

**ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА
ДАНЫХ ДЛЯ СВОБОДНОГО АССОЦИАТИВНОГО ЭКСПЕРИМЕНТА
(НА ПРИМЕРЕ ЦЕННОСТИ *ВОЛЯ*)**

Казаченко Оксана Васильевна,

Московский городской педагогический университет,
Финансовый университет при Правительстве Российской Федерации,
Москва, Россия

Шмалько Светлана Петровна,

Кубанский государственный университет,
Краснодар, Россия

Дорошенко Ольга Валерьевна,

Кубанский государственный университет,
Краснодар, Россия

Аннотация

Фокус современных как лингвистических, так и методических исследований сместился в направлении изучения человека как развивающейся языковой личности, что органически встраивается в антропоцентрическую и аксиологическую парадигмы развития научной мысли XXI века. Анализ статистически обработанных ассоциативных полей носителей русского языка необходим для рассмотрения изменений в языковом сознании личности, для понимания современной системы личностных ценностей, что может способствовать ценностно-ориентированному воспитанию подрастающих поколений носителей языка и культуры.

В качестве материала исследования выступили результаты свободного ассоциативного эксперимента на большой выборке испытуемых, проживающих в Российской Федерации. Подобный широко используемый эксперимент позволяет исследователю выделить ядро изучаемого стимула, характеризующего общепринятые ассоциации испытуемых, и периферию, указывающую на индивидуально-личностные ассоциации. При помощи статистических методов обработки массивов данных авторами предпринята попытка выделения главных ассоциатов, определяющих понимание изучаемой ценности *воля* в форме концептов.

Ко всем полученным реакциям на ценность *воля* был применен интеллектуальный анализ текстовых данных или Text Mining (TM) для определения оптимального числа концептов. С его помощью было осуществлено сингулярное разложение матрицы, заключающееся в нахождении SVD-коэффициентов. Важна интерпретация положительных и отрицательных знаков этих коэффициентов, значимых с точки зрения смысла ценности. Так, один из концептов ценности *воля*, получивший два наибольших коэффициента с противоположными знаками, интерпретируется нами как *сила*, но с противоположным знаком к *свободе*, т.е. *воля* как *безволие* / не желание действий. Далее авторами был использован метод латентно-семантического индексирования для кластеризации ассоциатов, который позволил описать восприятие ценности-стимула *воля* различными группами респондентов.

Статистические методы все чаще используют в прикладных филологических исследованиях, но содержательно интерпретировать числовые (положительные и отрицательные) результаты для анализа языковых единиц достаточно сложно. В нашем исследовании дается пример подобной интерпретации.

Ключевые слова: ассоциативный эксперимент, воля, интеллектуальный анализ данных, концепты, SVD-коэффициенты, ценности

Введение

Современные лингвистические исследования согласуются с общей научной антропологической парадигмой, изучающей человека не как объект, а как субъекта деятельности. Основы этого подхода были заложены в XX веке Л.С. Выготским в его теории речевой деятельности. Важно подчеркнуть, что развитие, обучение человека (безусловно, с разной степенью интенсивности) не прекращается всю его жизнь. Особенно это актуально для ранних периодов развития ребенка, поскольку обучение представляет собой внутренне необходимый и всеобщий момент в процессе развития у ребенка исторических особенностей человека, вызывающий к жизни ряд таких процессов, которые без него возникнуть не могут [Выготский 2001]. Те же самые тенденции определяют и развитие речемыслительных процессов.

При активном участии индивида, то есть в процессе деятельности, в сознании человека возникают представления об окружающем его мире (физическом, биологическом, социальном), объединенные при помощи различных связей в целостный образ мира, называемый лингвистами языковой картиной мира, языковым сознанием. Эти связи и отношения на сегодняшний день экспериментально выявляются посредством методики ассоциативного эксперимента, что позволило создать не только ряд ассоциативных словарей, но и проводить работу над тезаурусами различных языков.

Однако сложность языка как живого средства общения состоит в том, что любая фиксация языка в форме слов отражает его прошлое состояние, поскольку речь, а значит и язык, изменяется постоянно. Таким образом, даже зафиксировав какое-то состояние языка, мы можем судить только об определенном отрезке его существования и развития. Тем не менее, изучение динамики содержания слов в языковой картине мира современного человека важно для прогнозирования тенденций ее развития, ее культурной составляющей, а также для понимания степени значимости определенных понятий, номинированных словами языка. Понимание типов связей и отношений позволяет строить предположения о сходных процессах при изучении иностранных языков. Важно подчеркнуть, что для улучшения освоения иностранного языка необходимо изучить пути и тенденции формирования ассоциаций обучающихся на родном языке и их влияние на изучаемые языки. Это особенно существенно при исследовании ценностей, поскольку они вербализованы в речи в форме абстрактных существительных, представления о которых в полной мере складываются в позднем подростковом возрасте и являются моделью поведения личности.

Ценности как особый объект исследования были выделены в философии в 60-х годах XIX века. Это способствовало появлению в философской науке нового направления – аксиологии, определявшего ценности как значимости предмета или явления в противовес их онтологии или качественным характеристикам.

Проблема субъективности и объективности ценностей стояла для ученых с давних времен и, собственно, пока далека от своего решения и в настоящее время. Личность представляет собой источник ценностей. Важно то, что, будучи абстрактными сущностями, ценности существуют только в человеке, и только человек наделяет предметы или феномены значимостью. Убеждения и ценности лежат в основе оценки всех явлений человеческой жизни, что «придает оттенок субъективности всему, созданному человеком» [Федосюткина 2005: 4]. Эту же мысль подтверждает и М. Рокич, рассматривая ценности как положительные или отрицательные «абстрактные идеи», не связанные с определенным объектом или ситуацией, но содержащие убеждения личности о приемлемом поведении и предпочтительных целях [Rokeach 1973: 3].

Ценности локализируются в сознании и принадлежат двум уровням его структуры: уровню значения (знака) и уровню смысла, поскольку они обусловлены «эмоционально-переживаемым, потребностно-пристрастным отношением субъекта» [Каширский 2014]; таким образом, их исследование возможно осуществить при помощи изучения значений слов.

Кроме того, ценности неразрывно связаны с культурой этноса, в которой они функционируют. Культура обуславливает не только набор, но и иерархию ценностей, выделенных и выверенных в течение истории конкретного этноса. Это позволяет ценностям воспроизводиться в каждом конкретном человеке, что в совокупности обеспечивает выживание нации. Интересным представляется исследование К. Ситорама и Р. Когделла, в котором на основе анализа различных культур ценности распределены на первичные, вторичные, третичные и несущественные [Ситорам, Когделл 1992]. Разделив все культуры на несколько групп, включающих западные, восточные, американские и мусульманские, авторы ранжировали избранные ими ценности (материнство, спасение, мужественность и другие) и выяснили, что некоторые ценности более значимы для одних культур и менее значимы для других. Так, например, деньги как ценность более значимы в рамках западно-американской культуры, в то время как их релевантность для мусульманской культуры крайне низка.

Таким образом, поскольку ценность совмещает в себе разнопорядковые сущности и связана со многими феноменами человеческой жизни, то понятию ценности довольно сложно дать исчерпывающее определение. Ценности имеют множество дефиниций и смыслов, а потому эти определения в основном лишь указывают на область научного знания. Изучение ценностей довольно затруднительно, что создает дополнительную проблему выбора методики исследования.

Обзор литературы

Представляется, что анализ сущности ценностей необходимо проводить, используя междисциплинарные исследования, в том числе с лингвистической, психолингвистической и математической точек зрения, что было доказано в пилотном исследовании И.А. Бубновой [Бубнова 2021]. Поскольку ценности выражены вербально, они номинируют некую осознанную часть действительности, которая в языковом сознании человека представлена сложными образами.

Проблемой психолингвистических исследований является сложность обработки и последующей интерпретации полученных данных. Среди основных подходов различают качественный (квалитативный) и количественный (квантитативный) анализ данных лингвистических экспериментов; каждый из этих видов анализа обладает

своими преимуществами и недостатками [Rasinger 2008: 9]. Однако современная лингвистика уже на протяжении десятилетий все чаще оперирует количественными данными и реализует квантитативные методы при помощи специальных компьютерных программ [Johnson 2008, Алиакберова 2021]. Язык программирования R для статистической обработки данных с открытым исходным кодом выступает сильным инструментом для анализа текстов за счет широкого спектра специализированных пакетов [Ивин 2018; Kwartler 2017; Feinerer 2022].

Тестовые данные являются неструктурированным типом данных, для которых разработана технология интеллектуального анализа данных для преобразования их в пригодный для компьютерного анализа структурированный вид [Albright 2004: 13]. Выделяются следующие типичные задачи интеллектуального анализа текста или Text Mining (далее – ТМ): категоризация, кластеризация, извлечение концепта / сущности, создание детализированных таксономий, анализ настроений, обобщение документов и моделирование отношений элементов [Han et al. 2012]. Кроме того, существует латентно-семантическое индексирование или Latent Semantic Indexing (далее LSI), которое с помощью методов линейной алгебры пытается уловить скрытую структуру в данных [Papadimitriou et al. 1998].

Основными понятиями в ТМ и LSI являются «документы», «термины» и «концепты». Весь массив экспериментальных лингвистических данных рассматривается как коллекция документов, которыми могут быть фразы, предложения, абзац или просто набор слов. В зависимости от токенизации термин может быть просто словом, но также может относиться к таким элементам, как пунктуация, фразы или другие элементы, состоящие из нескольких слов. Концепты являются совокупностью терминов, взятых с определенным весом. В зависимости от целей исследования либо документы представляются наблюдениями, а термины – переменными, либо наоборот. Реализуются ТМ и LSI через сингулярное разложение матрицы или Singular Value Decomposition (SVD) терминов-документов или документов-терминов [Albright 2004; Papadimitriou et al. 1998]. ТМ использует метод SVD для обнаружения шаблонов в коллекции документов, а LSI, с другой стороны, предназначен для возврата подмножества документов, связанных с конкретным запросом.

Материалы и методы

Традиционно при эмпирическом подходе используется несколько методов или их групп в исследовании ценностей, таких как метод ассоциаций, метод ранжирования, методы субъективного шкалирования и парного сравнения, а также проективные методики. Метод ассоциаций заключается в том, что респондентам предлагается слово-стимул и они отвечают, не задумываясь, первыми пришедшими в голову ассоциациями. Ранжирование позволяет построить модель личных ценностных предпочтений. В проективных методиках испытуемых помещают в заданные экспериментатором проблемные ситуации, и они реагируют выбором типичного для них поведения. На основе такого выбора делаются выводы о ценностных приоритетах личности.

Довольно часто в качестве отправной точки для показа динамики ассоциативных полей, а значит и изменения языкового сознания используются различные ассоциативные словари. Так, например, наиболее частотными реакциями на стимул воля, согласно РАС, являются следующие: *вольная* 9; *свобода, сила, сильная* 8;

неволя 6; железная 5; ваша 4; Божья, доля, к победе, моя 3; земля, к жизни, твоя 2 [РАС 2002: 107].

Указание на ассоциативный эксперимент как на наиболее объективный метод выявления культурной спецификации лексики встречаем в работах А.А. Леонтьева [Леонтьев 1977: 14]. Этот метод в прошлом веке был усовершенствован для психолингвистических исследований А.Р. Лурией и О.С. Виноградовой и до сих пор остается успешным экспериментальным методом в психологических, социологических и лингвистических исследованиях. Например, метод ассоциативного эксперимента показал свою состоятельность в качестве эффективного психолингвистического метода для определения языковой картины мира [Андреева, Усейнова 2021]. Ассоциативный эксперимент применялся также в социопсихолингвистическом исследовании для анализа концептуальных образов, проявляющихся в языковом сознании [Богданович 2018: 10–16]. В работе [Levchenko et al. 2020] авторы визуализируют результаты ассоциативного теста и строят сеть словесных ассоциаций, которая является идеографическим словарем и тезаурусом. Поэтому в качестве рабочей методики в исследовании сущности ценностей был выбран ассоциативный эксперимент с последующей обработкой результатов статистическими методами.

Рассмотрим подробнее SVD как метод ТМ [Albright 2004]. Пусть в множестве данных имеется n документов, каждый из которых состоит не более чем из m терминов. Тогда каждый документ можно представить вектором, координаты этих векторов соответствуют терминам, а их значения равны числу вхождения терминов в документ. Другими словами, для любого документа в векторном представлении i -я координата является частотой i -го термина из рассматриваемых m терминов. Вектор для каждого документа, как правило, очень разрежен (содержит большую долю нулей), потому что немногие из терминов в коллекции в целом содержатся в каком-либо отдельно взятом документе. В результате получается разреженная матрица «термины–документы» A_{mn} размера $m \times n$ в которой термины представляют строки, а документы – столбцы. Для любой матрицы существует единственное разложение вида 1:

$$A_{mn} = U_{mm} \Sigma_{mm} V_{mn}, \quad (1)$$

где матрица Σ_{mm} – диагональная матрица сингулярных значений матрицы A_{mn} , записанных по убыванию. Тогда матрицу A_{mn} можно аппроксимировать с различной степенью точности, выбрав любое $r \leq m$ и урезав матрицу Σ до ранга r , то есть справедлива формула 2:

$$A_{mn} \approx U_{mr} \Sigma_{rr} V_{rn}, \quad (2)$$

Здесь матрица U_{mr} состоит из m терминов и r концептов, матрица V_{rn} состоит из r концептов и n документов, а элементы матрицы Σ_{rr} определяют «силу» соответствующего концепта. Обычно значение r выбирается намного меньше, чем m . Конечной целью применения SVD является размещение документов (или терминов) в пространстве гораздо меньшей размерности, чем определено пространством терминов (или пространством документа). Ортонормированные матрицы U_{mr} и V_{rn} образуют m -мерные ортогональные базисы пространств терминов или документов соответственно.

Любой документ d_m , представленный как m -мерный вектор терминов и состоящий из частот полной коллекции терминов в документе, может быть спроецирован на подпространство терминов более низкой размерности с помощью формулы 3:

$$\hat{d}_r = U_{mr}^T d_m \quad (3)$$

Таким образом, – это r -мерный вектор, i -й элемент которого формируется путем взятия линейной комбинации частот исходных терминов с i -м сингулярным вектором (столбцом) U_{mr} . Записи в столбцах U_{mr} можно рассматривать как веса, применяемые к отдельным терминам в коллекции. Заметим, что термины также можно проецировать в уменьшенное пространство. При последующей нормализации множества векторов $\{\hat{d}_r\}$ евклидово расстояние можно использовать как меру близости между документами.

Выбор количества используемых измерений может r быть важным аспектом многих решений интеллектуального анализа текста. При слишком малом количестве измерений модель не сможет объяснить заметные взаимосвязи в тексте. С другой стороны, использование слишком большого количества измерений добавит в модель ненужного шума и сделает построение эффективной модели практически невозможным. В некоторых случаях можно сделать выбор, основываясь на графике кумулятивных сингулярных значений, который строится на основе критерия информативности (4):

$$I(r) = \frac{\sum_{i=1}^r S_i}{\sum_{j=1}^n S_j} \quad (4)$$

Здесь S_i – сингулярные значения, взятые из матрицы Σ . Если этот график показывает, что скорость изменения сингулярных значений начинает выравниваться, то это указывает на величину r , при которой будет сохранено достаточное количество сингулярных значений. Если же график кумулятивных сингулярных значений не имеет уровня сглаживания, то размерность урезанной матрицы Σ выбирается согласно целям исследования. Заметим, что размерность матрицы Σ берется как минимум на порядок меньше, чем число рассматриваемых терминов.

Поскольку столбцы матрицы U_{mr} состоят из весов, которые применяются к записям в матрице частотности документа, то индивидуальный вес по его значению говорит о том, насколько «важен» соответствующий термин в данном столбце. Если есть несколько терминов с более высоким весом для данного измерения, и эти термины соответствуют одному понятию, то само понятие может использоваться как ярлык для данного измерения. Однако для интеллектуального анализа текста проблема интерпретации измерений усугубляется огромным количеством входных переменных, что обычно затрудняет интерпретацию. В результате в ТМ чаще всего используется SVD только для аспекта уменьшения размера. Любая интерпретация выполняется в конце процесса путем анализа частотности терминов, встречающихся в данном кластере или категории.

Результаты исследования

Для анализа ценностей, формирующих значение в языковом сознании, нами использовался метод свободного ассоциативного эксперимента, предусматривающий спонтанные ответы респондентов. Участники эксперимента не были ограничены во времени и количестве ответов, им предлагалось реагировать на стимул любым количеством слов, которые первыми приходят им в голову, без каких-либо ограничений в формальных или семантических особенностях ответа. Существует возможность искажений в представлении реальных связей в ментальном лексиконе,

поскольку последующая реакция может даваться уже не на стимул, а на предыдущую реакцию. Однако в работе [Яковлев 2018: 16] показано, что количество таких реакций не превышает 1% из общего количества, чаще возникает реакция смешанного типа – на стимул и предыдущую реакцию.

Стимулами в эксперименте были исследуемые слова-ценности, на которые респонденты предлагали свои ассоциации, называемые реакциями или ассоциатами. В эксперименте приняли участие 444 человека в возрасте 15–80 лет из разных регионов России; в результате были получены ответы 393 респондентов и 51 отказ. Выбор нижнего возрастного порога обусловлен концепцией Л.С. Выготского, согласно которой полноценные понятия формируются у человека в среднем подростковом возрасте, до этого осознание и оперирование абстрактными категориями невозможно [Выготский 2001]. В эксперименте участвовали 214 мужчин, что составляет около 55% от общего числа респондентов, и 179 женщин, то есть 45%, соответственно.

Территориальный признак проживания испытуемых представлен в исследовании весьма широко. Для последующего анализа вводятся 4 условных региона:

- 1) столицы (г. Москва и г. Санкт-Петербург) – 24%;
- 2) центральная часть России (г. Нижний Новгород, г. Липецк, г. Саранск, г. Брянск, области Московская, Новгородская, Владимирская и пр.) – 26%;
- 3) юг России (г. Ростов-на-Дону, г. Краснодар, г. Ставрополь, Республики Адыгея, Северная Осетия, Дагестан и пр.) – 30%;
- 4) Дальневосточный регион (г. Оренбург, г. Иркутск, Республика Бурятия, Алтайский край, Тюменская область и пр.) – 20%.

Среднее значение по территориальному признаку проживания составляет 2,47, что обозначает довольно ровное распределение респондентов по 4 условным классам регионов.

Возрастной диапазон респондентов представлен следующим образом: 43% юношества в возрасте от 17 до 22 лет; 36% молодежи до 35 лет; 14% взрослых людей до 50 лет и 7% людей более старшего поколения.

Проанализируем результаты ассоциативного эксперимента – реакции, полученные на слово-стимул воля. Всего было получено 1349 реакций.

Ассоциативное поле представлено следующими реакциями (см. табл. 1).

Таблица №1

Ассоциативное поле ценности воля

Частота реакции	Реакция
162	<i>свобода</i>
134	<i>сила</i>
31	<i>желание</i>
30	<i>победа</i>
26	<i>стремление</i>
25	<i>характер</i>
24	<i>выбор, упорство</i>
23	<i>Павел, терпение</i>
22	<i>дух, разум</i>
17	<i>власть</i>
16	<i>жизнь</i>

13	<i>простор, ветер</i>
12	<i>независимость, поле, счастье</i>
11	<i>мужество</i>
10	<i>усилие</i>
9	<i>дом, воля, смелость, спорт, степь, твердость, труд, честь</i>
8	<i>выдержка, кулак, небо, полет, старание, цель</i>
7	<i>вера, земля, интернет, раздолье, слава, сталь, стойкость</i>
6	<i>бог, испытание, мужчина, народ, неволя, преодоление, рабство, решимость, уверенность, успех, целеустремленность, человек</i>
5	<i>выносливость, горы, дружба, закалка, кремень, ограничение, отношение, порыв, правда, право, природа, стержень, ум, энергия</i>
4	<i>воздух, горе, мощь, отвага, песня, птица, решение, семья, тяга, харизма</i>
3	<i>боль, деньги, добро, достижение, качество, концентрация, машина, мир, напор, наслаждение, отпуск, слабость, слово, сложно, солнце, страсть, страх, уважение, усердие</i>
2	<i>бег, брат, время, гражданка, движение, действие, доля, доброта, дума, железо, железная воля, карандаш, колос, конь, контроль, легкость, могущество, море, мысли, необходимо, несгибаемая воля, ночь, обеспеченность, освобождение, осознание, отдых, откровение, подавление, подвиг, покой, путь, путешествие, радость, развитие, результат, родина, сдержанность, справедливость, статуя, тюрьма, увольнение, удовольствие, храбрость, чувство</i>
1	<i>авторитет, Бакунин, безграничность, беззаботность, безнаказанность, бесконечность, беспечность, благородство, большая, борьба, братство, будущее, веселье, воздержание, возможность, воин, война, вольность, воспитывать, вперед, герой, горький, граница, дело, делаю и говорю что хочу, держать, доблесть, добиваться, доблесть, доказательство, дорога, доступ, жажда, животные, воля к жизни, зависимость, заключение, законность, заставить, заточение, значимость, изъятие, исполнитель, казачество, камеди-клуб, каменный, клетка, к победе, край, красота, кругозор, крылья, кумир, лидер, любовь, место, мило, молодость, мотивация, мотоцикл, мощь, мудрость, мужественный, музыка, мышцы, надежда, непоколебимая, несгибаемая, несломимая, Ницше, нож, нужно, облегчение, обязанности, оковы, опыт, орел, ответственность, ответственная, отвлечение, политика, понимание, поступки, потребность, похоть, президент, преодоление трудностей, прижать, принуждение, принцип, прогулка, работа, равенство, ракета, раскрепощение, рвение, река, релакс, решительность, родная, русский, рыбалка, самодисциплина, самообладание, саморазвитие, самостоятельность, сверхчеловек, светлый, сердце, сильная, ситуация, сказка, скала, слабая, служба, смекалка, смех, снежок, совесть, сознание, сопротивление, способность, стойкий, страдание, страна, судьба, ТВ, театр, телевизор, терпимость, титаны, ТНТ, традиции, тренировка, трудно, трудолюбие, указ, упрямство, условие, утес, участие, хобби, холостяк, хочу, царская, цепи, честный, честность, Шопенгауэр, will.</i>

Ядро ассоциативного поля составляют частотные реакции в количестве 99, а обширную периферию – единичные реакции (набравшие три и менее голосов), они составили 294 единицы. Таким образом, четверть (25%) реакций являются общими для носителей русского языка и культуры, а остальные две трети (75%) характеризуются индивидуальными представлениями человека. На рис. 1 представлен частотный график десяти наиболее часто встречающихся ассоциатов.

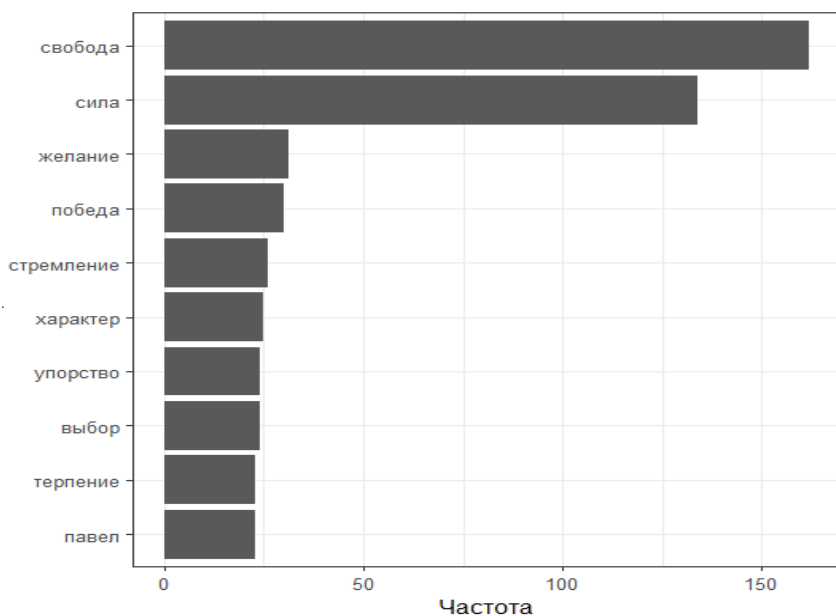


Рис. 1. Столбчатая диаграмма наиболее частотных ассоциатов к слову-стимулу *воля*

Как видно на графике, наиболее часто встречаемыми реакциями являются *свобода* (N=162) и *сила* (N=134). Все остальные реакции встречаются реже в несколько раз, например, реакция *желание* имеет всего 31 голос.

При анализе ассоциативного поля обращает на себя внимание преимущественно положительная окраска ассоциатов, таких как *счастье*, *победа*, *доброта*, *вера*, *успех*. Важно отметить также и наличие в ядре ценности прецедентного имени – *Павел*, обозначающего современного комедийного актера разговорного жанра Павла Волю, что подкрепляется присутствием таких единичных ассоциатов в периферии ассоциативного поля, как *камеди-клуб*, *снежок*, *ТВ* и *ТНТ*.

Распределения ядерных реакций свободного ассоциативного эксперимента в зависимости от пола, возраста или региона проживания респондентов представлены на рис. 2 и 3.

Из рис. 2 видно, что некоторые ассоциаты практически одинаково распределены у мужчин и женщин – это *свобода*, *сила*, *стремление* и др.; а есть ассоциаты с гендерным предпочтением – это, например, *желание* у женщин и *победа*, *дух*, *терпение*, *власть*, *разум* у мужчин. Выявленные ассоциаты носителей русского языка и культуры вполне соотносятся с данными этимологических словарей, подчеркивающих процессуальную

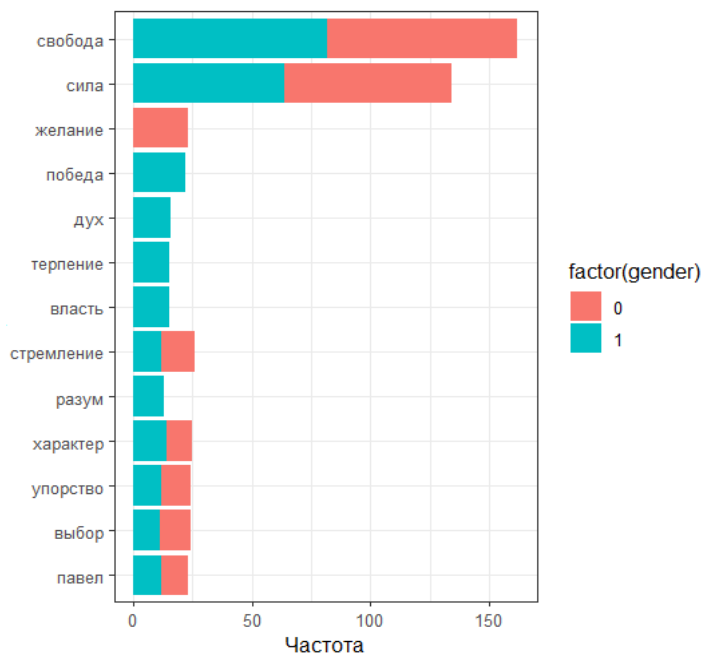


Рис. 2. Столбчатая диаграмма наиболее частотных ассоциатов в зависимости от пола респондента

семантику, тесно связанную с глаголом *велесть*, где ядерным смыслом ценности воля выступает «направленное желание», а производными являются *власть*, *сила*, *выбор*, *желание* [Вахненко 2018]. Эти совпадения говорят о преемственности поколений в понимании смысла ценности воля.

Из диаграмм рис. 3а и 3б можно заключить, что наиболее частотные реакции *свобода* и *воля* присутствуют у представителей всех возрастных категорий и во всех регионах. А далее наблюдаются некоторые различия. Например, *победа* наиболее часто встречается у юношества из Центральной части России, а *характер* и *выбор* у юношества Юга России. Для более тщательного анализа необходимо провести кластеризацию ответов-реакций.

Ко всей коллекции документов был применен интеллектуальный анализ текста. И первой задачей, рекомендации по решению которой достаточно обширны, было определение оптимального числа концептов. В настоящей работе мы опирались на сравнение с общей дисперсией, которое реализуется через критерий информативности. Рассмотрим график критерия информативности, рассчитанный по формуле (4), на рис. 4.

Для того чтобы описать 25% общих реакций для носителей русского языка и культуры на слово-стимул воля достаточно 8 концептов, что совпадает с общими рекомендациями выбора от 10 до 250 измерений [Albright 2004: 13]. Таким образом, 99 терминов было спроектировано на 8 концептов. В табл. 2 представлены SVD-коэффициенты, подтверждающие выбранные концепты.

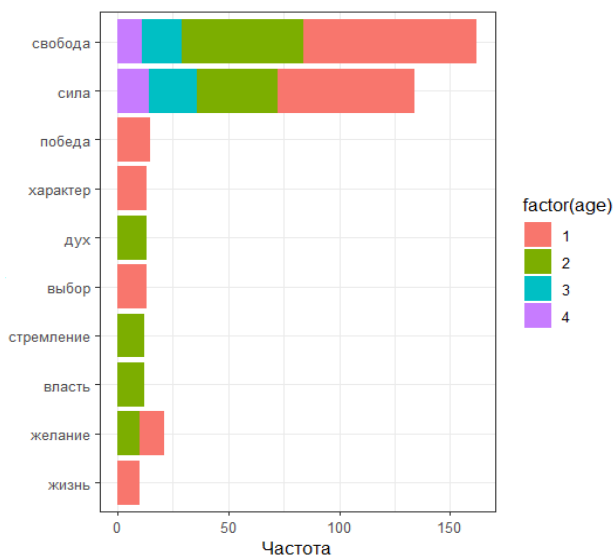


Рис. 3а. Столбчатая диаграмма наиболее частотных ассоциатов в зависимости от возраста респондента

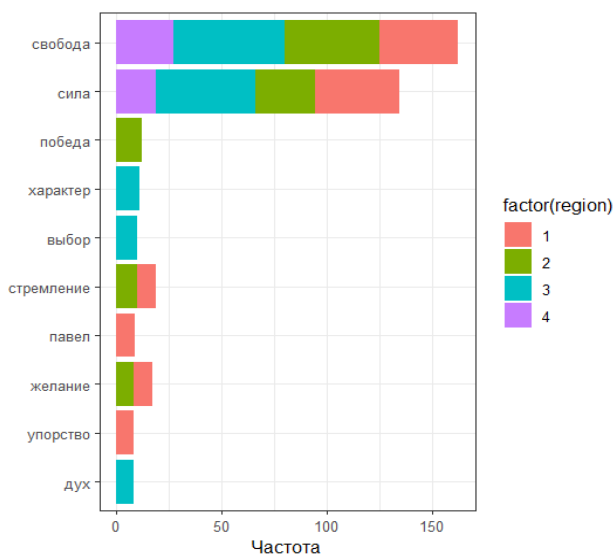


Рис. 3б. Столбчатая диаграмма наиболее частотных ассоциатов в зависимости от региона проживания респондента

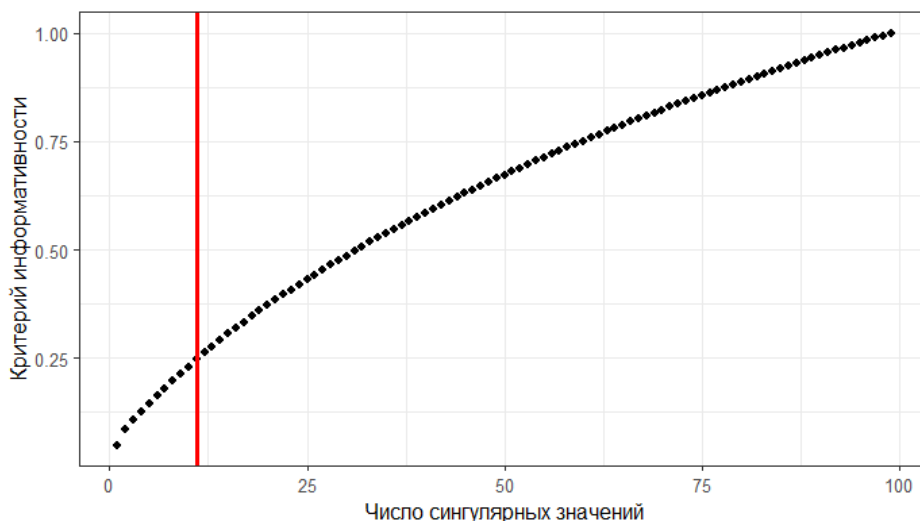


Рис. 4. График критерия информативности

Таблица №2

Фрагмент матрицы концептных нагрузок на стимул воля с выбором наибольших SVD-коэффициентов

Ассоциаты на стимул воля	Коэффициенты сингулярного разложения							
	Concept 1	Concept 2	Concept 3	Concept 4	Concept 5	Concept 6	Concept 7	Concept 8
желание	0,083	-0,024	0,642	-0,505	0,097	0,136	0,081	0,016
победа	0,046	0,078	0,456	0,735	0,072	0,143	-0,121	-0,165
разум	0,075	-0,005	-0,009	-0,041	0,598	0,191	-0,040	-0,264
свобода	0,566	-0,585	-0,049	0,051	-0,111	0,007	-0,026	0,041
сила	0,782	0,764	0,170	-0,097	0,066	-0,028	-0,048	-0,096
характер	0,068	0,114	-0,053	0,016	0,190	0,605	0,221	0,487
выбор	0,110	-0,054	-0,031	-0,059	0,658	-0,205	-0,036	-0,136
Павел	0,049	0,031	0,107	-0,046	-0,103	0,160	0,737	0,030
терпение	0,061	0,073	0,085	0,114	0,257	-0,516	0,255	0,518
стремление	0,063	0,115	0,438	-0,181	-0,112	0,058	-0,404	0,195
власть	0,032	0,013	0,003	0,062	-0,007	0,098	-0,007	0,357

Термины с более высокими значениями в матрице $U_{99,8}$ формулы 2 могут интерпретироваться как ярлык для данного концепта, которые представлены в столбце ассоциатов ценности воля в табл. 3. В правом столбце данной таблицы представлена интерпретация каждого концепта признакового поля.

Таблица №3

Ассоциаты и интерпретации концептов табл. 2

№ концепта	Ассоциаты ценности воля	Интерпретация ценности воля
Concept 1	{+свобода, + сила}	свобода действий, опирающаяся на силу, т.е. активная/агрессивная свобода
Concept 2	{+сила, – свобода}	сила без свободы, т.е. безволие/нежелание действий
Concept 3	{+желание, + победа, + стремление}	желание в сочетании с победой и стремлением, т.е. активно-осознанная свобода
Concept 4	{+ победа, – желание}	победа без желания, т.е. мечтательность
Concept 5	{+ выбор, + разум}	выбор, опирающийся на разум, то есть осуществление осознанного выбора
Concept 6	{+ характер, – терпение}	характер без терпения, т.е. проявление упрямства
Concept 7	{+ Павел, – стремление}	Павел (стендап-комик) без стремления, т.е. открытость, безнаказанность, беспечность
Concept 8	{+ терпение, + характер,+ власть}	терпение в сочетании с характером и властью, т.е. постепенное восхождение на пьедестал

Далее, пользуясь переменными признакового пространства – концептами, произведем кластеризацию документов, которая позволит описать восприятие ценности воля различными группами респондентов.

Множество документов представим 8-мерными векторами подпространства согласно формуле (3) в ортогональном базисе u_1, u_2, \dots, u_8 . Проекция данного множества документов на векторы u_1 и u_2 отчетливо демонстрирует разбиение на четыре кластера (рис. 5). Поэтому на следующем этапе проводится процедура кластеризации методом k средних [Kwartler 2017].

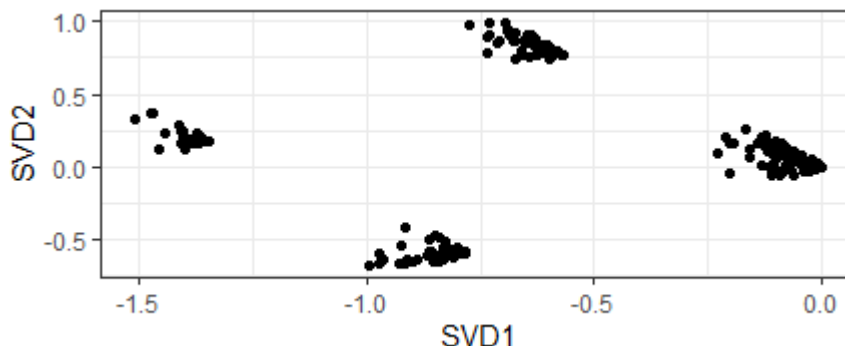


Рис. 5. Проектирование документов на u_1 и u_2 для определения числа кластеров

В результате кластеризации получилось четыре кластера с количеством документов 209, 126, 36 и 22 и количеством терминов 226, 138, 42 и 34 соответственно. Анализируя матрицу средних значений по результатам кластеризации, которая представлена в табл. 4, можно сделать вывод об однородности кластеров по таким признакам, как возраст респондента, пол и регион проживания. Это говорит о применимости предлагаемых выводов ко всей популяции России. Небольшое смещение наблюдается в третьем и четвертом кластерах в более взрослую группу населения и в третьем кластере в сторону женщин.

Таблица №4

Матрица средних значений признаков

№ кластера	Респонденты		
	возраст	пол	регион проживания
1	28,19	0,57	2,48
2	26,57	0,52	2,42
3	30,67	0,44	2,47
4	31,36	0,55	2,55

Наиболее частотными ассоциатами в первом наиболее многочисленном кластере является термин *сила* (N=85), который поддержан терминами *победа* (N=25) и *стремление* (N=23), во втором – свобода (N=126). В третьем кластере, который представляет собой более «женскую» группу населения, наиболее выражены ассоциаты *свобода* (N=36) и *сила* (N=36) с большим отрывом по частоте от остальных. В четвертом более «взрослом» и самом малочисленном кластере по частоте лидером является *характер* (N=22).

Таким образом, в современном русском языковом сознании в понимании ценности воля значимыми элементами являются вышеупомянутые кластеры, которые характеризуют волю как одну из важных ценностей русской культуры, подчеркивая положительные черты характера человека, обладающего целеустремленностью, мужеством, героизмом (кластер 1 и 3), свободой (кластер 2 и 3), а также настойчивостью (кластер 4).

Важно отметить, что ценности свобода и воля содержательно близки и, как доказано в диссертационном исследовании О.В. Казаченко [Казаченко 2021], определяются

посредством друг друга, объединяясь в диаду, что экспериментально подтверждает положение о систематичности ценностей, их тесной связи и взаимозависимости. У данной выборки респондентов наиболее частотными ассоциатами к слову СВОБОДА являются *воля* (N=69) и *независимость* (N=57).

Дискуссия

Выбор наиболее частотных реакций на стимул *воля* в языковом сознании носителей русского языка и культуры – *свобода, сила, победа, стремление, характер и желание* – может быть объяснен следующими факторами. Во-первых, лексема *воля* объясняется в толковых словарях через набор сем, которые включают *свободу, свободу действий, независимость, отсутствие всякого стеснения или принуждения, запретов в чем-либо, неподвластность, простор в действиях, самоволие, произвол, свободу от рабства, от крепостного состояния*. При этом отметим, что данная сема не является основной, а расположена на 3–5 местах после таких компонентов значения, как «способность осуществлять свои желания, поставленные перед собой цели» и «сознательное стремление к осуществлению». Тем не менее, связь словарной семы и компонента значения слова *воля* налицо.

Во-вторых, *воля* подразумевает под собой некую внутреннюю силу, стремление, которые выражены в словосочетании *сила воли*. Она является двигателем внутреннего настроя человека, направленного на выполнение каких-либо целей и задач. То есть в данном случае ценность *воля* способствует действию, а значит и поведению. Заметим также, что это важно для нашего исследования, поскольку ценности отличаются от других явлений человеческой жизни тем, что они направляют деятельность, поэтому обладают определенным прогностическим потенциалом.

В-третьих, появление реакции *победа*, как ключевой в понимании ценности *воля*, характеризуется, с нашей точки зрения, двумя взаимосвязанными причинами. Первая состоит в том, что при успешной реализации воли постигается результат, который можно назвать победой над собой (обстоятельствами, препятствиями и т.д.). Вторая, как нам представляется, связана с победой в Великой отечественной войне 1941–1945 гг., поскольку реализация воли народа в борьбе за свободу своей страны закончилась победой над фашизмом. В первом случае *воля* реализуется как личностное качество, а во втором – как чувство сопричастности к этносу.

В-четвертых, *желание* и *стремление* также являются важными ассоциатами, поскольку *воля* представляет собой целенаправленное действие и без осознанного желания и стремления к достижению определенной цели не осуществима.

И, наконец, *воля* связана с такой важной реакцией ассоциативного поля, как *характер*, поскольку, только имея сильный характер, человек может выступить против обстоятельств, обнаруживая *силу воли*, т.е. волен поступать в соответствии со своими принципами.

Резюме

При анализе результатов ассоциативного поля ценности *воля* были получены следующие результаты: самыми частотными реакциями являются *свобода* и *сила*. Эти ассоциаты вместе со *стремлением* практически одинаково распределены по гендерному признаку; но есть ассоциаты с гендерным предпочтением – *желание* у женщин и *победа, терпение, власть, разум, дух* у мужчин. Выявленные ассоциаты вполне соотносятся с данными этимологических словарей, подчеркивающих процессуальную семантику исследуемого стимула.

Изучая территориальное разнообразие, мы пришли к выводу, что все регионы также представлены в наиболее частотных реакциях *свобода* и *сила*, но что касается менее частотных ассоциатов, то здесь наблюдаются некоторые различия: *победа* наиболее часто встречается у юношества из Центральной части России, а *характер* и *выбор* у юношества Юга России. Для более тщательного анализа была проведена кластеризация реакций.

Кроме того, можно заключить, что использование количественных методов в анализе результатов свободных ассоциативных экспериментов помогает исследователю, во-первых, автоматизировать процесс анализа больших массивов данных; во-вторых, выявлять гендерную, возрастную и территориальную особенности ответов респондентов, пригодных для дальнейшего качественного анализа; в-третьих, строить ассоциативные поля изучаемой ценности с выделением ядра и периферии с интерпретацией числовых (положительных и отрицательных) результатов для анализа языковых единиц. А это уже поле действий для работы психологов и педагогов, которые корректируют программы по воспитанию подрастающих поколений.

© Казаченко О.В., Шмалько С.П., Дорошенко О.В., 2023

Литература

Алиакберова Л.Ф. Современные программные продукты для анализа данных // Молодой ученый. 2021. № 37 (379). С. 13–17. URL: <https://moluch.ru/archive/379/83970/> (дата обращения: 10.10.2023).

Андреева С.В., Усейнова Л.С. Понятие «любовь» как компонент языковой картины мира // Известия Саратовского университета. Новая серия. Серия: Филология. Журналистика. 2021. Т.1. Вып.3. С. 272–276. URL: <https://doi.org/10.18500/1817-7115-2021-21-3-272-276> (дата обращения: 09.10.2023).

Богданович Г.Ю. Ассоциативный эксперимент как прием интерпретации дискурса // Дискурс-Пи. 2018. №1 (30). URL: <https://cyberleninka.ru/article/n/assotsiativnyy-eksperiment-kak-priem-interpretatsii-diskursa> (дата обращения: 09.11.2023)

Бубнова И. А. Ценности и образ будущего поколения Z: специфика системы // Вестник РУДН. Серия: Теория языка. Семиотика. Семантика. 2021. №2. URL: <https://cyberleninka.ru/article/n/tsennosti-i-obraz-budushego-pokoleniya-z-spetsifika-sistemy> (дата обращения: 18.10.2023).

Вахненко А.П. Концепт «воля» и его отражения в абстрактном существительном // Вопросы науки и образования. 2018. №27 (39). URL: <https://cyberleninka.ru/article/n/kontsept-volya-i-ego-otrazheniya-v-abstraktnom-suschestvitelnom> (дата обращения: 13.10.2023).

Выготский Л.С. Мышление и речь: Психика, сознание, бессознательное. М.: Лабиринт, 2001. 366 с.

Ивин В.В. Применение языка R и среды RStudio для статистического анализа данных // Педагогический опыт: от теории к практике. 2018. С. 47–53. <https://interactive-plus.ru/e-articles/568/Action568-473126.pdf> (дата обращения: 30.10.2023).

Казаченко О.В. Структура и содержание аксиологической сферы русского образа мира в XXI веке: дисс...д-ра филол. наук. М., 2021. 399 с.

Каширский Д.В. Психология личных ценностей: дисс... д-ра психол. наук. М., 2014. 550 с.

РАС – Русский ассоциативный словарь: [В 2 т.] / Ю.Н. Караулов, Г.А. Черкасова, Н.В. Уфимцева [и др.]. М.: АСТ: Астрель, 2002. 781 с.

САНРЯ – Леонтьев А.А. Словарь ассоциативных норм русского языка. Изд-во Моск. Ун-та, 1977. 192 с.

Ситорам К., Когделл Р. Основы межкультурной коммуникации // Человек. 1992. № 4. С. 106–116.

Федосютина Н.С. Слова-ценности как средство доступа к ценностной картине мира: Экспериментальное исследование: дисс... канд. филол. наук. Курск, 2005. 160 с.

Яковлев А.А. Об одном методологическом изъеме при проведении свободного ассоциативного эксперимента // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2018. Т.16. Вып. 4. С. 16–25. DOI 10.25205/1818-7935-2018-16-4-16-25 (дата обращения 13.10.2023)

Albright R. Taming text with the SVD. SAS Institute Inc, Cary. 2004. 17 p.

Feinerer I. Introduction to the tm Package Text Mining in R. 2022. Available at: URL: <https://cran.r-project.org/web/packages/tm/vignettes/tm.pdf> (дата обращения: 10.10.2023).

Han J., Micheline K., Jian P. Data mining trends and research frontiers. In: Data mining, 3rd edition. Morgan Kaufmann, 2012. P. 585–631.

Johnson K. Quantitative method in linguistics. Malden, MA: Blackwell, 2008. 277 p.

Kwartler T. Text mining in practice with R. John Wiley & Sons, 74, 2017. 307 p.

Levchenko O., Tyshchenko O., Dilai M., Gajarsky L. A model of the information system of the associative verbal network presentation // Advance in Intelligent Systems and computing. 2020. Vol. 1293. P. 71–83.

Papadimitriou C.H., Raghavan P., Tamaki H., Vempala S. Latent semantic indexing: a probabilistic analysis. In: Proceedings of the 17th ACM Symposium on Principles of Database Systems (PODS). 1998. P. 159–168.

Rasinger S.M. Quantitative research in linguistics: an introduction. Research methods in linguistics. Bloomsbury Academic, 2008. 305 p.

Rokeach M. The Nature of Human Values. N.Y.: Free press, 1973. 322 p.

Сведения об авторах:

Казаченко Оксана Васильевна – доктор филологических наук, доцент, профессор департамента филологии ГАОУ ВО МГПУ; доцент департамента английского языка и профессиональной коммуникации ФГБОУ ВО Финансовый университет при Правительстве Российской Федерации,

Контактная информация:

129226, Москва, 2-й Сельскохозяйственный проезд, дом 4; 125167, Москва, пр-кт Ленинградский, д. 49/2,

ORCID: 0000-0001-9208-6785

e-mail: kazachenko_07@mail.ru

Шмалько Светлана Петровна – кандидат педагогических наук, доцент, доцент кафедры информационных образовательных технологий ФГБОУ ВО КубГУ,

Контактная информация:

350040, г. Краснодар, ул. Ставропольская, д. 149.

ORCID: 0000-0002-6456-9648

e-mail: shmalko_sis@mail.ru

Дорошенко Ольга Валерьевна – кандидат физико-математических наук, научный сотрудник института математики, механики и информатики ФГБОУ ВО КубГУ

Контактная информация:

350040, г. Краснодар, ул. Ставропольская, д. 149

ORCID: 0000-0002-8037-2976

e-mail: oldorosh@mail.ru

Для цитирования:

Казаченко О.В., Шмалько С.П., Дорошенко О.В. Использование методов интеллектуального анализа данных для свободного ассоциативного эксперимента (на примере ценности **ВОЛЯ**) // Вопросы психолингвистики №4(58) 2023, С. 75–93, doi: 10.30982/2077-5911-2023-58-4-75-93

UDC 81'(23+27)

LBC 81

DOI 10.30982/2077-5911-2023-58-4-75-93

Research article

**TEXT MINING METHODS FOR A FREE ASSOCIATIVE EXPERIMENT
(ON THE VALUE *WILL*)**

Oksana V. Kazachenko,

Moscow City University

Financial University

Moscow, Russia

Svetlana P. Shmalko,

Kuban State University

Krasnodar, Russia

Olga V. Doroshenko,

Kuban State University

Krasnodar, Russia

Abstract

The focus of modern linguistic and methodological research has shifted towards the study of a person as a developing linguistic personality, which is organically integrated into the anthropocentric and axiological paradigms of the development of scientific thought in the 21st century. Statistically processed analysis of the associative fields of Russian speakers is necessary to consider changes in the linguistic consciousness of the individual; to understand the modern system of personal values, which can contribute to the value-oriented education of the younger generations of native speakers of the language and culture.

The material of the study was the results of a free associative experiment on many people living in the Russian Federation. This widely used experiment allows the researchers to identify the core of the studied stimulus, which characterizes the generally accepted associations of people, and the periphery, indicating individual-personal associations. Using statistical methods for processing data, the authors attempted to identify the main associates that determine the understanding of the studied value *will* in the form of concepts.

Text Mining (TM) was applied to all reactions to the value *will* to determine the optimal number of concepts. With its help, a singular value decomposition of the matrix was carried out, which consists in finding the SVD coefficients. It is important to interpret the positives and negatives of these coefficients, which are significant from the point of view of the meaning of value. So, one of the concepts of the value *will*, which received the two largest coefficients with opposite signs, is interpreted as a force, but with the opposite sign to freedom, i.e. will as lack of will / not desire for action. Further, the authors used the method of Latent Semantic Indexing for the clustering of associates, which made it possible to describe the perception of the value-stimulus *will* by various groups of respondents.

Statistical methods are increasingly used in applied philological research, but it is rather difficult to meaningfully interpret numerical (positive and negative) results for the analysis of language units. Our study provides an example of such an interpretation.

Keywords: associative experiment, will, data mining technology, concepts, SVD-coefficients, values

© Kazachenko O.V., Shmalko S.P., Doroshenko O.V., 2023

Bionotes:

Kazachenko Oksana V. – Doctor of Philology, Associate Professor, Moscow City University; Financial University under the Government of the Russian Federation

Contact information:

4 Vtoroy Selskohoziaystvenny proezd, Moscow, 129226, Russia

ORCID: 0000-0001-9208-6785

e-mail: kazachenko_07@mail.ru

Shmalko Svetlana P. – Candidate of Pedagogy, Associate Professor, Department of Information Technologies, Kuban State University.

Contact information:

350040 ul. Stavropolskaya, 149, Krasnodar, Russian Federation

ORCID: 0000-0002-6456-9648

e-mail: shmalko_sis@mail.ru

Doroshenko Olga V. – Candidate of Physics and Mathematics, research officer, Institute of Mathematics, Mechanics and Computer Sciences, Kuban State University

Contact information:

350040 ul. Stavropolskaya, 149, Krasnodar, Russian Federation

ORCID: 0000-0002-8037-2976

e-mail: oldorosh@mail.ru

For citation:

Kazachenko O.V., Shmalko S.P., Doroshenko O.V. Text mining methods for a free associative experiment (on the value WILL) // Journal of Psycholinguistics. 4(58), 2023. P. 75–93. Available from: doi: 10.30982/2077-5911-2023-58-4-75-93 (In Russian)