преподаватель кафедры теории, истории и методики преподавания русского языка и литературы, ORCID ID: 0000-0002-2586-410X, e-mail: svetlana_lyubova_2305@mail.ru

Аннотация. Ассоциативный эксперимент является одним из наиболее эффективных психолингвистических методов моделирования значения слова в индивидуальном языковом сознании. В связи с трудоемкостью проведения ассоциативных экспериментов и важностью полученных эмпирических данных для решения различных лингвистических задач, а следовательно, для широкого круга исследователей особую значимость представляют общедоступные корпуса, содержащие результаты подобных экспериментов. Отечественными и зарубежными исследователями созданы и поддерживаются ряд подобных ресурсов, однако они, как правило, не содержат информации об индивидуальном ассоциативном поведении, а также результаты лингвистического аннотирования. В настоящей статье представлено описание созданной в ВГПУ при поддержке гранта РНФ и не имеющей аналогов в мировой лингвистике базы данных ассоциаций, содержащей метаданные разных видов, процесс и результаты аннотирования ассоциативных данных по типам отношений «стимул – реакция» по разработанной авторами классификации, а также сведения о взаимосвязи различных типов метаданных.

Ключевые слова: психолингвистическое значение слова, ассоциативный эксперимент, ассоциативная реакция, лингвистическая разметка, корпус текстов.

Для цитирования: Заварзина В.А., Котлярова Е.С., Любова С.Г. Открытая многокомпонентная база вербальных ассоциаций, аннотированная по типам отношений «стимул – ассоциат» и времени реакции: связь характеристик и метаданных // Известия Воронежского государственного педагогического университета. 2024. № 1. С. 215–223. DOI: 10.47438/2309-7078_2024_1_215.

Введение

Изучение словесных ассоциаций, специфики речевых действий, репрезентированных в соотношении «стимул – реакция», особенностей ассоциативного поведения уже давно является одной из актуальных проблем современной ассоциативной лексикографии

и психолингвистики [14, с. 4]. Высказывается мнение о том, что вербальные ассоциации являются не просто парадигмой исследования речевой деятельности, но механизмом, лежащим в ее основе [20].

Для исследования ассоциаций необходимы специальные ресурсы. Существующие ассоциативные сло-

вари [4] и открытые базы данных [13] ориентированы на изучение ментального лексикона «усредненного» носителя языка, но они не позволяют исследовать индивидуальные особенности ассоциативно-вербальной сети конкретного человека.

С 2022 г. в Воронежском государственном педагогическом университете ведется работа по созданию базы данных ассоциативных реакций, включающей в себя метаразметку, в которую входит следующая информация: устойчивые характеристики респондентов, их эмоциональное состояние в момент эксперимента, а также данные об индивидуальных ассоциативных рядах [10; 11].

Кроме того, названная база данных содержит информацию об особенностях клавиатурного набора ассоциатов, а именно о длительности пауз между стимулом и каждой ассоциативной реакцией (латентный период вербального ответа), что является важным психолингвистическим маркером силы ассоциативной реакции.

Следует отметить, что важной составляющей создания базы данных является разработка и апробация классификатора типов отношений «стимул - реакция». Проблема подобного лингвистического аннотирования является предметом обсуждения многих отечественных и зарубежных лингвистов, в чьих трудах находим различные классификации [6, 22], однако сам процесс разметки, как правило, не описывается детально (не указывается число разметчиков, данные о согласованности ответов разметчиков и т.д.). Кроме того, результаты подобного аннотирования не доступны, как правило, для других пользователей.

В единичных работах подробно описывается процесс разметки ассоциативных данных с применением методов краудсорсинговой разметки и автоматического аннотирования, а также представлена ссылка на итоговый датасет [17].

Этап выделения типов отношений «стимул - реакция» является чрезвычайно значимым в ходе анализа данных ассоциативного эксперимента, а также для решения различных фундаментальных проблем языкознания и когнитивных наук.

В настоящий момент нет единой, универсальной, устоявшейся классификации, которая позволила бы объективно оценить ассоциаты и определить тип отношений, несмотря на то, что данный вопрос является предметом обсуждения уже долгое время. Наличие большого числа классификаций объясняется тем, что в их концептуальной основе лежат различные принципы.

Проведенный анализ научной литературы показал, что большинство ученых выделяют синтагматические, парадигматические и тематические ассоциации [6, 22]. На основе обширного анализа научной литературы, а также на основе данных о согласованности разметчиков по различным типам отношений, выделенным в [17], мы разработали классификатор отношений «стимул-ассоциат». В ходе исследования была поставлена задача разработать различные эвристики для объективной классификации результатов эксперимента, которая позволила бы минимизировать субъективизм разметки.

Далее мы сопоставили результаты проведенной разметки с данными о длительности пауз, а также о косинусовой близости между стимулом и реакцией в дистрибутивных языковых моделях.

Представленный датасет является, насколько нам известно, наиболее крупным из единичных представленных в открытом доступе¹ датасетом ассоциаций, содержащим несколько типов метаданных и лингвистическую разметку.

1. Описание процесса разметки. В рамках проведенного эксперимента, в котором приняли участие 49 респондентов, всего было собрано 25224 ассоциативных реакции на 50 высокочастотных слов, представляющих собой ядро русского языкового сознания. Из 49 испытуемых 10 участвовали в трехэтапном эксперименте. С деталями проведенного исследования можно ознакомиться в следующей работе [19].

Анализ полученных данных, представляющий собой интерпретацию реакций, которые дали респонденты, проходил в два этапа.

На первом этапе анализа три аннотатора, дипломированных лингвиста, независимо друг от друга составляли предложение-объяснение причин ассоциирования, которое бы содержало слово-ассоциат и слово-стимул, а также выбирали один из наиболее подходящих типов отношений из составленного заранее списка [19]. Таким образом было размечено 500 ассоциативных реакций. Однако сопоставление результатов аннотирования показало существенные расхождения как на этапе свободных пояснений, так и на этапе выделения типа отношений. При этом для ряда отношений (прежде всего парадигматических) была достигнута существенная сходимости (90 %) уже на этом этапе. Основную трудность представляло выделение отношений, которые являются контекстно-тематическими.

На следующем этапе были сформулированы модели, речевые формулы, которые упрощают и уточняют процесс отнесенности того или иного ассоциата к типу отношений. Данные модели были сформированы на основе результатов первого этапа аннотирования. Типы отношений, вызывающие наибольшие расхождения аннотаторов, были объединены в укрупненные группы. Подобная работа на втором этапе (тот же материал, что и на первом этапе; те же аннотаторы) привела к существенному повышению сходимости результатов аннотирования, которая составила 98,1%. В случае расхождений в разметке участвовал еще один независимый интерпретатор. Следует отметить, что в случаях сомнений разметчики оставляли два и более возможных варианта ответов (доля таких случаев в нашем материале не превышает 10 %) [8, с. 101]. Подобный подход является общепринятой практикой в исследованиях ассоциаций и позволяет сгладить влияние субъективных факторов при подведении итогов эксперимента [1, с. 50].

В результате двухэтапного анализа была выработана классификация, показавшая высокую сходимость результатов работы разметчиков.

¹ Датасет находится в открытом доступе:
https://github.com/Litvinova1984/word_association_dataset

На первом этапе мы выделили **2 типа отношений** (в скобках приведены их англоязычные обозначения):

I. Формальные, связанные непосредственно с формой. Они были разделены на 2 группы:

• **Фонетические (PHON)**, имеющие звуковое сходство со стимулом. Речевая модель: **ЕСТЬ ЛИ СХОДСТВО ПО ЗВУЧАНИЮ, НО НЕ ПО ЗНАЧЕНИЮ?** (*мир – пир, мир-сыр, путь-ругуть, друг-круг*).

Всего обнаружено 10 случаев фонетических реакций;

• **Морфологические (MORPH)**, связанные с морфологической общностью, куда входят реакции, имеющие общую лемму со стимулом (*смысл – иметь смысл*), общий корень (*смелый – смелость, душа – бездушный, друг-дружба*) (словообразовательные реакции). В эту же группу были отнесены формообразующие реакции, представляющие собой формы слова-стимула (*человек – люди*), а также повторы стимула (*ненавидеть – ненавидеть*). Всего в данную категорию вошли 75 случаев, при этом словообразовательные реакции составляют 95% от общего числа реакций данной группы.

Отметим малое число (относительно общего числа реакций) формальных реакций (как фонетических, так и морфологических) в нашем датасете, что говорит об успешности проведенного эксперимента с точки зрения достижения его целей и задач, поскольку свидетельствует о вовлеченности и мотивированности респондентов [9].

II. Содержательные, связанные с семантическими отношениями стимула и реакции.

В группу **содержательных** реакций вошли два основных блока:

1) Собственно языковые:

- **синтагматические** предикативные реакции, в которых испытуемый доканчивает целую, единую структуру (4232 случая от общего числа). Примерная речевая модель: **ЕСТЬ ЛИ ПРЕДИКАЦИЯ?** (в том числе идиоматическая – *душа болит*). Сюда мы отнесли собственно предикативные реакции (**PRED**), представленные 171 примером (*помощь – дорогого стоит, жизнь – одна, душа – радуется*); реакции – определения, или характеристики (**CHAR**), которые доминируют в данном сегменте классификации, – 2372 случая (*бабушка – старая, маленький – ребенок*); реакции – дополнения (**OBJ**), которые оказались также высокочастотными, – 1403 примера (*встретить – тебя, увидеть – сон*); реакции – обстоятельства (**ADC**), представленные 286 примерами (*думать – усердно, любить – сильно*).

В этом блоке разграничивались синтаксически корректные сочетания (в условных наименованиях к ним добавляется **DIR**), примеры которых приведены выше, и **семантически конвенциональный** тип синтагматических отношений (в условных наименованиях к ним добавляется **IND**), то есть такой способ связи стимула и реакции, когда синтагматические отношения не выражены грамматически корректно, но подразумеваются: *дорогой – друзья, родной – место, встретить – враг* [6];

- **парадигматические** (19675 от общего числа), которые предполагают включение слова в ту или иную категорию. Следует отметить, что они преобладают в

нашем материале и составляют 82% от общего числа реакций.

Среди парадигматических выделяются следующие отношения:

• отношения сходства в широком смысле, то есть некое тождество, смежность (2565 случаев) (**SIMIL**); речевая модель: **А ЕСТЬ В, А ПРЕДПОЛАГАЕТ В: смысл – истина, настоящий – искренний, начало – старт, история – рассказ;**

• родо-видовые отношения (**HYP**) (411 пример); речевая модель **А ВХОДИТ В В И НАОБОРОТ**; примеры: *Россия – государство, дерево – клен, семья – мама;*

• антонимические отношения (**ANT**) (245 примеров); речевая модель: **А – АНТОНИМ В: ненавидеть – любить, глупый – умный, добро – зло.**

2) Тематические:

- среди данной группы самым частотным является тип отношений «фрейм» (контекстуально-тематический тип отношений, **FRAME**), который представлен 6171 примером. Фрейм предполагает экспликацию возможного образа фрагмента действительности. Здесь стимул и реакция выступают актантами ситуации, их нельзя распределить по категориям по отношению друг к другу. Их объединяет место, время, сценарий, они являются компонентами одной ситуации. Сюда могут быть отнесены реакции, которые воспроизводят непосредственные явления сосуществования и смежность предметов, наблюдаемые испытуемым в окружающем мире. Примерные модели и примеры: 1) **А – ГДЕ?** – В: *бабушка – деревня, учиться – школа, место – билет*; 2) **А – КОГДА? В КАКОЙ ПЕРИОД?** – В: *бабушка – лето*; 3) **А ЧАСТО ИМЕЕТ В (В ТИПИЧНОЕ, РАСПРОСТРАНЕННОЕ, ХАРАКТЕРНОЕ ДЛЯ А):** *бабушка – морщины, смерть – гроб, путь – приключение*; 4) **А/В – ЭЛЕМЕНТ СИТУАЦИИ В/А:** *путь – развилка, дом – семья; думать – учеба.*

- причинно-следственные отношения (**CAUSE**) (1459 случаев):

• **А – СЛЕДСТВИЕ В/ПОЭТОМУ В** (*хотеть – стремиться, делать – успех, обещать – надеяться*); **А – ПРИЧИНА В/ ТАК КАК В** (*глупый – маленький, делать – трудолюбивый, хотеть – жажда*); **СНАЧАЛА А, ПОТОМ В / КОГДА А, ТОГДА В** (*обещать – делать, обещать – выполнить, увидеть – захотеть, увидеть – понять*);

- предметно-функциональные (**FUNC**) (1045 примеров), отражающие функциональную связь стимула и реакции. Примерные модели и примеры:

• **РОЛЬ А В В / А, ЧТОБЫ В** – или наоборот (*глаз – видеть, делать – силы, думать – решение, думать – мозг, хотеть – стимул, бог – наказывать*);

• **А – ИСТОЧНИК В / ИСТОЧНИК А – В** (*мама – жизнь, огонь – свет, огонь – тепло, связь – телефон, свет – лампочка*);

- ментально-функциональные (**MENTAL**) (5288 примеров), при которых ассоциат содержит указание на какую-либо ментальную категорию, связанную со стимулом. Примерные модели и примеры: **А ПРОЯВЛЯЕТ В** (*бабушка – забота / мама – любовь*), **А ОКАЗЫВАЕТ В** (*друг – помощь, семья – поддержка*), **А ПРИВОДИТ В СОСТОЯНИЕ В, А ГЕНЕРИРУЕТ В** (*враг – зло / мама – переживание / счастье – смех /*

дом – уют / жизнь – свет / лес – дыхание / добро – счастье / друг – взаимопомощь);

- аффективно-эмотивные (**AFFECT**) (72 примера), при которых реакции имеют эмоционально-оценочную коннотацию, в них репрезентируется отношение к предъявленному стимулу. Примеры: *война – ужас, ненавидеть – плохо!*

- реакции пояснения и уточнения (**SPECIFIC**) (2419 случаев), то есть идентифицирующие, когда стимул и реакция выражают отношения контекстной синонимии или реакция иллюстрирует слово-стимул. Примерная модель: А ДЛЯ АВТОРА – ЭТО В, А – ЭТО, НАПРИМЕР, В (*плохо – болеть, счастье – дети, хотеть – мечтать, счастье – верить, плохо – насилие, много – миллиард*);

- прецедентные (**PRECED**) (929 примеров), то есть цитатные ассоциации (перенос цитат из литературных произведений, кинофильмов, пословиц, песен и т.п.), символы – все, что предполагает знание культурного контекста в широком смысле. В эту группу

реакций входят все реакции, имеющие фольклорные, литературные, теле- и киноисточники. Примеры: *дерево – Война и мир, мир – голубь, глупый – «Цветы для Элджернона».*

III. Индивидуальные (**INDIV**) (303 случая), то есть такие ассоциации, которые не смог объяснить и отнести к какому-либо типу отношений ни один из разметчиков (*помощь – белое / много – мира / враг – черная дыра*). Подобные случаи как правило демонстрируют, что семантическая связь между стимулом и реакцией является индивидуальной.

Отметим, что в ходе интерпретации в **1295** случаях разметчики отнесли реакции к различным типам отношений. Приведем несколько примеров: *мир – новый, новый – год* (PRECED+ CHAR); *жизнь – творчество* (SPECIFIC+ FRAME); *счастье – любовь* (MENTAL+ SPECIFIC); *помощь – вера* (FUNC+ FRAME) [12].

Данные по распределению типов отношений в датасете представлены на рис. 1.

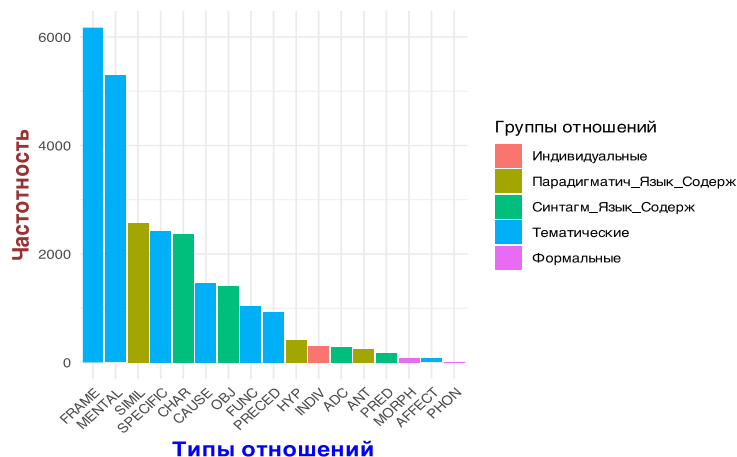


Рисунок 1 – Распределение типов отношений в датасете

2. Связь типов отношений «стимул - ассоциат» с другими типами метаданных, содержащимися в датасете.

Как указывалось выше, ассоциативный эксперимент проходил в три этапа. В первом этапе принимали участие все отобранные для исследования респонденты, в связи с чем ниже представим результаты анализа, направленного на установление связи между типами отношений и видами метаданных, содержащимися в

сформированном датасете, на примере данных, собранных на первом этапе.

Общее число реакций по первому этапу составило 14700. Для дальнейшего анализа мы отобрали наиболее частотные из них (представленные свыше чем 50 примерами; всего 12949 реакции). Распределение отношений «стимул - реакция» по типам для первого этапа эксперимента представлено на рис. 2.

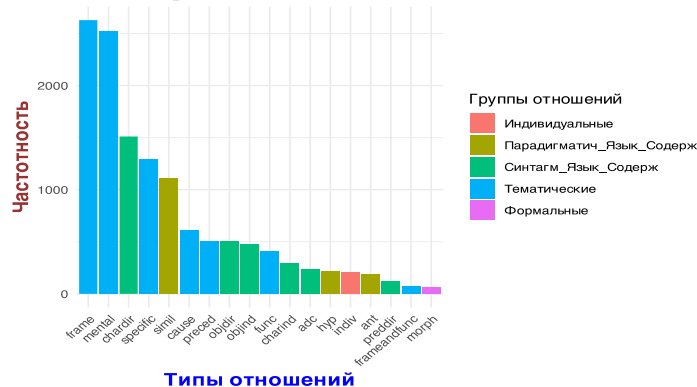


Рисунок 2 – Распределение типов отношений в датасете по первому этапу эксперимента

Как мы видим, распределение типов отношений в целом по корпусу и по первому этапу оказывается схожим.

Наш датасет содержит данные о времени реакции, то есть и времени паузы между стимулом и первым ассоциатом, первым и вторым ассоциатом и так далее (в миллисекундах). После удаления выбросов было установлено, что среднее значение времени реакции составляет 2512 мс, медианное – 1830. Как свидетельствует анализ литературы, данное значение соответствует когнитивным паузам [21].

Тест Крускала-Уоллиса показал наличие значимых различий между типами отношений по времени

реакции ($p = 0.0001765$, Kruskal-Wallis chi-squared = 32.302, $df = 9$). Чтобы определить, между какими именно группами присутствуют различия, мы использовали Т-критерий Уилкоксона. Названный тест показал наличие различий по времени реакции между такими типами отношений, как MENTAL и CHARDIR ($p=0.016$), MENTAL и OBJIND ($p=0.016$), FRAME и OBJIND ($p=0.016$), FRAME и PRECED ($p=0.040$), FRAME и SIMIL ($p=0.020$), MENTAL и SIMIL ($p=0.016$), MENTAL и PRECED ($p=0.026$).

На рис. 3 представлены данные о времени реакции на стимулы по типам отношений.

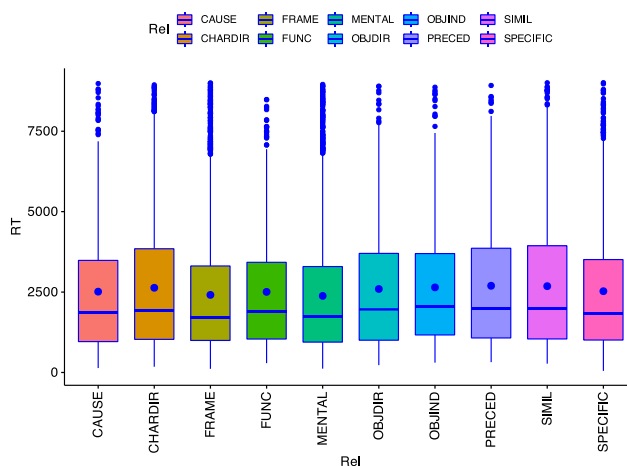


Рисунок 3 – Данные о времени реакции на стимулы для выделенных типов отношений

Следующим этапом нашего анализа явилось исследование различий типов отношений по значению косинусовой близости (мера семантической близости) в дистрибутивных семантических моделях, обученных на многомиллионных корпусах текстов: чем больше косинусовая (семантическая) близость между векторами слов, тем чаще слова встречаются в схожих контекстах. Подобные исследования являются чрезвычайно важными для понимания механизма ассоциирования. Как известно, единого мнения среди ученых по данному вопросу до сих пор не выработано, однако совместная встречаемость слов рассматривается как один из важных факторов, лежащих в основе ассоциирования [20]. Мы рассматривали две языковые модели разных типов (ruwikiruscorpora_upos_cbow_300_10_2021 и

symanto/sn-xlm-roberta-base-snli-mnli-anli-xnli). Данные о косинусовой близости между стимулом и реакцией, извлеченные из двух моделей, также содержатся в сформированном датасете.

В модели ruwikiruscorpora_upos_cbow_300_10_2021 отсутствовали 474 реакции, что составляет менее 5 % от общего числа реакций. Среднее значение семантической близости составило 0.2101, медианное - 0.2282.

Тест Крускала-Уоллиса Kruskal-Wallis rank sum test показал связь между типом отношений и косинусовым расстоянием (Kruskal-Wallis chi-squared = 1328.6, $df = 9$, $p\text{-value} < 2.2e-16$).

Попарное сравнение, проведенное с использованием Т-критерия Уилкоксона (рис. 4), показало наличие различий среди большинства пар, кроме SPECIFIC и CAUSE, OBJIND и CHARDIR, OBJIND и OBJDIR, FUNC и CAUSE, FUNC и FRAME, FRAME и CAUSE.

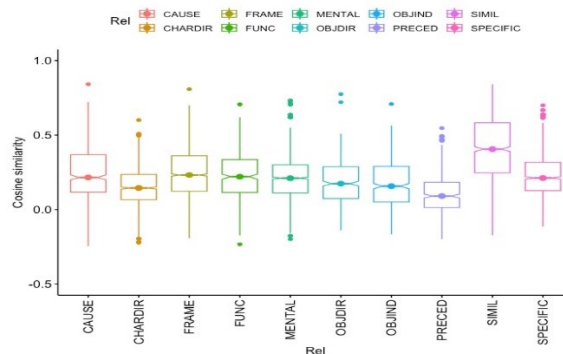


Рисунок 4 – Косинусовая близость между стимулом и реакцией для разных типов отношений (модель ruwikiruscorpora_upos_cbow_300_10_2021)

Мы повторили подобный анализ для другой модели - xlm-roberta-base-snl1-mnli-anli-xnli. Показатели семантической близости между стимулом и реакцией по данной модели в среднем выше, чем в первой (среднее значение - 0.46176, медианное - 0.44854).

Тест Крускала показывает наличие статистически значимых различий между группами (Kruskal-Wallis

chi-squared = 2937.7, df = 9, p-value < 2.2e-16). Тест парных различий показал наличие различий между всеми парами типов отношений, кроме CAUSE и FUNC, FUNC и FRAME, FRAME и CAUSE (рис. 5), что схоже с результатами, полученными на первой модели.

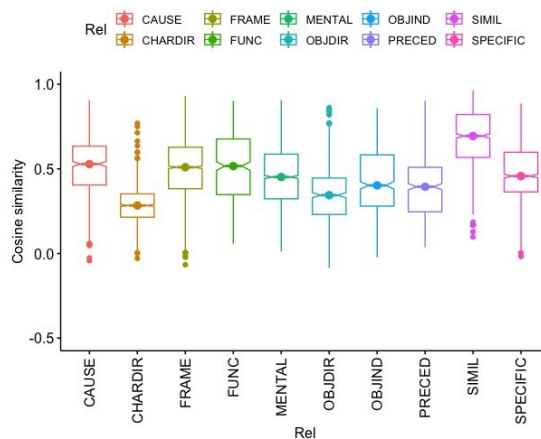


Рисунок 5 – Косинусовая близость между стимулом и реакцией для разных типов отношений (модель ruwikiruscorpora_upos_cbow_300_10_2021)

Указанное наблюдение может свидетельствовать о том, что выделение причинно-функциональных типов является факультативным; аннотированные подобным образом пары «стимул – реакция», с точки зрения дистрибутивных языковых моделей, могут быть отнесены к типу «Фрейм».

В целом полученные данные о различиях по семантической близости свидетельствуют в пользу выработанной классификации типов отношений «стимул – реакция».

Выводы

Ассоциативный эксперимент, несмотря на долгую историю использования, продолжает оставаться эффективным исследовательским инструментом, применяемым в различных областях знаний (см. результаты метаанализа научной литературы, посвященной вербальным ассоциациям, выполненной с использованием структурного тематического моделирования, в нашей работе [19]). Прогресс в области исследований вербальных ассоциаций, в связи с трудоемкостью проведения соответствующих экспериментов и аннотирования полученных данных, во многом связан с разработкой соответствующих открытых датасетов с различными слоями разметки, в том числе лингвистической. В работе представлено описание процесса аннотирования ассоциативных данных по типу отношений «стимул – реакция» профессиональными лингвистами, а также результаты

исследования связи выделенных типов отношений с временем реакции респондента и семантической близостью по данным языковых моделей разных типов. Проведенное исследование подтвердило правомерность выделения названных типов отношений.

Создаваемая база данных, содержащая различные слои разметки, не имеет аналогов в мировой лингвистике. Наличие большого типа метаданных позволит исследователям эффективно решать различные актуальные задачи, связанные с анализом ассоциативного материала, в частности, сопоставить типы ассоциативных отношений и время реакции в зависимости от постоянных и темпоральных характеристик респондентов; установить наиболее типичные для того или иного респондента типы ассоциативных отношений; рассмотреть связь семантического состояния, извлеченного из дистрибутивных моделей, и типов отношений – с одной стороны и времени реакции – с другой, а также многие другие задачи.

Источник финансирования

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

Конфликт интересов

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

Библиографический список

1. Береснева Н.И. Модель внутреннего лексикона в позднем онтогенезе (ассоциативный эксперимент) : дисс. ... канд. филол. наук. Пермь, 1997. С. 50.
2. Борисова Ю.А. Ассоциативный эксперимент в современных психолингвистических исследованиях // Психология. Историко-критические обзоры и современные исследования. 2019. Т. 8, № 1А. С. 265–275.
3. Бутакова Л.О. Слова с ценностной семантикой в языковом сознании представителей поколения Y // Вестник РУДН. Серия: Теория языка. Семиотика. Семантика. 2022. Т. 13, № 4. С. 937–955.

4. Возрастной дифференциальный психолингвистический словарь русского языка / науч. ред.: И.А. Стернин, А.В. Рудакова. Воронеж : Центр коммуникативных исследований ВГУ, 2020. 159 с.
5. Григорьева Е.И. Роль ассоциативного онлайн-словаря в современных психолингвистических исследованиях // Информационные технологии в науке, бизнесе и образовании. проблемы обеспечения цифрового суверенитета государства : материалы XII Междунар. науч.-практ. конф. студентов, аспирантов и молодых ученых. М. : МГУ, 2021. С. 235–239.
6. Горошко Е.И. Интегративная модель свободного ассоциативного эксперимента. М. : РА–Каравелла, 2001. URL: <https://www.textology.ru/razdel.aspx?ID=38> (дата обращения: 17.02.2023).
7. Караулов Ю.Ю., Коробова М.М. Индивидуальный ассоциативный словарь // Вестник языкознания. 1993. № 5. С. 5–15. URL: <https://vja.ruslang.ru/ru/archive/1993-5/5-15> (дата обращения: 27.01.2023).
8. Караулов Ю.Н. Русский язык и языковая личность / отв. ред. Д.Н. Шмелев. М. : Наука, 1987. С. 101.
9. Курганова Н.И. Ассоциативный эксперимент как метод исследования значения живого слова // Вопросы психолингвистики. 2019. № 3 (41). С. 24–37.
10. Литвинова Т.А., Котлярова Е.С., Заварзина В.А. Фактор гендера в ассоциативных связях слов: данные словаря и дистрибутивно-семантической модели // Научный диалог. 2022а. Т. 11, № 5. С. 136–156.
11. Литвинова Т.А., Заварзина В.А., Любова С.Г. База данных словесных ассоциаций, содержащая информацию о клавиатурном поведении респондентов // Известия Воронежского государственного педагогического университета. 2022б. № 4 (297). С. 240–249.
12. Литвинова Т.А., Любова С.Г. Разметка ассоциативных реакций по типам отношений «стимул – ассоциат» как этап создания аннотированного многокомпонентного корпуса ассоциативных реакций индивида // Проблема изучения живого русского слова на рубеже тысячелетий : материалы XI Междунар. науч. конф. Воронеж : ВГПУ, 2023. С. 49–58.
13. Русская региональная ассоциативная база данных (СИБАС): сайт: URL: <http://adictu.nsu.ru/stimlist> (дата обращения: 27.01.2023).
14. Степыкин Н.И. Сопоставительный анализ ассоциативных полей вежливость и *рущиршцщрпнрпнн* // Известия Юго-Западного государственного университета. Серия: Лингвистика и педагогика. 2022. 12 (1). С. 4.
15. Пищальникова В.А. Интерпретация ассоциативных данных как проблема методологии психолингвистики // Russian Journal of Linguistics. 2019. Т. 23, № 3. С. 749–761. DOI: 10.22363/2312-9182-2019-23-3-749-761 (дата обращения: 17.02.2023).
16. Уфимцева Н.В. Языковое сознание и образ мира славян // Языковое сознание и образ мира : сборник статей / отв. ред. Н.В. Уфимцева. М., 2000. 320 с. URL: https://iling-ran.ru/library/psylingva/sborniki/Book2000/html_204/4-2.html (дата обращения: 19.12.2022).
17. Chunhua Liu, Trevor Cohn, Simon De Deyne, Lea Frermann. A new dataset for word association explanations // School of Computing and Information Systems Melbourne School of Psychological Sciences The University of Melbourne. Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), November 20–23. 2022. P. 106–120.
18. Minto-Garcna A., Arias-Trejo N. & Vargas-Garcna E.M. Lexical Relations in Spanish-Speaking Older Adults // J Psycholinguist Res. 2020. № 49. P. 663–716 DOI: 10.1007/s10936-020-09708-5.
19. Tatiana A. Litvinova. Mapping the field of word association research using text mining approach / Tatiana A. Litvinova, Viktoriya A. Zavarzina, Elena S. Kotlyarova, Svetlana G. Lyubova // Proceedings of the 2023 5th International Conference on Information Technology and Computer Communications. June 2023. P. 90–98. URL: <https://dl.acm.org/doi/10.1145/3606843.3606858>. (дата обращения: 11.08.2023).
20. Planchuelo C., Buades-Sitjar F., Hinojosa J. A., Ducabeitia and J. A. The Nature of Word Associations in Sentence Contexts // Experimental Psychology. 2022. Vol. 69, iss. 2. DOI: 10.1027/1618-3169/a000547
21. Wengelin, E., Johansson, V. (2023). Investigating Writing Processes with Keystroke Logging. // O. Kruse et al. Digital Writing Technologies in Higher Education. Springer, Cham. DOI: 10.1007/978-3-031-36033-6_25
22. Wulff D.U., Aeschbach S., De Deyn S., & Mata R. Data From the MySWOW Proof-of-Concept Study: Linking Individual Semantic Networks and Cognitive Performance // Journal of Open Psychology Data. 2022. 10: 5. P. 1–8. DOI: <https://doi.org/10.5334/jopd.55>.

References

1. Beresneva, N.I. (1997) *Model' vnutrennego leksikona v pozdnem ontogeneze (assotsiativnyi eksperiment)*. Diss. kand. filol. nauk [The model of the internal lexicon in late ontogenesis (associative experiment). Candidate philol. sci diss.]. Perm', p. 50. (In Russian)
2. Borisova, Yu.A. (2019) *Assotsiativnyi eksperiment v sovremennykh psikholingvisticheskikh issledovaniyakh* [Associative experiment in modern psycholinguistic research]. *Psikhologiya. Istoriko-kriticheskie obzory i sovremennye issledovaniya*. 8(1A), 265–275. (In Russian)
3. Butakova, L.O. (2022) *Slova s tsennostnoi semantikoi v yazykovom soznanii predstavitelei pokoleniya Y* [Words of understandable semantics in the linguistic consciousness of the president]. *Vestnik RUDN. Seriya: Teoriya yazyka. Semiotika. Semantika*. 13 (4), 937–955. (In Russian)

4. Sternin, I.A., Rudakova, A.V. (ed.) (2020) *Vozrastnoi differentsial'nyi psikholingvisticheski slovar' russkogo yazyka* [Age-related differential psycholinguistic dictionary of the Russian language]. Voronezh, Tsentr kommunikativnykh issledovaniy VGU publ. 159 p. (In Russian)
5. Grigor'eva, E.I. (2021) The role of an associative online dictionary in modern psycholinguistic research. In: *Information technologies in science, business and education. problems of ensuring the digital sovereignty of the state : materials of the XII International Scientific and Practical Conference of Students, postgraduates and young Scientists*. Moscow, MGU, pp. 235–239. (In Russian)
6. Goroshko, E.I. (2001) *Integrativnaya model' svobodnogo assotsiativnogo eksperimenta* [Integrative model of free associative experiment]. Moscow, RA–Karavella publ. Available from: <https://www.textology.ru/razdel.aspx?ID=38> (accessed: 17.02.2023). (In Russian)
7. Karaulov, Yu.Yu., Korobova, M.M. (1993) Individual'nyi assotsiativnyi slovar' [Individual associative dictionary]. *Vestnik yazykoznaniiya*. (5), 5–15. Available from: <https://vja.ruslang.ru/ru/archive/1993-5/5-15> (accessed: 27.01.2023). (In Russian)
8. Karaulov, Yu.N. (1987) *Russkii yazyk i yazykovaya lichnost'* [Russian language and linguistic personality]. Moscow, Nauka publ., p. 101. (In Russian)
9. Kurganova, N.I. (2019) Assotsiativnyi eksperiment kak metod issledovaniya znacheniya zhivogo slova [Associative experiment as a method of studying the meaning of a living word]. *Voprosy psikholingvistiki*. 3 (41), 24–37. (In Russian)
10. Litvinova, T.A., Kotlyarova, E.S., Zavarzina, V.A. (2022) Faktor gendera v assotsiativnykh svyazyakh slov: dannye slovary i distributivno-semanticheskoi modeli [Gender factor in associative relations of words: dictionary data and distributive semantic model]. *Nauchnyi dialog*. 11 (5), 136–156. (In Russian)
11. Litvinova, T.A., Zavarzina, V.A., Lyubova, S.G. (2022) Baza dannykh slovesnykh assotsiatsii, sodержashchaya informatsiyu o klaviaturnom povedenii respondentov [A database of verbal associations containing information about the keyboard behavior of respondents]. *Izvestiya Voronezhskogo gosudarstvennogo pedagogicheskogo universiteta*. 4 (297), 240–249. (In Russian)
12. Litvinova, T.A., Lyubova, S.G. (2023) Razmetka assotsiativnykh reaktzii po tipam otnoshenii «stimul – assotsiat» kak etap Marking of associative reactions by types of "stimulus – associate" relations as a stage in creating an annotated corpus of multicomponent associative reactions of an individual. In: *The problem of studying the living Russian word at the turn of the millennium : materials of the XI International Scientific Conference*. Voronezh, VGPU, pp. 49–58. (In Russian)
13. *Russkaya regional'naya assotsiativnaya baza dannykh (SIBAS)* [Russian regional database (SIBAS)]. Available from: <http://adict.ru/nsu.ru/stimlist> (accessed: 27.01.2023). (In Russian)
14. Stepykin, N.I. (2022) Sopostavitel'nyi analiz assotsiativnykh polei vezhливost' i politenosti [Comparative analysis of the associative fields politeness and вежливости]. *Izvestiya Yugo-Zapadnogo gosudarstvennogo universiteta. Seriya: Lingvistika i pedagogika*. 12 (1), 4. (In Russian)
15. Pishchal'nikova, V.A. (2019) Interpretatsiya assotsiativnykh dannykh kak problema metodologii psikholingvistiki [Interpretation of associative data as a problem of methodology of psycholinguistics]. *Russian Journal of Linguistics*. 23 (3), 749–761. Available from: doi:10.22363/2312-9182-2019-23-3-749-761 (accessed: 17.02.2023). (In Russian)
16. Ufimtseva, N.V. (2000) Linguistic consciousness and the image of the world of the Slavs. In: *Linguistic consciousness and the image of the world : a collection of articles*. Moscow. 320 s. Available from: https://ilingran.ru/library/psylingva/sborniki/Book2000/html_204/4-2.html (accessed: 19.12.2022). (In Russian)
17. Chunhua Liu, Trevor Cohn, Simon De Deyne, Lea Frermann (2022). A new dataset for word association explanations. In: *School of Computing and Information Systems Melbourne School of Psychological Sciences The University of Melbourne. Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing* (Volume 1: Long Papers), November 20–23, pp. 106–120. (In English)
18. Minto-Garcia A., Arias-Trejo N. & Vargas-Garcia E.M. (2020) Lexical Relations in Spanish-Speaking Older Adults // *J Psycholinguist Res*. (49), 663–716. Available from: doi:10.1007/s10936-020-09708-5. (In English)
19. Tatiana A. Litvinova. Mapping the field of word association research using text mining approach / Tatiana A. Litvinova, Viktoriya A. Zavarzina, Elena S. Kotlyarova, Svetlana G. Lyubova // *Proceedings of the 2023 5th International Conference on Information Technology and Computer Communications*. June 2023. P. 90–98. Available from: <https://dl.acm.org/doi/10.1145/3606843.3606858>. (дата обращения: 11.08.2023). (In English)
20. Planchuelo C., Buades-Sitjar F., Hinojosa J. A., Ducabeitia and J. A. The Nature of Word Associations in Sentence Contexts // *Experimental Psychology*. 2022. Vol. 69, iss. 2. Available from: doi:10.1027/1618-3169/a000547 (In English)
21. Wengelin, E., Johansson, V. (2023). Investigating Writing Processes with Keystroke Logging. // O. Kruse et al. *Digital Writing Technologies in Higher Education*. Springer, Cham. Available from: doi:10.1007/978-3-031-36033-6_25 (In English)
22. Wulff D.U., Aeschbach S., De Deyn S., & Mata R. Data From the MySWOW Proof-of-Concept Study: Linking Individual Semantic Networks and Cognitive Performance // *Journal of Open Psychology Data*. 2022. 10: 5. P. 1–8. DOI: <https://doi.org/10.5334/jopd.55>. (In English)

Поступила в редакцию 30.01.2024

Подписана в печать 28.03.2024

Original article

UDC 81'33

DOI: 10.47438/2309-7078_2024_1_215

**AN OPEN MULTICOMPONENT DATABASE OF VERBAL ASSOCIATIONS,
ANNOTATED BY TYPES OF INCENTIVE–ASSOCIATE RELATIONSHIPS
AND REACTION TIME: RELATIONSHIP OF METADATA CHARACTERISTICS**

Victoria A. Zavarzina¹, Elena S. Kotlyarova², Svetlana G. Lyubova³

*Voronezh State Pedagogical University^{1, 2, 3}
Voronezh, Russia*

¹*Senior Lecturer of the Department of Theory,
History and Methods of Teaching Russian Language and Literature,
Laboratory Assistant of the Research Laboratory of Computer Semasiology,
ORCID ID: 0000-0001-6484-4606, e-mail: zva0604@yandex.ru*

²*Postgraduate Student, Senior Laboratory Assistant of the Research Laboratory of Computer Semasiology,
ORCID ID: 0000-0002-4162-5002, e-mail: kray-voronezh@mail.ru*

³*Cand. Philol. Sci., Senior Lecturer of the Department of Theory, History
and Methods of Teaching Russian Language and Literature,
ORCID ID: 0000-0002-2586-410X, e-mail: svetlana_lyubova_2305@mail.ru*

Abstract. The associative experiment is one of the most effective psycholinguistic methods of modeling the meaning of a word in an individual linguistic consciousness. Due to the complexity of conducting associative experiments and the importance of the empirical data obtained for solving various linguistic problems, and therefore, for a wide range of researchers, publicly available corpora containing the results of such experiments are of particular importance. A number of such resources have been created and maintained by domestic and foreign researchers, but they usually do not contain information about individual associative behavior, as well as the results of linguistic annotation. This article presents a description of the association database created at Voronezh State Pedagogical University with the support of a Russian Scientific Fund grant and which has no analogues in world linguistics, containing metadata of various types, the process and results of annotating associative data by types of "stimulus-reaction" relations according to the classification developed by the authors, as well as information about the relationship of various types of metadata.

Key words: psycholinguistic meaning of a word, associative experiment, associative reaction, linguistic markup, corpus of texts.

Cite as: Zavarzina, V.A., Kotlyarova, E.S., Lyubova, S.G. (2024) An open multicomponent database of verbal associations, annotated by types of incentive–associate relationships and reaction time: relationship of metadata characteristics. *Izvestia Voronezh State Pedagogical University*. (1), 215–223. (In Russ., abstract in Eng.). DOI: 10.47438/2309-7078_2024_1_215.

Received 30.01.2024

Accepted 28.03.2024