Санкт-Петербургский государственный университет
Филологический факультет
Кафедра математической лингвистики

Курочкина Юлия Николаевна

ОЦЕНКА АВТОМАТИЧЕСКИХ МЕТОДОВ ВЫЯВЛЕНИЯ УСТОЙЧИВЫХ СЛОВОСОЧЕТАНИЙ

Магистерская диссертация

Направление «Лингвистика»

Образовательная программа «Прикладная и экспериментальная лингвистика»

Профиль «Компьютерная лингвистика и интеллектуальные технологии»

Научный руководитель:
доц., к.ф.н. Захаров В.П.

Санкт-Петербург
2017

Оглавление

Аннотация

Данная работа посвящена оценке мер ассоциации, используемых для выявления коллокаций. Поставлена цель выяснить, насколько эффективны и релевантны эти меры ассоциации, и показать это на примере популярных мер. В работе рассматриваются теоретические вопросы извлечения коллокаций, классификации устойчивых словосочетаний, дано описание мер ассоциаций. В работе описан эксперимент по выявлению коллокаций из корпуса Araneum Russicum Russicum Maius объемом 1,2 млрд токенов в системе NoSketch Engine, представлена оценка работы мер ассоциации.

Ключевые слова: корпуса, извлечение коллокаций, меры ассоциации**.**

ВВЕДЕНИЕ

Важность и роль выражений из нескольких слов, т.е. устойчивых словосочетаний, в прикладной лингвистике и в сфере обработки естественного языка давно признаны. Тем не менее, эти единицы требуют дальнейшего изучения. Ранее, без помощи компьютерных устройств, научное исследование было сопряжено с многими техническими трудностями, но при наличии соответствующих инструментов исследования стало проще, быстрее и удобнее.

Наш научный интерес сосредоточен на устойчивых сочетаниях, в состав которых входят сочетания разных типов.

Существуют различные методы автоматического выявления устойчивых словосочетаний (коллокаций) на базе больших корпусов текстов. В целом процедура заключается в отборе кандидатов в коллокации на основе выбранных критериев.

Помимо выявления устойчивых словосочетаний, нужна еще и оценка методов их выявления. Суть исследования заключается в том, чтобы проанализировать возможные и доступные автоматические методы, сравнить их, выявить положительные и отрицательные стороны и предложить вариант их улучшения или применения в зависимости от полученного результата.

Актуальность работы состоит в том, что эта тема в современной компьютерной лингвистике важна как в теоретическом, так и в практическом плане. Выявление устойчивых словосочетаний значимо для составления словарей, а также для использования их в самых разных прикладных задачах. Кроме того, эта задача представляет собой еще и теоретическое исследование, изучающее закономерности сочетаемости языковых единиц.

Объект изучения - устойчивые словосочетания.

Предмет исследования – методы автоматического извлечения на основе мер ассоциации и способы оценки их применимости и эффективности.

Материалом исследования послужили данные различных корпусов и инструменты корпусных систем.

Цель исследования - выяснить, насколько эффективны и релевантны меры ассоциации.

Сформулированная таким образом цель определила ряд стоящих перед нами **задач**:

1. описать понятие сочетаемости
2. рассмотреть методы выявления устойчивых словосочетаний;
3. описать меры ассоциации
4. выбрать метрики для оценки методов выявления устойчивых словосочетаний;
5. провести эксперименты по выявлению коллокаций;
6. провести оценку эффективности мер ассоциации;
7. наметить пути повышения эффективности методов выявления коллокаций.

Методы исследования включают использование корпусных инструментов, предназначенных для извлечения словосочетаний, их настройку, статистическую обработку данных путем сопоставления их с ассоциативными, толковыми, фразеологическими словарями, экспертную оценку.

Существует несколько точек зрения на определение термина *устойчивое словосочетание*, в данной работе мы будем рассматривать устойчивые словосочетания с точки зрения корпусной лингвистики, то есть опираясь на статистические методы. Также в нашей работе слово «словосочетание» будет заменяться на выражение *коллокация*, в соответствии с зарубежным термином *collocation* или *multiword expression.*

Практическая значимость данной работы заключается в том, что полученные результаты могут быть применены при решении различных задач прикладной лингвистики. Можно предположить, что итоги исследования окажутся полезными при составлении словарей, корпусов, снятии семантической неоднозначности.

Работа состоит из введения, четырех глав, заключения, списка литературы и приложений. **В первой главе** рассказывается про отношения в языке, дается

определение понятию "словосочетания" и приводится классификация устойчивых словосочетаний. **Во второй главе** обсуждаются методы извлечения коллокаций и дается классификация и описание мер лексической ассоциации. **Третья глава** посвящена методам и критериям оценки как таковой в целом и оценки методов извлечения устойчивых словосочетаний. **Четвертая глава** содержит описание эксперимента и оценки работы мер ассоциации различными способами.

ГЛАВА 1. СЛОВОСОЧЕТАНИЯ В ЯЗЫКЕ

1.1.Отношения между словами

Что такое словосочетание? Это осмысленное сочетание слов, связанных по смыслу и грамматически. Словосочетания не следует путать с n-граммами, которые представляют собой обычную последовательность слов в тексте. Последовательность из двух элементов называют биграммой, последовательность из трёх элементов - триграммой**.** В отличие от n-грамм, словосочетания связаны отношениями языка, а именно парадигматическими и синтагматическими.

Язык как знаковую систему принято изучать с точки зрения парадигматики и с точки зрения синтагматики.

"СИНТАГМАТИКА -1) один из двух аспектов исследований языка – изучение языковых единиц в линейном ряду, в тех реальных отношениях, которыми они связаны в тексте; противополагается парадигматике". [Лингвистический энциклопедический словарь, http://tapemark.narod.ru/les/447d.html].

"ПАРАДИГМАТИКА - 1) один из двух аспектов системного изучения языка, определяемый выделением и противопоставлением двух типов отношений между элементами и/или единицами языка — парадигматических и синтагматических; раздел науки о языке, занимающийся парадигматическими отношениями, их классификацией, определением области их действия и т. п.; противопоставляется синтагматике по типу изучаемых отношений и их группировок; 2) в более широком смысле — то же, что система языковая, понимаемая как совокупность лингвистических классов — парадигм; противопоставляется синтагматике как синониму понятия лингвистического процесса и текста (Л. Ельмслев)". [Лингвистический энциклопедический словарь, http://tapemark.narod.ru/les/366b.html]. Законы построения предложений и сочетания слов основываются на этих связях.

Существует мнение, что в голове у человека находятся уже готовые синтагмы. Когда человек создает текст, в этом тексте имеют место как

синтагматические, так и парадигматические связи, однако принято считать, что в тексте присутствуют только синтагмы.

Язык - это парадигматическая система. Слова, расположенные близко друг к другу в тексте, могут быть не связаны по смыслу и не образовывать словосочетание, но при этом они будут связаны парадигматически, так как относятся к одной теме. Приведем пример: некий инструмент извлечения коллокаций выдает нам сочетание «болезнь таблетка» в качестве коллокации. Он сделал это потому, что эти понятия (болезнь и таблетка) находятся в одном семантическом поле, и это парадигматическая связь. В данной ситуации нам покажется это сочетание бессмысленным, хотя по сути инструмент сработал правильно - эти два слова действительно связаны по смыслу.

1.2. Понятие сочетаемости

Сочетаемость - это "свойство языковых единиц сочетаться при образовании единиц более высокого уровня; одно из фундаментальных свойств языковых единиц, отражающее синтагматические отношения между ними" [Лингвистический энциклопедический словарь, 1990]. Лингвисты обычно выделяют универсальные и конкретно-языковые законы и тенденции сочетаемости, отступив от которых говорящий или пишущий нарушает норму или провоцирует изменение свойств языковых единиц. Однако намеренное нарушение правил сочетаемости может быть средством художественной выразительности.

Существует классификация сочетаемости в зависимости от позиции — *контактная* (когда языковые единицы соположены) и *дистантная* (находятся на расстоянии); в зависимости от факторов сочетаемости — *обусловленная* (определяется наличием у языковых элементов различительных черт) и *произвольная* (определяется лишь принятой нормой); в зависимости от уровня языка — формальная и семантическая.

Понятие сочетаемости относится к разным уровням языка. На уровне фонем обусловленная сочетаемость заключается в совместимости или в несовместимости их дифференциальных признаков (например, во многих

языках не могут сочетаться глухие и звонкие согласные, и т. п.). На морфологическом уровне языка сочетаемость зависит от комбинации морфем и различается на формальном и семантическом уровнях. На более высоком уровне языка сочетаемость слов определяется грамматическими, лексическими, семантическими факторами и изучается теорией словосочетаний. Лингвисты обычно связывают словосочетания с лексикологией и синтаксисом. Грамматическая сочетаемость заключается в принадлежности слов к частям речи: "например, для многих языков подчинительное сочетание из двух существительных не характерно, и зависимое существительное либо проявляет тенденцию к адъективации (англ. a *stone* wall), либо его адъективная функция подкрепляется его морфологическим изменением (англ. my brother's friend) или использованием транспонирующего служебного слова (the friend of my brother)". Что касается лексического фактора, сочетаемость определяется избирательностью лексем, ср. «оказать услугу, внимание», но «не оказать заботу, интерес». На семантическом уровне сочетаемость слов проявляется в *семантическом согласовании* —сочетающиеся компоненты не должны иметь противоречащие семы, например, глагол или прилагательное должны сочетаться с существительными одушевленными («люди разговаривают», «больной человек»), так как обозначают действие или свойство живого существа; в противном случае нарушается норма или переосмысляется один из компонентов («больная совесть», «весь дом говорил об этом»).

Изучение сочетаемости помогает идентифицировать языковые элементы, определять их принадлежность к таксономическим классам, выявлять их варианты, определять условия образования переносных значений. Идея сочетаемости как одного из основных факторов структуры и функционирования языка активно разрабатывается в 20 в., в частности после работ Ф. де Соссюра, в фонологии (комбинаторная фонетика, работы Н. С. Трубецкого), в связи с теорией функциональной транспозиции (Ш. Балли) и теорией словосочетаний (В. В. Виноградов). Дистрибутивная грамматика использовала сочетаемость как основу лингвистического анализа, рассматривая её исключительно на

формальном уровне. Интерес к семантической стороне языка побудил изучать закономерности семантической синтагматики, играющей важнейшую роль в образовании смысла высказывания. Сочетаемость изучается как в формальном, так и в семантическом аспектах [Лингвистический энциклопедический словарь, 1990].

1.3. Устойчивые словосочетания

От понятия сочетаемости мы переходим к фразеологизмам, или устойчивым словосочетаниям. В.В. Виноградов пишет [Виноградов 1972:154], что Ш. Балли дал фразеологизмам определение общего характера: «сочетания, прочно вошедшие в язык, называются фразеологическими оборотами». Исследователи В.Л. Архангельский, С.Г. Гаврин, В.Н. Телия определяют устойчивое словосочетание как языковую единицу, для которой характерны такие второстепенные признаки как метафоричность, эквивалентность и синонимичность слову. Но, по мнению Н.М. Шанского [Шанский 1985:223], метафоричность присуща также и многим словам, а эквивалентность — не всем устойчивым сочетаниям. Поэтому включение этих второстепенных и зависимых признаков в определение фразеологизма не совсем корректно. Ученый также подчеркивал, что «правильная дефиниция фразеологизма невозможна без учета его отличий от слова и свободного сочетания».

Правила, по которым формируются устойчивые словосочетания, уникальны, что обусловлено идиоматичностью и принципом экономии. Идиоматичность отличается несколькими особенностями – **переинтерпретация** (*почесать нос* и *стоять на носу корабля*), то есть одно значение создается на основе другого значения, **непрозрачность** (бить баклуши), так как сложно вычислить настоящее значение идиомы из-за отсутствия правила, позволяющего выделить это значение, и **усложнение способа указания на денотат** (*обманывать* и *вешать лапшу на уши*), когда можно сказать одно и то же по-разному. Выражение считается идиоматичным, если в нем есть одна из этих особенностей [Баранов, Добровольский 2014:44-61].

1.4. Классификации словосочетаний

1.4.1. Классификация устойчивых словосочетаний по В.В.Виноградову

Существует множество классификаций, сошлемся, по крайней мере, на следующие: [van der Wouden 1997], [Čermák 2006], [Sag et al. 2002:3-7], [Mel'ˇcuk 1998] и [Виноградов 1972].

В данной работе мы будем опираться на классификацию В.В. Виноградова в связи с ее полнотой и законченностью. Итак, по [Виноградов, 1972] словосочетания делятся на 3 типа.

Фразеологические сращения – это семантически неделимые обороты, «значение которых совершенно независимо от их лексического состава, от значений их компонентов».

К числу сращений относятся, например, такие словосочетания, как *бить баклуши*, *валять дурака*, *поминай как звали*, *во всю ивановскую*, *у черта на куличиках* и др. Из отдельных компонентов сращений нельзя вывести значение всей единицы.

Фразеологические единства – фразеологизмы, общее значение которых вытекает из значений составляющих частей. Большая часть фразеологизмов этого разряда образовалась в результате метафорического переосмысления свободных словосочетаний: *взваливать на плечи*, *видеть насквозь*, *вить гнездо*, *белая ворона*.

Фразеологические сочетания – словосочетания, состоящие из двух знаменательных слов, из которых одно имеет самостоятельное, а другое – связанное значение: *обращать внимание* («внимание» будет всегда иметь одно и то же значение, а «обращать» будет менять смысл в зависимости от словосочетания, в котором оно употребляется: *обращать внимание – обращать в другую веру*), *оказывать помощь*, *впадать в нужду* и др.

Впоследствии Шанский добавил еще один тип – фразеологические выражения. Это «устойчивые в своем составе и употреблении фразеологические обороты, которые не только являются семантически членимыми, и состоят целиком из слов со свободными значениями» [Шанский

1964:201]. Эти выражения делятся еще на два типа - «фразеологические выражения коммуникативного характера» (*хрен редьки не слаще, человек — это звучит гордо*) и «фразеологические выражения номинативного характера»(*кот наплакал*, *руки не доходят*, *куры не клюют*). Поскольку к выражениям коммуникативного характера, например, относятся и пословицы, и крылатые слова, такая классификация не вполне точна. Отсюда можно сделать вывод, что задача распределения фразеологизмов по классам зависит от исследования.

С учетом различных исследований А.Н. Барановым и Д.О.Добровольским была создана классификация, несколько упрощенная, и отвечающая традиции, и включающая в себя новшества. Она состоит из семи типов словосочетаний [Баранов и Добровольский 2014:67-96].

1.4.2. Классификация устойчивых словосочетаний по А.Н. Баранову и Д.О. Добровольскому

1) Идиомы

«Идиомы — это сверхсловные образования, которым свойственна высокая степень идиоматичности и устойчивости». Они выделяются в соответствии с параметрами переинтерпретации, непрозрачности и устойчивости, о которых речь шла выше (*шишка на ровном месте; работать спустя рукава; выпустить джинна из бутылки; не мытьем, так катаньем; сойти с ума; хоть ты тресни*). Среди идиом выделяются те или иные типы:

a) Речевые формулы. Некоторые фразеологизмы рассматриваются с точки зрения их связи с моментом речи. В этом подклассе появляется непосредственная отсылка к коммуникативной ситуации: *старость не радость; где наша не пропадала; не гони лошадей; избави бог; дурак или родом так?; поживем — увидим; как только, так сразу; где уж нам, дуракам, чай пить!*

b) идиомы-комментарии, с их помощью говорящий выражает отношение к происходящему: *дурак и уши холодные; не лаптем щи хлебаем; лед тронулся; дела идут — контора пишет; за что боролись, на то и напоролись.* Помимо комментариев выделяется еще один тип

речевых формул. Полуавтономные речевые формулы синтаксически зависимы, это не законченные предложения: *хоть [ты] убей, хоть [ты] тресни, хоть [ты] лопни,* но несмотря на это, они могут выделяться в речи интонационно и синтаксически.

c) идиомы-перформативы выражают речевые действия**.** Например, идиомы *вот те / тебе крест, зуб даю, век свободы не видать* и т.п. реализуют речевой акт клятвы. В свою очередь, идиомы *чтоб глаза повылезали / вылезли (у кого-л.), чтоб пусто было (кому-л.), чтоб руки отсохли (у кого-л.)* связаны с выражением речевого акта проклятия. Имеются речевые формулы, реализующие и другие типы речевых актов: отказ(*шёл бы ты своей дорогой, нашёл дурака, скатертью дорога, и думать забудь, катись колбаской, много будешь знать — скоро состаришься, спешу и падаю*), просьба(*не в службу, а в дружбу; позолоти ручку!; не корысти ради*), обещание(*будет вам и белка, будет и свисток; дай срок*).

d) Формулы ответа – заранее заготовленные ответы на определенные вопросы: *{– Ну?}— Баранки гну; {– Где?} — У тебя на бороде; {– Откуда?} — От верблюда; {– Куда?} — На кудыкину гору; {– Почему?} — По кочану; {– Говорят...} — Кур доят; {– Привет!} — Привет от старых штиблет; {– Как дела?} — Как сажа бела.* А также **формулы вопроса -** (*дурак или родом так?; какая муха тебя укусила?*

2) **Пословицы** (Примеры: *цыплят по осени считают; без труда не вытащишь и рыбку из пруда; не подмажешь — не поедешь; волков бояться — в лес не ходить; назвался груздем — полезай в кузов*)

«Пословица — это фразеологизм, имеющий структуру предложения, с семантикой всеобщности, выражающий рекомендацию (совет, нравоучение или запрет) и/или объясняющий обсуждаемое положение дел с точки зрения правил наивной логики». Так, в пословицах часто встречаются слова *все, всё, всякий, каждый,* а также употребляется обобщенно-личная форма глагола. Отличительной особенностью пословиц является определенного рода

независимость от контекста или ситуации. Например, Писатель А.Н. Островский использовал в названиях своих пьес много пословиц, и читателями воспринимается это совершенно естественно. Пословицы специфичны, и это позволяет отделить их от других речевых формул. Главное отличие – рекомендательная сила пословиц и отсылка к общему знанию носителей данного языка. Еще несколько особенностей: во-первых, смысл пословицы имеет возможность «расширения», во-вторых, предложение с пословицей должно сочетаться с вводными выражениями: *как известно, как учит народная мудрость* и т.п. Если в предложении они не сочетаются, то перед нами поговорка. Однако, между пословицами и поговорками нет четкой грани.

3) **Грамматические фразеологизмы** (Примеры: *во что бы то ни стало; по крайней мере; по меньшей мере; едва не; хотя бы; потому что; из-за того, что; вследствие того, что*).

Виноградов выделял среди фразеологизмов «союзные речения», включающие предлоги, союзы, указательные местоимения а также некоторые модальные частицы, ср. *до тех пор пока, с тех пор как, между тем как, после того как, подобно тому как, едва только / лишь, чуть лишь*. С формальной точки зрения большую часть фразеологизмов из этой группы составляют соединения служебных слов. «Грамматические фразеологизмы — это неоднословные выражения, которые с содержательной точки зрения характеризуются идиоматичностью значения (т.е. их план содержания не вычисляется по регулярным правилам) и которые связаны с нерегулярным выражением грамматических (в том числе модальных) смыслов и/или представляют собой сочетания различных служебных слов.» Эти фразеологизмы используются в широком списке случаев: при выражении времени (чуть что), пространства (*из-под, из-за, по-над*), частиц (*как раз, ну вот, да и*). Некоторые из них напоминают по функции идиомы (*чуть что, как раз, по крайней мере, по меньшей мере)*, а также выполняют метатекстовые функции. Например, фразеологизм *и вот. Он может выражать вывод из повествования,.— Я здоров как бык. Но эти проклятые семейные сцены...*

*Короче говоря, я поссорился с баронессой — **и вот** я здесь*[А. и Б. Стругацкие. Трудно быть богом], *или вводить новую сцену: Виктор целый год не был в родном городе. **И вот** он снова дома.*

4) **Фразеологизмы-конструкции**

«Фразеологизмы-конструкции — это синтаксически автономные выражения устойчивого состава, в которых пропущены некоторые элементы (актанты — обычные или пропозициональные). Причем фиксированные элементы конструкции, вместе с ее синтаксисом, характеризуются единым значением, приближающимся к лексическому.»

Например:

X — *он и в Африке* X (*Кризис — он и в Африке кризис*; *Работа — она и в Африке работа*; *Блондинка — она и в Африке блондинка*);

***тоже мне* X** (*Тоже мне подарок*; *Охотнички тоже мне*; *Тоже мне Европа*).

Идиоматичность сосредоточена в самой структуре и фиксированной части, однако, на заполняемые места также действуют некоторые ограничения. Например, вместо «*всем глазам глаза»* лучше сказать *не глаза, а глазищи.* Эти и подобные фразеологизмы изучаются в рамах Грамматики конструкций. Существуют также «фразеосхемы», как назвал их Д.Н. Шмелев. Это продуктивные синтаксические конструкции русского языка, которые, однако, не являются фразеологизмами-конструкциями в рассматриваемом понимании. Они отличаются от идиом тем, что слова, используемые в «пустых» местах, используются в прямых значениях. Например, *туча тучей* и *дура дурой*. В первом случае слова туча использовано в переносном значении. Использующий это выражение имеет в виду, что у него плохое настроение. Во втором же словосочетании смысл прозрачен, и в эту конструкцию можно поставить другие слова, выражающие тот же смысл – *дурак дураком*, *болван болваном* и т.п. Таким образом, *туча тучей* – это идиома.

5) Ситуативные клише

«Ситуативные клише — это слабоидиоматичные или неидиоматичные словосочетания, фразеологичность которых определяется преимущественно устойчивостью и прежде всего прямой зависимостью от правил («писаных и неписаных»), действующих в конкретной ситуации.» Такие фразеологизмы употребляются в определенных ситуациях – когда того требует, например, традиция, этикет или какой-либо устав. Ср. *добрый день; до свидания; спокойной ночи; руки вверх; ко мне; стой, кто идет.* Также есть совершенно не идиоматичные выражения – это давно зафиксированные конструкции, изменяющиеся от языка к языку. Например, надпись на упаковке продукта, свидетельствующая о его свежести*:*

рус. *Годен до...*;

англ. *Best before...* букв. «Лучше всего до...»;

нем. *Mindestens haltbar bis...* букв. «Способен храниться по крайней мере до...» или *Zu verbrauchen bis...* букв. «Употребить до...»;

фр. *À consommer de préférence avant le...* букв. «Предпочтительно употребить до...»;

Ситуативные клише воспроизводятся в конкретных ситуациях как единое целое, поэтому они относятся к фразеологизмам. Наиболее близки они к коллокациям.

б) **Крылатые слова**

«Крылатыми словами принято называть различные в структурном отношении устойчивые сочетания слов, в большинстве случаев афористического характера, источник возникновения которых (литературный, публицистический, имеющий мифологическую основу и т.п.) мыслится как общеизвестный».

Отнесение выражений к крылатым словам зависит от знаний людей, их употребляющих. Как квалифицировать, к примеру, такое выражение, как *время жить, время умирать*? Для одних оно относится к крылатым словам потому, что известен его источник — Книга Екклесиаста; для других — потому, что так называется один из романов Э.М. Ремарка; для третьих это вообще не крылатое

слово, а просто слабоидиоматичное устойчивое русское выражение афористического характера. То же касается античных выражений. Те, кто учили латынь, знают латинские крылатые слова, например, *жребий брошен*. Поэтические тексты с большим содержанием крылатых выражений и аллюзий на них могут быть прочитаны по-разному, в зависимости от знаний читателя.

Когда источник высказывания известен, выражение трактуется как цитата и поэтому употребляется именно таким образом: *как говорил Остап Бендер...*

Для русской действительности источником крылатых слов стали басни И.А. Крылова, кинофильмы («Служебный роман»), популярные романы («Двенадцать стульев») .

7) **Коллокации**

«Коллокации — это слабоидиоматичные фразеологизмы преимущественно со структурой словосочетания, в которых семантически главный компонент (**база**) употреблен в своем прямом значении, а сочетаемость со вспомогательным компонентом (**коллокатором**) может быть задана в терминах семантического класса, но выбор конкретного слова предопределен узусом». Примеры: *проливной* [коллокатор] *дождь* [база], *принимать* [коллокатор] *решение* [база], *зерно* [коллокатор] *истины* [база], *ставить под* [коллокатор] *сомнение* [база], *топорная* [коллокатор] *работа* [база], *трескучий* [коллокатор] *мороз* [база]. Коллокатор обычно произволен, и со временем может заменяться на такой же по смыслу. Ср. *полагать надежду* в «Пиковой даме» Пушкина и современную форму *возлагать надежды*: *Она описала ему самыми черными красками варварство мужа и сказала наконец, что* ***всю свою надежду полагает*** *на его дружбу и любезность*. Хотя, например, в коллокации *проливной дождь* конструкция постоянна и ограничена. Очень часто один коллокатор может сочетаться с разными базами. Например, *принимать: принимать решение*, *принимать соболезнования, принимать участие*. Но установить, с каким классом слов сочетается этот глагол, нельзя – выбор базы для возможного коллоката непредсказуем. К примеру, можно сказать *принимать решение*, но нельзя сказать **принимать заключение* (надо сказать *прийти к*

заключению), можно сказать *принимать соболезнования*, но нельзя сказать **принимать сожаления*, соответственно, можно сказать *принимать участие*, но нельзя сказать **принимать членство*, можно только *быть членом. Коллокации группируются на основе лексических функций – выделяются* **коллокации-magn, коллокации-oper-func**, **коллокации-real-fact**, **коллокации-sing, коллокации-mult**.

a) Коллокации-magn являются словосочетаниями с компонентом, необычным способом выражающие смысл **magn,** *жгучий брюнет, закадычный друг, заклятый враг, закоренелый преступник, проливной дождь.*

b) *Коллокации*-oper-func содержат компонент, уникальным образом выражающий смыслы OPER или FUNC, ср. *принимать решение, ставить вопрос, одержать победу, потерпеть поражение, взять реванш*,

c) коллокации-real-fact — это, соответственно, устойчивые словосочетания, нестандартным образом передающие смыслы REAL или FACT (также в комбинации с другими смыслами типа CAUS, FIN), ср. *желание сбывается, справедливость торжествует, долг велит (кому-л. сделать что-л.).*

d) Коллокации-sing передают значение одного экземпляра, а MULT – наоборот, множества. ср. *порыв ветра, кочан капусты и отара овец, стадо коров, косяк трески*.

Также существуют **метафорические коллокации** – один компонент в них употребляется в прямом смысле, а второй – как метафора. Например, *зерно истины* и *червь сомнения.*

Часто коллокации судят на основе критерия семантической связанности, то есть, выражен ли смысл всей конструкцией или отдельными словами.

Коллокации описывают тремя критериями:

- Некомпозиционность. Значение коллокации не состоит прямо из связей смыслов частей коллокации

- неизменность . Многие коллокации не могут быть свободно изменены с помощью добавления лексики или грамматических изменений
- незаменимость. Мы не можем заменить ближайшие синонимы компонентами коллокации.

В корпусной лингвистике коллокации определяют как статистически устойчивые словосочетания на основании близости слов в тексте.

Выводы по главе 1

1. Словосочетания следует отличать от n-грамм. Сочетания всегда осмысленны. Эта осмысленность проявляется в том, что сочетания слов и связи между ними могут иметь как синтагматический, так и парадигматический характер.
2. Сочетаемость - это "свойство языковых единиц сочетаться при образовании единиц более высокого уровня". Сочетаемость существует на всех уровнях языка.
3. Синтагматические словосочетания изучаются в первую очередь в синтаксисе и в лексикологии. Особо выделяют устойчивые словосочетания. Существует много типов устойчивых словосочетаний. И много их классификаций.
4. Особо выделяют коллокации "слабоидиоматичные фразеологизмы преимущественно со структурой словосочетания, в которых семантически главный компонент употреблен в своем прямом значении, а сочетаемость со вспомогательным компонентом может быть задана в терминах семантического класса, но выбор конкретного слова предопределен узусом";
5. В корпусной лингвистике коллокации определяют как статистически устойчивые словосочетания.

ГЛАВА 2. МЕТОДЫ ВЫЯВЛЕНИЯ УСТОЙЧИВЫХ СЛОВОСОЧЕТАНИЙ

2.1. Корпуса текстов как исходный материал для выявления коллокаций

Методика извлечения устойчивых словосочетаний на 99% связана с корпусной лингвистикой, которая напрямую работает с корпусами.

Лингвистический корпус текстов это большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач. Понятие «корпус текстов» включает также систему управления текстовыми и лингвистическими данными, которая называется *корпусным менеджером* (или корпус-менеджером) (англ. corpus manager). Она является специализированной поисковой системой, содержащей программные средства для поиска данных в корпусе, получения статистической информации и предоставления результатов пользователю в удобной форме [Захаров 2005: 5].

Смысл создания и удобство использования корпусов определяется следующими причинами:

1) достаточно большой объем корпуса обеспечивает полноту представления всего спектра языковых явлений;

2) Содержание в корпусе данных разного типа в естественной контекстной форме позволяет использовать их в целях всестороннего и объективного изучения;

3) Возможность использовать однажды собранный корпус многократно в различных целях.

Практически все современные лингвистические исследования и работы по составлению словарей ориентированы на использование корпусов текстов. Современные интеллектуальные программные системы (и их создание), предназначенные для обработки текстов на естественном языке, также требуют

большого массива лингвистических данных. Корпусные данные востребованы в связи с появлением соответствующих технических возможностей.

Признаки хорошего корпуса:

1. репрезентативность - Под репрезентативностью понимается необходимо-достаточное и пропорциональное представление в корпусе текстов различных периодов, жанров, стилей, авторов и т.п. Можно сказать, что данное понятие относительно, и у него много определений, и применительно к общеязыковому (национальному) корпусу это понятие сложно рассчитать и описать строго математически, однако к этому можно и нужно стремиться в процессе создания корпуса;
2. сбалансированность (жанров, стилей, текстов каких-либо авторов и т.д.);
3. уникальная разметка (например, мультимедийная);
4. хорошая документированность;
5. дружественность по отношению к пользователю.

Существует большое количество видов корпусов [Захаров 2005: 5-6].

Наличие текстов не решает различные лингвистические задачи - для адекватного их решения нужно, чтобы в массиве содержалась лингвистическая информация. Для того, чтобы извлечь нужную информацию из текста, например, коллокации, в корпус нужно добавить лингвистическую информацию. Такое действие называется разметкой или лингвистической предобработкой текста. Под лингвистической предварительной обработкой мы имеем в виду морфологическую разметку и синтаксическую разметку на этапе создания корпуса и снятия неоднозначности, анализ и устранение неоднозначности на уровне морфологии и синтаксиса. *Разметка* (tagging, annotation) заключается в приписывании текстам специальных меток , или тэгов(tags): экстралингвистических (сведения об авторе и сведения о тексте: автор, название, год и место издания, жанр, тематика - *метаразметка*), структурных (глава, абзац, предложение, словоформа) и собственно лингвистических маркеров. Лингвистические тэги содержат в себе

информацию о лексических и грамматических свойствах компонентов текста. Характер разметки обычно определяет и способ использования данного корпуса. Существуют различные лингвистические типы разметки.

Для извлечения коллокаций важнее всего морфологическая, синтаксическая и семантическая разметки.

Как правило, лингвистическая предварительная обработка не является обязательной для извлечения коллокаций, особенно при работе с языками с простой морфологией (например, английский), или, если мы ориентируемся, например, только на фиксированные соседние и немодифицируемые словосочетания. Тем не менее, если мы имеем дело со сложной морфологией (например, в русском языке) и если мы хотим извлечь синтаксически ограниченные словосочетания со свободным порядком слов, эта информация является весьма полезной. Языковая информация также может быть использована на последующей стадии для фильтрации потенциальных коллокаций и выявления дополнительных особенностей в методах, сочетающих статистические и лингвистические данные в более сложных моделях [Pecina 2009:27-28].

На данном этапе информация о тексте, морфологических категориях, и синтаксисе предложения формируется в целях выявления потенциальных коллокаций и всех их вхождений - независимо от формы слов и позиции в предложении.

2.2. Коллокации и их извлечение

Автоматическое извлечение коллокаций обычно выполняется как процесс, состоящий из нескольких шагов [Evert and Kermes 2003:83-86]:

Во-первых, корпус в виде набора машиночитаемых текстов на одном языке лингвистически предварительно обрабатывается (как уже было сказано выше) - размечается морфологически и, возможно, синтаксически и снимается неоднозначность.

Во-вторых, все сочетания, которые могут быть коллокацией выявляются, и их статистика встречаемости извлекается из корпуса.

В-третьих, кандидаты фильтруются для повышения точности (на основе грамматических моделей и / или частоты встречаемости).

В-четвертых, выбирается мера ассоциации и применяется к статистическим данным встречаемости, полученным из корпуса.

И, наконец, кандидаты в коллокации классифицируются в соответствии с количественной оценкой их сочетаемости и эта оценка сравнивается с определенным порогом - кандидаты выше этого порога классифицируются как словосочетания, кандидаты ниже этого порога - не являются словосочетаниями.

Задача извлечения коллокаций далее сводится к ранжированию кандидатов в коллокации. Цель состоит не просто в извлечении ограниченного набора словосочетаний из данного корпуса, а в ранжировании всех потенциальных словосочетаний в зависимости от силы связи элементов словосочетания, так что те кандидаты, в которых наблюдается наиболее крепкая связь, оказываются в верхней части списка [Pecina 2009:26].

Не следует также забывать, что слова, которые имеют тенденцию к расположению рядом друг с другом, в любом случае не могут быть найдены в произвольном порядке, поскольку существуют грамматические правила языка. Существуют также методы, учитывающие синтаксическую природу коллокаций.

Б. Дай утверждает, что лингвистические знания резко улучшают качество "стохастических" (случайных) систем [Daille 1994 192]. Одним из методов учета синтаксиса являются так называемые ворд скетчи (эскизы слов), которые представляют собой списки статистических сочетаний, где каждое слово имеет по отношению друг к другу синтаксическую связь [Kilgarriff, Tugwell 2004].

Кроме того, оценка силы связи зависит от типа единиц (лемм или словоформ), статистика которых используется для расчетов. Иногда извлечение коллокации статистическими мерами должна производиться на уровне словоформы, а не на уровне лемм. Анализ, описанный в [Захаров, Хохлова 2014: 340], показал, что в некоторых случаях показатель силы связи для

словоформ получает значительно большее значение для всех мер ассоциации, причем таких словосочетаний много.

Само количество вычисленных коллокатов и значение меры ассоциации также зависят от «диапазона» между базой и коллокатом, который был выбран для вычислений. Когда диапазон увеличивается, помимо значимых синтагм, слова из общего лексико-семантического поля находятся в качестве кандидатов коллокации [Захаров 2017:2-3].

2.3.Факторы, от которых зависит качество работы методов

1. Исследуемый материал:
   a) тип конструкций, которые мы хотим выявить;
   b) язык, для которого мы это делаем;
   c) область знаний, к которой относятся устойчивые словосочетания (в художественной литературе будет много метафор и т.д., в научной литературе много терминов);
2. Характеристики корпуса, с которым мы работаем:
   a)размер
   b)репрезентативность и сбалансированность
   c)уровень анализа
   d)«изощренность» дистрибутивно-статистического аппарата,
   e)учет зависимости между методами и текстовым материалом.

2.4.Меры лексической ассоциации

В настоящее время существует несколько способов рассчитать силу связи частей коллокации. Естественно предположить, что одним из способов определения устойчивости словосочетания является частота их совместной встречаемости. Встречаемость, в свою очередь, связана с частотой отдельных компонентов коллокации. Было создано много формул (или заимствовано из других наук) для интеграции различных факторов, которые определяют связь между компонентами коллокации. Обычно такие формулы называются мерами ассоциации. Большинство из них основано на сравнении частот для пар слов,

извлеченных из фактического корпуса с относительными частотами, взятыми из гипотетического корпуса, в котором все слова случайно расположены. Это делается для выявления статистически значимых колебаний между наблюдаемыми и ожидаемыми частотами [Dunning 1993: 61-74].

«Меры ассоциации – статистические формулы, вычисляющие силу синтагматической связи элементов в составе устойчивого словосочетания на основе частоты совместной встречаемости, частот в данном корпусе каждого отдельного слова и других характеристик»[Захаров, Масевич 2014:49].

Меры ассоциации часто основаны на гипотетическом статистическом критерии. Работает это так. Есть две гипотезы, нужно выбрать одну правильную. При нулевой гипотезе – u и v независимы(где u и v обозначают лексические элементы, отображаемые в таблице сопряженности). При альтернативной гипотезеH1 - u и v взаимно зависимы. H1 выбирается системой. Если гипотеза нулевая, то кандидаты не могут быть названы коллокацией.

Ошибки, которые могут быть сделаны при работе с гипотезами:

Тип 1 – в случае, когда неправильно отвергли нулевую гипотезу, хотя она на самом деле верна (кандидат неверно считается коллокатом, это ложный результат);

Тип 2 – когда не отвергли нулевую гипотезу, хотя она неверная (кандидат не засчитывается как коллокат, хотя должен, это также ложный результат).

Мера может быть односторонней и двусторонней. Двусторонняя мера не различает положительную и отрицательную силу связи.

Статистический критерий может быть параметрическим и непараметрическим:

- Параметрические критерии (например, t-score, z-score, log- likelihood) включают числовые данные, и чаще всего их использование предполагает, что данные нормально или биноминально распределены.
- непараметрические критерии (например, χ^2) включают в себя порядковые данные (ранги) и являются более эффективными, чем параметрические

критерии, там, где не соблюдены некоторые условия относительно совокупности.

Существуют различные меры, основанные на вычислении степени близости слов в тексте. P.Pečina приводит 82 меры, описывает их математические основы, включая их формулы и ключевые ссылки [Pecina 2009: 44-45, 48]. Наиболее популярными мерами, по-видимому, являются MI, t-score и log- likelihood.

Лексические меры ассоциации применяются к вхождению ключевого слова (узла). Список кандидатов, ранжированных по количественным значениям мер, является результатом всего процесса. Верхняя часть списка представляет собой словосочетания, которые, как предполагается, имеют наибольшую связь друг с другом и, следовательно, являются наиболее вероятными кандидатами на коллокации. В целом, все они учитывают частоту совместной встречаемости ключевого слова (узла) и его коллоката, тем самым отвечая на вопрос о том, насколько случайна сила связи между соседними словами. Но формулы отличаются друг от друга, и они демонстрируют разную силу связи для одного и того же сочетания, поэтому коллокационные ранги, полученные разными мерами, не совпадают. Известно также, что некоторые меры выдают аналогичные результаты, а другие значительно отличаются [Křen 2006: 246-247].

T-score и z-score не рекомендуется использовать для низкочастотных кандидатов. Известно, что t-score извлекает наиболее частые коллокации. z-score не используют на малых выборках, для этого лучше применить t-score.

Log--likelihood предпочитают использовать, так как эта мера хорошо показывает себя на всех размерах корпусов, а также продвигает менее частотных кандидатов.

Напротив, мера MI позволяет выявить низкочастотные терминологические сочетания из нескольких слов и имена собственные. Кроме того, следует отметить, что частота совместной встречаемости также может быть хорошим показателем , но она имеет недостаток, заключающийся в том,

что не может идентифицировать редкие термины [Daille 1994: 172-173]. χ^2 не использует нормальное распределение. Данная мера менее чувствительна к низким частотам. χ^2 не точен, когда выборка маленькая [Seretan 2011:43].

Несмотря на такое разнообразие мер и их подробное описание, имеются некоторые неразрешенные задачи. Есть идея, что для разных функциональных стилей нужны разные меры ассоциаций. Ягунова Е.В. и Пивоварова Л.М. в статье "Природа коллокаций в русском языке"сравнивали работу мер ассоциаций на материале коллокаций в новостных текстах. Они выяснили, что мера MI больше всего подходит для выявления терминов, объектов, сложных номинаций. T-score, напротив, лучше работает при выделении «общеязыковых устойчивых сочетаний» (производных служебных слов, дискурсивных слов) и «устойчивых конструкций», где и те, и другие характеризуют именно стилистические особенности текстов рассматриваемого типа (в данном случае – новостных текстов) [Ягунова, Пивоварова 2010].

Также существует мнение, что можно использовать сразу все меры, а потом найти их среднее арифметическое их рангов.

Есть еще другие подходы, как использовать меры:

- Исследовать, являются ли биграммы (Или триграммы) разрывными или нет. Это значит, что нас интересует, вклиниваются ли между словами другие слова, но такие подсчеты более сложны.
- Использовать синтаксический метод, метод шаблонов. Указывается, какие должны быть коллокации – какая часть речи, согласование, разрывные или нет. Об этом можно найти информацию в документации NoSketch Engine [https://www.NoSketchEngine.co.uk/documentation]. В языке, например, могут встречаться сочетания A+N, N+N и т.д. Первое – слова согласовываются, могут быть разрывными (много других прилагательных между ними). Возможные варианты сочетания N+N - это может быть сочинительная связь (союз и), второе слово может быть в родительном падеже или, если между ними предлог, то зависимое слово будет стоять в падеже согласно предлогу.

Отдельные шаблоны можно записывать не только в терминах синтаксиса, но и семантики, если у нас есть семантически размеченный корпус.

2.5. Классификация мер ассоциации

В нашей работе мы использовали инструмент NoSketch Engine. NoSketch Engine - это запросная веб-система корпусов, которая поддерживает ряд функций на основе морфологически аннотированных текстов. Эти функции включают в себя конкорданс, частотные списки, распределительный тезаурус и word sketches (ворд скетчи, одностраничная сводка грамматического и разговорного поведения слова). Ворд скетчи могут восприниматься как типичные фразы, определяемые, с одной стороны, синтаксисом, который ограничивает «коллективность» слов в данном языке, а с другой - вероятностью, тесно связанной с использованием слов. [Kilgarriff 2014:105-116].

В NoSketch Engine реализовано 7 мер: T-score, MI, MI3, log-likelihood, min.sensitivity, log-Dice, MI log_f. Дадим их краткую характеристику [https:// www.sketchengine.co.uk/documentation/statistics-used-in-sketch-engine].

Условные обозначения:

N– размер корпуса,

f_A– сколько раз встретилось ключевое слово во всем корпусе (конкорданс),

f_B – сколько раз встретился коллокат во всем корпусе,

f_{AB}– сколько раз встретился коллокат в конкордансе (количество вхождений)

T-Score

T-score выражает определенность, с которой мы можем утверждать, что существует связь между словами, то есть их совпадение не является случайным. Значение зависит от частоты всей коллокации, поэтому очень частые словосочетания имеют тенденцию достигать высокого значения T-score, несмотря на то, что не являются значимыми в качестве коллокаций.

T-score использует критерий Стьюдента. Является односторонней параметрической мерой, которая предполагает, что выборка сделана из нормально распределенной совокупности. Он сравнивает среднее значение

выборки, X (то есть, наблюдаемое среднее), со средним значением выборки, μ (т.е. средняя оценка в предположении нулевой гипотезы). Высокая разность указывает на то, что образец не был составлен из совокупности, в которой имеет место нулевая гипотеза. Таким образом, в случае лексических данных, высокое значение T свидетельствует о том, что образец не был составлен из совокупности, в которой две лексические единицы являются независимыми, и, следовательно, указывает на сильную положительную ассоциацию

В большинстве случаев показатель T-score является более надежным или более полезным, чем показатель MI.

Рассчитывается по формуле:

$$\frac{AB - \frac{AB}{N}}{\sqrt{AB}}$$

MI Score

Mutual Information (взаимная информация) отражает степень совпадения слов по сравнению количества раз, когда они появляются отдельно. Показатель MI сильно зависит от частоты, низкочастотные слова обычно имеют высокий показатель MI, что может вводить в заблуждение. Вот почему NoSketch Engine позволяет установить лимит, а слова с частотой ниже этого предела не будут включены в расчет.

Рассчитывается по формуле:

$$\log\text{-}_2\frac{ABN}{AB}$$

MI³-Score отличается тем, что формула возведена в куб**.** **MI³** использует более высокий показатель в числителе, чтобы еще больше увеличить оценку силы связи высокочастотных слов, представляет собой чисто эвристический подход.

$$\log\text{-}_2\frac{3ABN}{AB}$$

log-likelihood относится к мерам наибольшего правдоподобия, известна как логарифмическая функция правдоподобия. Получает результат на основе таблицы вероятности**.**

Рассчитывается по формуле:

$2*(xlx(f_{AB})+xlx(f_A-f_{AB})+xlx(f_B-f_{AB})+xlx(N)+xlx(N+-f_{AB}\ -f_A\ -f_B)-xlx(f_A)-xlx(f_B)-xlx(N-f_A)-xlx(N-f_B))$,

где $xlx(f)$ это $f\ln(f)$.

Min.sensitivity

Еще одной мерой ассоциации, относящейся к группе точечной оценки силы связи, которая не получила широкого распространения, является минимальная чувствительность (MS).

Рассчитывается по формуле:

$$\min(\frac{AB}{B}, \frac{AB}{A})$$

Коэффициент **Dice**, мера, также относящаяся к группе точечно й оценки, интересна тем, что, как считают некоторые лингвисты (Smadja), что она идентифицирует пары слов с особенно высокой степенью силы связи (т. е. с силой связи почти 100%).

Рассчитывается по формуле:

$$\frac{2AB}{A + AB}$$

MI. log_f

Рассчитывается по формуле:

MI-Score*$\ln(f_{AB}+1)$.

Выводы по главе 2

1. Методика извлечения коллокаций напрямую связана с корпусной лингвистикой. Корпуса текстов и корпусные менеджеры – незаменимый материал и инструмент для выявления коллокаций. Для выявления коллокаций на базе корпуса большое значение имеют лингвистическая

разметка текстов и предобработка данных в корпусе, а также снятие неоднозначности.

2. Устойчивость словосочетаний (коллокаций) определяется разными факторами: лексическими, грамматическими, узуальными.

3. Есть несколько факторов, от которых зависит качество работы методов извлечения коллокаций: репрезентативность и сбалансированность корпуса, «изощренность» дистрибутивно-статистического аппарата, учет зависимости между методами и текстовым материалом.

4. Меры ассоциации – статистические формулы, вычисляющие силу синтагматической связи элементов в составе устойчивого словосочетания на основе частоты совместной встречаемости. Существует большое количество различных мер и исследований по их оценке.

5. Автоматическое извлечение коллокаций представляет собой несколько последовательных шагов. Первый - предобработка текстов, из которых извлекаются коллокации. Второй шаг- все потенциальные коллокации и статистика их встречаемости выявляются на основе грамматики заданного языка, третий шаг- кандидаты в коллокации фильтруются для повышения точности, четвертый шаг - к статистическим корпусным данным применяется мера лексической ассоциации и затем кандидаты ранжируются в соответствии с порогом истинности.

ГЛАВА 3. МЕТОДЫ И КРИТЕРИИ ОЦЕНКИ

К автоматическим способам оценки автоматизированных систем относится вычисление полноты, точности, F-меры и средней точности, которые часто используются многими исследователями.

Процедура извлечения коллокаций включает в себя получение списка кандидатов в устойчивые словосочетания. Оценка результатов выделения коллокаций заключается в том, чтобы среди выделенных кандидатов определить настоящие коллокации и оценить результаты по выбранным метрикам.

Способы оценки можно делить по нескольким основаниям [Ramisch 2012:70-72]:

1. по природе используемых мер:
   a. количественные. Такой вид оценки предполагает использование полноты, точности, F-меры, а также среднее арифметическое точности;
   b. качественные. При качественной оценке производится обзор полученного списка с учетом таких критериев, как частеречные цепочки, частотное распределение и контекст. Такая оценка возможна как вручную, так и при помощи статистического анализа. Обычно она имеет рекурсивный характер – получение списка кандидатов, его оценка, учет ошибок, прогон заново, и т.д. до получения приемлемого результата;
2. по типу доступных ресурсов для оценки:
   a. оценка вручную. При такой оценке носители языка или эксперты в той предметной области, для которой выделялись устойчивые словосочетания, вручную оценивают получившийся список кандидатов в устойчивые словосочетания и отбирают среди них действительные и ложные устойчивые словосочетания. К сожалению, такой вид оценки очень затратный по временным и человеческим ресурсам;
   b.автоматическая оценка. Такой вид оценки проводится при наличии золотого стандарта, который является некоторым эталоном, списком, который содержит только «правильные» устойчивые словосочетания. Для подобной

оценки необходимым условием должно быть полное или значительное покрытие устойчивых словосочетаний золотым стандартом.

3.1.Точность и полнота

Для нашего эксперимента потребуются такие формулы, как точность (precision) и полнота (recall), которые являются метриками и используются при оценке большей части алгоритмов извлечения информации. Они могут использоваться также в качестве основы для производных метрик, например, для F-меры.

Точность – это доля документов (в нашем случае коллокаций), действительно принадлежащих данному классу (совпадающих с золотым стандартом) относительно всех документов, которые система отнесла к этому классу. Полнота – это доля найденных классификатором документов (коллокаций), принадлежащих классу относительно всех документов этого класса в тестовой выборке.

Эти значения рассчитываются на основании таблицы контингентности, которая составляется для каждого класса отдельно.

Категория i		Экспертная оценка	
		положительная	отрицательная
Оценка системы	Положительная	TP	FP
	Отрицательная	FN	TN

В таблице содержится информация, сколько раз система приняла верное и сколько раз неверное решение по документам заданного класса. А именно:

- *TP*(true positive) — истино-положительное решение;
- *TN* (true negative)— истино-отрицательное решение;
- *FP* (false positive)— ложно-положительное решение;
- *FN* (false negative)— ложно-отрицательное решение.

Тогда, точность и полнота определяются следующим образом:

$$Precision=\frac{TP}{TP+FP}$$

$$Recall=\frac{TP}{TP+FN}$$

В работе потребуется так называемая "условная" полнота, так как в нашем материале невозможно найти нужные компоненты для вычисления истинной полноты. Пример вычисления условной полноты приведен в главе 4.

3.2.F-мера

F-мера, или гармоническое среднее, часто используется как единая метрика, объединяющая в себе метрики полноты и точности, являясь, таким образом, их усредненным значением.

Вычисляется по следующей формуле:

$$F = \frac{\text{Precision} + \text{Recall}}{2}$$

F-мера это необязательно среднее арифметическое. Точности и полноте в этой формуле можно приписывать различные коэффициенты, в зависимости от целей исследования, но в нашей работе мы этого не делали. Именно поэтому мы используем в формуле F-меры среднее арифметическое.

3.3.Средняя точность

Средняя точность (mean average precision) учитывает приоритет словосочетаний, имеющих высокий ранг перед словосочетаниями, находящимися в конце списка, позволяя более точно оценить качество работы того или иного метода выявления устойчивых словосочетаний.

3.4.Составление золотого стандарта

Для оценки автоматизированных систем нужен золотой стандарт или список действительных устойчивых словосочетаний и метод, который составляет ранжированный список возможных устойчивых словосочетаний.

Сложность составления золотого стандарта заключается в том, что часто бывает неясно, откуда взять данные, и часто это очень трудоемкая задача. Поэтому иногда составляют золотой стандарт только для конкретного

эксперимента. Но важность и удобство золотого стандарта давно широко признаны.

Выводы по главе 3

1. Метрики оценки включают в себя вычисление полноты, точности, F-меры, средней точности.
2. Для адекватной оценки результатов извлечения требуется наличие так называемого золотого стандарта.
3. В качестве золотого стандарта используются данные имеющихся словарей лексикографических материалов, предварительно размеченные корпусы и экспертная оценка получаемых результатов.
4. В нашей работе мы будем оценивать получаемые результаты в целом и эффективность отдельных мер ассоциации на основе оценок экспертов и золотого стандарта на базе словарей, подготовленного нами специально для данной работы.

ГЛАВА 4. ОЦЕНКА АВТОМАТИЧЕСКИХ МЕТОДОВ ИЗВЛЕЧЕНИЯ КОЛЛОКАЦИЙ

Для исследования разных методов выявления коллокаций, а именно статистических, мы провели ряд экспериментов.

Цель экспериментов – оценить эффективность статистических мер путем сравнения результатов автоматического выделения коллокаций с «золотым стандартом», а также экспертной оценкой.

Задачи:

- Выбрать метрики для их оценки
- Создать золотой стандарт
- Оценить меры

Инструменты:

Исследование проводилось с помощью системы NoSketch Engine на корпусе Araneum Russicum Russicum Maius объемом в 1,20 млрд токенов с использованием различных словарей русского языка. Для извлечения коллокаций есть готовые инструменты, в частности, функция Collocations, встроенная в NoSketch Engine (NoSkE). Данная система управления корпусом способна работать с чрезвычайно крупными корпусами и способна предоставить платформу для вычисления широкого диапазона лексической статистики.

Система спроектирована по модульному принципу. Она содержит библиотеку индексирования для сжатия, создания и извлечения индексов, модуль оценки запросов с классами для различных операций запроса, анализатор запросов, который преобразует запросы в абстрактные синтаксические деревья, набор инструментов командной строки для построения и обслуживания корпусов, два графических пользовательских интерфейса. [Rychlý 2007:65-70].

Материалы:

Эксперимент проводился на корпусе Araneum Russicum Russicum Maius объемом в 1,20 млрд токенов. Этот корпус был разработан в рамках проекта [Benko 2014: 257-264] и также доступен в сайте братиславского университета

(http://ucts.uniba.sk). Наш выбор пал именно на данный корпус, так как, например, в корпусе НКРЯ невозможно выделение коллокаций. Название корпусов происходит от латинского названия "лингвистически нейтральный", и обозначения языка, соответственно, Araneum Anglicum, Araneum Germanicum, Araneum Russicum и т.д. Каждый корпус существует в четырех вариантах размеров, из которых основных два: Maius (от лат. "больше") объемом 1,2 млрд. токенов и Minus (от лат. "меньше"), составляющий 10% от корпуса Maius. Также есть версия Maximum, содержащая столько данных, сколько можно загрузить из Интернета для конкретного языка, а их размер в основном определяется конфигурацией сервера.

Сбор всех исходных данных для корпусов Aranea осуществляется с помощью SpiderLing, веб-сканера, оптимизированного для сбора текстовых данных из Интернета. Система содержит встроенный модуль кодирования символов (chared.py) и распознавания языка (trigrams.py), а также инструмент для удаления шаблона (jusText). В корпусе устранены дубли. Для автоматического аннотирования текста используется теггер под названием TreeTagger. Чтобы упростить создание совместимых грамматик, все собственные тэг-таблицы внесены в универсальный набор тегов Araneum. Для всех корпусов были написаны совместимые скетч-грамматики. Их основная идея состоит в том, чтобы иметь одинаковое количество граммем (и отображать таблицы ворд скетчей) для всех классов слов во всех языках [Benko 2014, 257-264].

Методы:

Метод исследования заключается в сопоставлении статистических мер и других параметров, предназначенных для извлечения словосочетаний.

Приводим подробный план работы:

- Выбрать слова разных частей речи для эксперимента;
- Выбрать словари для составления золотого стандарта;
- Из словарных статей взять коллокации, в которых встречаются эти слова; это и будет золотой стандарт;

- Сравнить золотой стандарт с коллокациями, извлеченными мерами ассоциации;
- Вычислить корреляцию между мерами;
- Попросить экспертов оценить выданные мерами коллокации;
- Выбрать лучшую меру по результатам эксперимента.

В ходе эксперимента были использованы следующие словари[1]:

- Словарь сочетаемости русского языка под редакцией Денисова и Морковкина,
- МАС (малый академический словарь),
- СИБАС (Сибирский ассоциативный словарь русского языка)
- БТС (Большой Толковый Словарь),
- Русский ассоциативный словарь,
- Ассоциативная база данных УрРАС,
- Словарь-тезаурус ЕВРАС,
- Толковый словарь Ушакова,
- Толковый словарь Ожегова,
- Толковый словарь Ефремовой,
- Лингво-страноведческий словарь Русские фразеологизмы В.П. Фелициной и В.М. Мокиенко;
- Фразеологический словарь русского языка под редакцией А.И. Молоткова.

4.1.Эксперимент

С помощью названных выше словарей был создан "золотой стандарт" ad hoc под 7 слов, которые мы анализировали (*сердце, вода, рука, белый, скакать, семь, свой*). См. приложения 1-7. Мы также сравнили, в скольких словарях встретилась та или иная коллокация. Например, словосочетание *питьевая вода* встретилось в 5 из 12 словарей*, правая рука*- 10 из 12*, доброе сердце* - 9 из 12 (таблица 1).

[1] Библиографические описания словарей см. в Источники

Таблица 1. Словосочетания для слова вода и их присутствие в словарях золотого стандарта (фрагмент)

	словарь сочетаемости	МАС	Сибас	БТС	русский ассоциативный словарь	Русский ассоциативный словарь	Ассоциативная база данных УрРАС	Словарь-тезаурус ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремова	Словарь Молоткова	Словарь Мокиенко
питьевая	+		+		+	+			+				
горячая	+		+		+	+							
живая		+		+	+	+							+
как рыба в воде		+		+					+			+	
как с гуся вода		+		+					+			+	+

В нашем материале есть также очень редкие коллокации. Говоря «редкие», мы имеем в виду словосочетания, встретившиеся лишь в одном или двух словарях. Для слова *рука* мы выбрали из всех словарей 246 коллокаций со словом *рука*, для слова *вода* - 328, для слова *сердце* – 245, для слова *белый* – 90, для слова *скакать* – 62, для слова *семь* – 24, для слова *свой* – 75. Мы полагаем, что имеем основания опираться на данную подборку в качестве золотого стандарта. Более подробно коллокации можно посмотреть в приложениях 1-7.

В NoSketch Engine реализовано 7 мер: T-score, MI, MI3, log-likelihood, min.sensitivity, log-Dice, MI. log- _f. Из них для своей работы мы выбрали 4 меры. Этот выбор обусловлен тем, что:

1) Из трех MI-подобных мы взяли MI3, считающуюся наиболее оптимальной;

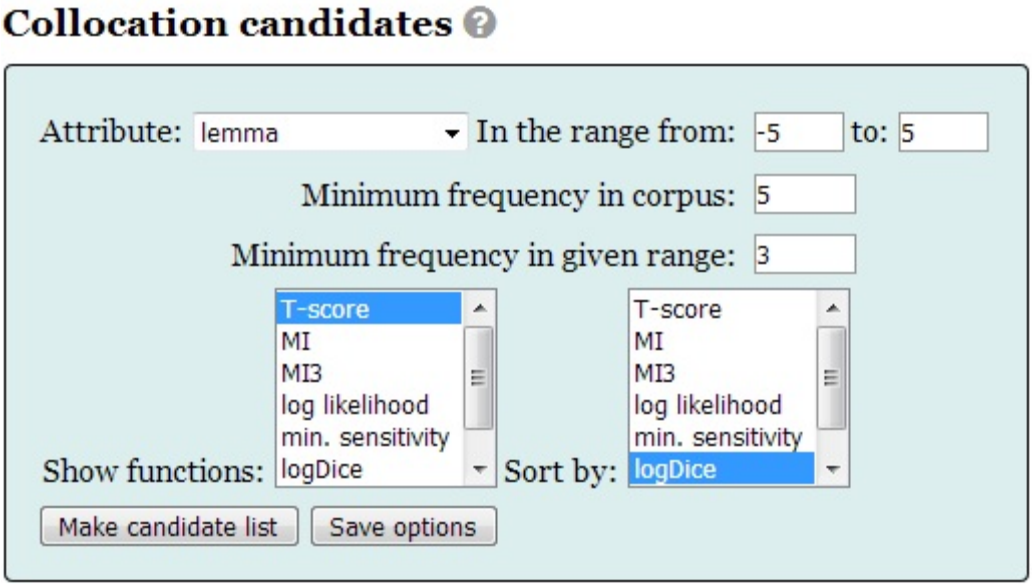
2) На мере log-Dice, которая в NoSketchEngine вообще является основной, строятся основные сервисы NoSketchEngine, такие как wordsketches, thesaurus, Differences;

3) Мера T-score и log-likelihood являются противоположными MI .

Коллокации могут вычисляться для разных диапазонов, поэтому нужно с ними определиться. Этот вопрос не такой простой, как кажется с первого взгляда, поэтому даже может стать темой отдельного исследования. Само число вычисленных коллокатов и значение меры ассоциации также зависят от «диапазона» между ключевым словом и коллокатом, который был выбран для вычислений. Когда диапазон увеличивается, помимо значимых синтагм система находит в качестве кандидатов коллокации слова из общего лексико-семантического поля. По умолчанию в NoSketch Engine предлагается диапазон -5..+5, но мы считаем, что при его использовании выдается много «шума», то есть много коллокаций, не являющихся синтагмами. Однако, при очень маленьком диапазоне также могут возникнуть проблемы. В статье Efficiency of the Sketch Engine grammar приводится пример словосочетания «*запуск двигателя по будильнику*», при этом система выделяет коллокацию "*двигатель по будильнику*", которая является бессмысленной. Такая проблема возникает по причине заданного маленького диапазона, не охватывающего рядом стоящие слова [Khokhlova, Zakharov 2016]. Мы в своей работе нашли оптимальный диапазон от-3 до 3, хотя иногда и он варьируется в зависимости от части речи. В корпусной лингвистике вообще считается, что для большинства случаев это не сильно принципиально.

В системе NoSketch Engine есть инструмент Collocations, с помощью которого пользователь имеет возможность извлекать коллокации. Для этого нужно ввести запрос в основное окно корпуса, получить результат и перейти во вкладку Collocations. В открывшемся окне (см. рисунок 1) задать параметры извлечения (выбрать меру, сортировку, диапазон, минимальную частоту и т.д.).

Рисунок 1. Инструмент извлечения коллокаций в NoSketch Engine



Б ы л и исследованы кандидаты в коллокаты для слов *рука, вода* и *сердце* с диапазонами от 0 до 1,от 0 до 2,от 0 до 3,от -1 до 0, от -2 до 0, от -3 до 0 и для слов белый от 0 до +3, скакать от -1 до 1, семь от -3 до 3, свой от -1 до 1. При этом мы сравнили выдачу меры T-score с при диапазонах от 0 до 1 и от 0 до 3 для слова *белый*. Можно видеть, что результаты одинаковые, то есть размер диапазона не очень влияет на работу мер ассоциации. См. таблицу 2.

Мы сравнили наш золотой стандарт с коллокациями, выданными разными мерами, проанализировав, встречаются ли в выдаче мер коллокаты, извлеченные нами из словарей. Для примера возьмем список коллокатов для слова *сердце* (для всех семи слов проделано то же самое). Словосочетание *доброе сердце* встретилось в 9 словарях (Словарь сочетаемости, Малый академический, Сибирская ассоциативная база и т.д.), и все 4 меры его выдали. Словосочетание *с замиранием сердца* встретилось в 3 словарях и выдано 2 мерами (см. таблицу 3). И далее по списку. Полный список представлен в приложении 1.

Полные списки коллокатов каждой меры совпадают, но мы используем только верхнюю, более релевантную их часть, поэтому они отличаются рангами коллокатов. То есть, может оказаться так, что в выдаче какой-либо меры хорошие коллокаты (то есть коллокаты из золотого стандарта) "спустились" в самый низ, но остальные меры выдали их в начале своих списков. Как правило, большая часть «хороших» коллокатов оказывается в верхней части списка.

Поэтому кажется целесообразным оценивать только эту верхнюю часть. Мы определяли пороговое значение каждой меры эмпирическим путем, хотя есть исследования, предлагающие такие пороговые значения (напр., Ф. Чермак для меры MI предлагает порог, равным 8 [Čermák 2006:223-248]).Была сделана попытка проверить эту гипотезу на нашем материале. Мы сравнили две таблицы - верхняя часть списка (значение log-Dice ≈9-8) и нижняя (значение log-Dice ≈3), таблицы 4 и 5 соответственно. В первой таблице из 10 коллокатов 6 составляют устойчивые словосочетания (*левая рука*, *палец руки*, *правая рука*, *взять в руки*, *под руку* и т.д.), согласно мнению экспертов (подробнее об экспертной оценке речь пойдет далее). Во второй таблице, напротив, из 10 кандидатов ни один не формирует устойчивое словосочетание. Из этого следует, что не стоит рассматривать нижнюю часть списков коллокатов, так как коллокаты, образовывающие хорошие устойчивые словосочетания, в основном сосредоточены в верхней части списка.

Таблица 2. Диапазоны меры T-score для слова *белый*

	снег	флаг	фон	халат	цвет	человек	шум	море	песок	золото	глина	конь	палет	краска	роза	рубашка	полоса	список	одежда	стена	мрамор	шоколад	город	тигр	зал	порошок	экран	голубь	потолок	сахар	зависть	акула	зарплата
От 0 до 1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
От 0 до 3	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Таблица 3.Коллокаты для слова *сердце*.

	словарь сочетаемости	MAC	Сибас	БТС	Русский СОПОСТАВИТЕЛЬНЫЙ ассоциативный словарь	Ассоциативная база данных УрРАС	Словарь-тезаурус ЕВРАС	словарь ушакова	словарь ожегова	словарь ефремова	молотков	Мокиенко	log-dice	MI3	T-score	log-likelihood
биение	+		+										+	+	+	+
доброе	+		+	+	+	+	+	+	+		+		+	+	+	+

завоевать												+	+	+	+
Как ножом по сердцу		+		+						+					
Львиное	+		+	+			+					+	+	+	+
Ничто не шевельнулось в моём с.				+											
остановка												+	+		
положа руку на сердце	+	+		+				+	+	+					
Предложение руки и сердца	+	+		+				+				+	+	+	+
разбитое			+		+	+	+					+	+		
с замиранием		+		+						+			+		+
собачье			+		+		+					+	+	+	+
Холодное			+		+		+					+	+	+	+
Храброе			+				+							+	+

Таблица 4. Кандидаты в коллокаты из верхней части списка log-Dice

	Частота совместной встречаемости	log-Dice	Эксперт 0	Эксперт 1	Эксперт2	Эксперт3	Среднее арифетическое	Стандартное отклонение
держать	13124	9,734	0	0	1	1	0,5	0,57735027
ногами	13313	9,540	2	2	0	2	1,5	1

левый	8235	9,109	2	2	2	2	2	0
палец	8056	9,068	0	0	0	0	0	0
правый	922	9,042	2	1	2	2	1,75	0,5
рука	12985	8,977	0	0	0	0	0	0
взять	9588	8,877	2	0	1	0	0,75	0,95742711
под	16757	8,761	1	1	1	1	1	0
свой	46443	8,728	2	2	2	2	2	0
кисть	3861	8,142	1	1	1	1	1	0

Таблица 5. Кандидаты в коллокаты из нижней части списка log-Dice

	Значение совместной встречаемости	log-Dice	Эксперт 0	Эксперт 1	Эксперт2	Эксперт3	Среднее арифметическое	Стандартное отклонение
располагаться	200	3,756	0	0	0	0	0	0
определенный	287	3,756	0	0	0	0	0	0
паспорт	199	3,753	0	0	0	0	0	0
христос	192	3,750	0	0	0	0	0	0
ухаживать	181	3,750	0	0	0	0	0	0
хоть	214	3,747	0	0	0	0	0	0
называться	213	3,746	0	0	0	0	0	0
напряжение	206	3,745	0	0	0	0	0	0
кружок	178	3,744	0	0	0	0	0	0
множество	252	3,740	0	0	0	0	0	0

Как было сказано выше, в нашем исследовании мы вычисляем **условную** полноту. Мы не можем вычислить истинную полноту, но можно утверждать, что на достаточно большом корпусе должно встречаться большинство коллокаций из золотого стандарта.

Условная полнота была вычислена следующим образом (на примере слова *рука)*: коллокация считается хорошей, если она встретилась более в чем 2 словарях и выдана более чем 1 мерой. см. таблицу 6.

Таблица 6. Пример оценки коллокатов к слову *рука*.

	Количество словарей	Количество мер	
как рукой сняло	2	2	хорошая
мужчины	2	1	хорошая
рукой подать	2	2	хорошая
умывать руки	2	1	хорошая
Всё валится из рук	*3*	*2*	хорошая
кривые	3	2	хорошая
матери	3	2	хорошая
нечистые	3	2	хорошая
под	3	2	хорошая
под рукой	3	1	хорошая
подать руку	3	2	хорошая
просить руки	3	1	хорошая
целовать	3	2	хорошая
вытянуть	1	3	
девушки	1	3	
кисть	1	4	
кожа	1	4	
обе	1	3	
прибрать к	1	3	

ребенка	1	3	
с пустыми	1	3	
трогать	1	3	
чешутся	1	4	
в четыре руки	*2*	*3*	хорошая
не доходят	2	3	хорошая
Рука на пульсе	2	3	хорошая

Дано: всего b коллокаций, a – количество «хороших» коллокатов,c - количество коллокатов в золотом стандарте. Для нашего слова *рука* это a=72, b= 246,c= 153. Полнота = $\frac{a}{b}$.

Таким образом, мы вычислили точность, условную полноту и F-меру (см. главу 3) для наших слов (таблица 7).

Таблица 7. Полнота, точность, F-мера для мер ассоциации

Слово	Полнота	Точность	F-мера
Вода	0,2	0,3	0,25
Сердце	0,4	0,3	0,35
Рука	0,3	0,32	0,31
Белый	0,6	0,3	0,45
Скакать	0,3	0,1	0,2
Семь	0,4	0,2	0,3
Свой	0,12	0,2	0,12

Также мы разделили все получившиеся словосочетания по группам: группа 1 - кандидаты, не встретившиеся в словарях, и выданы 1-2 мерами; группа 2 - кандидаты, не встретившиеся в словарях, и выданы 3-4 мерами; группа 3 - кандидаты, встретившиеся в 1-3 словарях, и выданы 1-2 мерами; группа 4 - кандидаты, встретившиеся в 1-3 словарях, и выданы 3-4 мерами; группа 5 - кандидаты, встретившиеся в 4-12 словарях, и выданы 1-2 мерами; группа 6 - кандидаты, встретившиеся в 4-12 словарях, и выданы 3-4мерами; группа 7 - кандидаты, встретившиеся в 1-3 словарях, но не выданы ни одной

мерой; группа 8 - кандидаты, встретившиеся в 4-12словарях,но не выданы ни одной мерой. Приводим ниже таблицы 8-9 для наглядности.

Таблица 8. Распределение количества коллокаций для слов *рука*, *вода*, *сердце*

рука				вода				сердце		
словари	меры			словари	меры			словари	меры	
0	1--2	44		0	1--2	25		0	1--2	53
0	3--4	30		0	3--4	45		0	3--4	31
1--3	1--2	21		1--3	1--2	10		1--3	1--2	26
1--3	3--4	21		1--3	3--4	49		1--3	3--4	33
4--12	1--2	8		4--12	1--2	1		4--12	1--2	14
4--12	3--4	27		4--12	3--4	15		4--12	3--4	22
1--3	0	134		1--3	0	229		1--3	0	135
4--12	0	31		4--12	0	25		4--12	0	14

Таблица 9. Распределение количества коллокаций для слов *семь*, *свой*, *белый*, *скакать*

семь				свой		
словари	меры			словари	меры	
0	1--2	8		0	1--2	1
0	3--4	7		0	3--4	24
1--3	1--2	2		1--3	1--2	4
1--3	3--4	1		1--3	3--4	1
4--12	1--2	0		4--12	1--2	0
4--12	3--4	0		4--12	3--4	1
1--3	0	4		1--3	0	40
4--12	0	0		4--12	0	4

белый				скакать		
словари	меры			словари	меры	
0	1--2	12		0	1--2	1
0	3--4	26		0	3--4	40
1--3	1--2	7		1--3	1--2	2

1--3	3--4	11		1--3	3--4	5
4--12	1--2	8		4--12	1--2	0
4--12	3--4	10		4--12	3--4	1
1--3	0	13		1--3	0	12
4--12	0	2		4--12	0	0

Такое разное количество коллокатов объясняется тем, что последние 4 слова (*белый*, *скакать*, *семь*, *свой*) являются низкочастотными, и взяты намеренно с целью проанализировать работу статистических мер на разных типах слов.

Далее мы воспользовались услугами экспертов. Мы опросили трех экспертов с филологическим образованием, а также двоих с его отсутствием. Кандидаты к коллокации были рассортированы по группам в зависимости от количества случаев встречаемости в словарях и мер, которые выявили эти коллокации. Экспертам было необходимо оценить каждый пример по шкале от 0 до 2, где 0 - не коллокация, 1- слабая коллокация/затрудняюсь ответить, 2 - абсолютно точно коллокация.

Также экспертам был представлен текст, поясняющий, что есть коллокация. Он звучит так: "Коллокации - это словосочетания, в которых главный по смыслу компонент (**база**) употреблен в своем прямом значении, а вспомогательный компонент (**коллокатор**) сочетается в рамках смыслового класса, но выбор конкретного слова предопределен общепринятым употреблением. Например: *проливной* [коллокатор] *дождь* [база], *принимать* [коллокатор] *решение* [база], *зерно* [коллокатор] *истины* [база], *ставить под* [коллокатор] *сомнение* [база], *топорная* [коллокатор] *работа* [база], *трескучий* [коллокатор] *мороз* [база]"[Баранов и Добровольский 2014:73].

В таблице 10 представлен пример экспертной оценки коллокатов для слова *вода*. С результатами экспертной оценки всех слов можно ознакомиться в приложениях 8-14.

Как мы видим, в большинстве случаев ответы экспертов распределились в соответствии с сортировкой кандидатов в коллокации в зависимости от выдачи мер и количества случаев встречаемости в словарях. Также мы посчитали для

каждой коллокации среднее арифметическое ответов экспертов и стандартное отклонение. Например, сочетание «выпрямить руки» было в группе №1, то есть оно встретилось 0 раз в словарях и 1-2 раза оно было выдано мерами ассоциации. Эксперты посчитали его «не коллокацией» и поставили 0. Это подтверждает гипотезу о том, что данное словосочетание коллокацией не является. Аналогично, выражение *газированная вода*, которое встретилось 8 раз в словарях и было выдано всеми мерами, набрало максимальное количество баллов, то есть все эксперты единодушно поставили 2. На основе этого мы можем сделать вывод, словосочетание является «качественной» коллокацией. Однако, не всегда мнение экспертов совпадает со словарями - в эксперименте встретились сочетания, входящие в группы с редкой встречаемостью, (например, в первую (0|1-2) , вторую(0|3-4), седьмую(1-3|0) и восьмую (4-12|0) группы), которые также получили максимальное количество оценок «2» - *в самое сердце*, *водой не разольешь*, *под горячую руку*, *прижать к сердцу, наложить руки*. Основываясь на эксперименте, можно сделать вывод, что не стоит полагаться на наш золотой стандарт, так как многие коллокации, не встретившиеся в словарях, признаны экспертами «хорошими». Для примера возьмем коллокации со словом *сердце*. В первой группе приблизительно 20% коллокаций, получивших оценку 2, во второй 15%, в третьей 20%, в четвертой 25%, в пятой 38%, шестой 45%, седьмой 15% и восьмой 40%. Первая и вторая группы содержат словосочетания, мало встретившиеся в словарях, но при этом в них достаточно много (20 и 15%) коллокаций, отмеченных экспертами оценкой 2. Также мы используем получившееся среднее арифметическое оценок для каждого слова - предположим, что среднее арифметическое выше 1.6 - это хороший результат. Посмотрим, сколько коллокаций являются хорошими с этой точки зрения. Результаты получились такие: *сердце* - 81 из 327 (24.77 %), *вода* - 44 из 398(11.06 %), *рука* - 50 из 314(15.92 %), *белый* - 49 из 91(53.85 %), *скакать* - 18 из 64(28.13 %), *семь* - 11 из 25(44 %), *свой* - 42 из 76(55.26 %).

Еще один способ сравнить меры ассоциации – это оценить их качество по отношению друг к другу. Можно сделать это с помощью коэффициента корреляции Спирмена. Для этого мы проранжировали слова и их значения, выданные мерами. Каждое слово имеет разный ранг для разных мер. Например, коллокация махнуть рукой - в T-score она на 101 месте, в MI3 на 19, log-likelihood на 39 и так далее. См. таблицу 11. Далее, опираясь на эти ранги, мы посчитали Коэффициент корреляции Спирмена, где диапазонами считаются ранги двух мер. Каждая из семи мер сравнивается с остальными шестью, а также с рангом значения совместной встречаемости (Cooccurrence count rank, второй столбец таблицы 12).

Таблица 10. Оценка экспертов коллокатов для слова *вода*.

	Количество коллокаций в словарях	Количество коллокаций, выданных мерами ассоциации	Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	Среднее арифметическое	Стандартное отклонение
туалетная	3	4	2	2	2	2	2	2	0,595119
родниковая	4	1	2	2	2	2	2	2	0,640095
газированная	8	4	2	2	2	2	2	2	1,699673
горячая	4	4	1	2	1	2	0	1,2	1,247219
жесткая	4	3	1	1	1	1	2	1,2	0,953794
живая	5	4	2	2	2	2	2	2	0,953794
минеральная	7	4	1	1	1	1	2	1,2	1,753964
морская	7	4	1	1	1	1	2	1,2	1,753964
питьевая	5	4	1	2	1	2	2	1,6	1,187317
прозрачная	4	4	0	0	1	0	0	0,2	1,462494
святая	4	4	2	2	2	2	2	2	0,745356

стакан	7	4	2	2	2	2	0	1,6	1,738054
теплая	4	4	1	1	1	1	0	0,8	1,247219
холодная	4	4	1	1	1	1	0	0,8	1,247219
чистая	5	4	0	1	0	0	2	0,6	1,63936
чистейшей	4	4	2	2	2	2	2	2	0,745356

Таблица 11. Сводная таблица коллокаций для слова *рука* с рангами

	T-score		MI3		log-likelihood		log-Dice		MI		min, sensitivity		MI,log-_f	
держать	14	114,306	5	36,176	7	136804,5	1	9,734	13	8,816	2	0,031	1	83,594
махнуть	101	42,662	19	32,531	39	25427,65	33	7,134	2	10,869	53	0,004	2	81,604
протянуть	67	53,911	15	33,042	28	35907,6	16	7,793	3	10,027	19	0,007	3	79,984
левый	30	90,489	9	34,474	11	81367	3	9,109	17	8,459	7	0,02	4	76,27
палец	32	89,482	10	34,312	12	78424,69	4	9,068	20	8,36	8	0,019	5	75,197
покладать	199	26,934	37	30,398	86	11216,63	182	5,814	1	11,39	152	0,002	6	75,047
нога	12	114,883	8	35,254	8	119937,3	2	9,54	28	7,852	1	0,032	7	74,572
кисть	56	61,991	18	32,56	25	39668,71	10	8,142	15	8,731	15	0,009	8	72,108
умелый	105	42,17	26	31,179	48	20651,91	36	7,09	8	9,579	54	0,004	9	71,715
пожать	139	34,125	40	30,242	72	14063,21	75	6,489	5	9,865	93	0,003	10	69,675

вытянутый	130	36,845	35	30,455	65	15883,54	55	6,706	7	9,634	92	0,003	11	69,525

Таблица 12. Коэффициент корреляции Спирмена между различными мерами для слова *рука*.

Коэффициент корреляции Спирмена		T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,5680	X	-0,4390	0,708277	0,820777	0,648558	0,567985	-0,15824
MI	0,3976	-0,4390	X	0,214894	0,080569	0,189818	0,39761	0,942273
MI3	0,7339	0,7083	0,214894	X	0,97637	0,631589	0,734003	0,497478
log-likelihood	0,7777	0,8208	0,080569	0,97637	X	0,707565	0,777751	0,379206
min. sensitivity	0,9085	0,6486	0,189818	0,631589	0,707565	X	0,908553	0,369471
log-Dice	1,0000	0,5680	0,39761	0,734003	0,777751	0,908553	X	0,597898
MI.log-_f	0,5978	-0,1582	0,942273	0,497478	0,379206	0,369471	0,597898	X

Корреляцию для остальных слов можно посмотреть в приложениях 15-21.

Чтобы оценить эффективность каждой меры, мы использовали метод Харина-Ашманова [Ashmanov et al. 1997], который оценивает релевантность возвращенной информации. На основе экспертной оценки выделенных коллокатов и их места в ранжированном списке в отношении каждой меры был сформирован набор характеристик. Набор характеристик означает количество истинных коллокаций, полученных с различным количеством коллокатов из ранжированного списка точности). Согласно [Ashmanov et al. 1997], мы выбираем характеристические множества, которые содержат 5 элементов - значения точности для первых 20, 50, 100, 150 и 200 коллокаций в верхней части списка. В таблице 12 показано распределение количества истинных коллокаций в разных мерах:

Таблица 12. Распределение количества истинных коллокаций в разных мерах

	t-score	MI	MI3	Log-likelihood	Min. sensitivity	Log--Dice	MI.log-_f
1-20	1	8	3	2	5	4	7
1-50	5	15	7	6	10	10	15
1-100	10	22	16	16	19	14	21
1-150	14	25	17	19	24	17	24
1-200	22	26	21	24	25	22	25

Вес присваивается каждому элементу набора признаков (1, 2, 3, 4 и 5 соответственно). Каждый элемент «взвешивается»: каждое из пяти значений точности умножается на его вес и делится на 15 (сумма весов). Сумма взвешенных элементов - это результирующая точность характеристического множества. Приведем пример для MI3.Количествоистинных коллокаций в мере MI3 в первых20 примерах- 3 (точность 0,15), в первых50 - 7 (точность 0.14), в первых 100- 16 (точность 0.16),в первых 150 - 17 (точность 0,113),в первых 200

- 21 (точность 0.105). Средняя точность получается 0.15*1/15+0.14*2/15+0.16*3/15+0.113*4/15+0.105*5/15=0,01+0,019+0,032+0,03+0,035=0,126. В таблице 13 можно ознакомиться с результатами для остальных мер ассоциации.

Таблица 13. Значения точности для разных мер.

	t-score	MI	MI3	Log-likelihood	Min.sensitivity	Log--Dice	MI.log-_f
Количество настоящих коллокаций	22	26	21	24	25	22	25
Точность	0,099	**0,199**	0,080	0,129	**0,166**	0,135	**0,190**
Место	6	**1**	7	5	**3**	4	**2**

Итак, лидирует мера MI. На втором месте MI.log-_f, за ней следует Min.sensitivity. Интересно заметить, что лучше всего оказались меры семейства MI. Однако, родственная им мера MI3 оказалась на 7 месте, последнем. Возможно, это связано с тем, что в формуле использовано возведение функции в куб.

4.2. Оценка результатов

Актуальность данной работы состоит в том, чтобы сравнить результаты проведенных исследований с похожими, уже имевшими место ранее в других исследованиях. До проведения эксперимента мы считали, что словарные статьи (золотой стандарт) производили впечатление достаточно полных, для того, чтобы черпать из них коллокации. Как оказалось, это не так. Словари неполны, а словосочетания, приведенные в них, являются, скорее идиомами, чем устойчивыми словосочетаниями в широком смысле. Также мы считали возможным, что мера ассоциации Log-Dice окажется лучшей и самой эффективной среди остальных мер, но после наших расчетов стало видно, что первое место завоевала мера MI, второй после нее оказалась MI.log-_f, а предполагаемый лидер занимает всего лишь 4 место из 7. После объявления

результатов сравнения реальности и ожиданий можно перейти к соотнесению похожих работ на данную работу.

Похожее исследование было проведено Захаровым В.П. [Zakharov 2017:9], в этой работе также сравнивались меры ассоциации с помощью метода [Ashmanov et al. 1997], в результате оказалось, что лучшая мера MI.l-og_f. Кроме того, в исследовании сказано, что точность меры log--likelihood ниже, чем точность меры min. sensitivity. Меру же MI тоже нужно отметить, так как она оказалась эффективной и отличается от остальных. В нашей работе мы выяснили, что лучшей мерой(по точности) оказалась MI. Точность меры log--likelihood равняется 0,129, тогда как точность min. sensitivity - 0,166.Поэтому мы согласны с утверждением, что точность первой меры меньше. Также в приведенном исследовании сказано, что мера MI незаменима при извлечении редких терминологических словосочетаний.

Еще одна похожая работа, автором которой является Хохлова М.В. [Хохлова 2008:353-355], гласит, что в результате эксперимента выяснилось, что меры извлекают словосочетания, не зафиксированные в словаре. Это соотносится с нашим выводом о том, что словарные статьи неполны и устарели. Также в этой статье замечено, что мера T-score выделила большее число биграмм, в которых компонентом являются знаки препинания. Мы можем сказать, что частично это подтвердилось и в нашем эксперименте, данная мера действительно выделяет много знаков препинания.

Извлечением коллокаций помимо вышеназванных работ занимались Кормачева Д., Пивоварова Л., Копотев М. [Kormacheva, Pivovarova, Kopotev 2014:4].

Все авторы сходятся в едином мнении, что результаты ручного аннотирования совпадают с результатами оценки, полученными с помощью золотого стандарта. В этом мы можем согласиться с авторами. Также авторы статьи [Kormacheva, Pivovarova, Kopotev 2014:4] делают вывод, что мера t-score справляется с извлечением устойчивых словосочетаний лучше, чем Dice и MI,

хотя в целом количество коллокаций, полученных с использованием этих мер, высокое. Это утверждение спорное и требует проверки.

Выводы по главе 4

1) В результате эксперимента мы имеем возможность сравнить выбранные меры ассоциации с золотым стандартом и с оценками экспертов, оценить их качество и эффективность с помощью значений точности.

   Важно отметить, что эксперимент был проведен на большом репрезентативном корпусе, для всех слов были учтены и оценены все выданные коллокации. Поэтому полученные результаты можно считать достоверными.
2) Была проделана работа по выявлению корреляции между данными автоматического извлечения коллокаций и наполнением нашего золотого стандарта. При этом учитывалось, на основе какого количестве мер была извлечена коллокация и в каком количестве словарей она присутствует.
3) В отношении золотого стандарта можно сказать, что словари неполны. Наш золотой стандарт показал себя слабо: многие словосочетания, выданные мерами и которые можно причислить к коллокациям (устойчивым или идиоматичным), отсутствуют в словарях.
4) На основе работы, указанной в п. 2, можно сделать еще один вывод, что в словарях в подавляющем большинстве содержатся фразеологизмы, а эксперты оценивают просто устойчивые словосочетания.
5) Сравнение мер лексической ассоциации дало следующие результаты: наибольшую эффективность показала мера MI, далее следуют MI.log-_f и Min.sensitivity.
6) На наш взгляд, эффективность меры MI может быть объяснена тем, что в процедуре извлечения коллокаций мы задавали ограничение по частоте коллокаций, причем довольно высокое.
7) В отношении экспертной оценки можно сказать, что не стоит полагаться на наш золотой стандарт, так как многие коллокации, не встретившиеся в словарях, признаны экспертами «хорошими", причем единогласно. Судя

по оценке, можно сказать, что какие-то коллокации "сильнее" других, так как эксперты поставили им больше баллов. Например, *газированная вода* сильнее, чем *питьевая вода*, потому что первая коллокация получила от экспертов самые высокие баллы в отличие от второй.

8) Вычислив полноту и точность мер, мы можем заявить, что результаты получились примерно одинаковыми и особых расхождений не наблюдается, поэтому мы делаем данные метрики основными в нашей работе.

ЗАКЛЮЧЕНИЕ

В данном исследовании были описаны меры лексической ассоциации и проведена оценка их эффективности с помощью золотого стандарта, оценки людей-экспертов и вычисления точности. В качестве золотого стандарта была собрана база устойчивых словосочетаний на основе различных толковых и фразеологических словарей.

Была проделана работа по выявлению корреляции между данными автоматического извлечения коллокаций и наполнением нашего золотого стандарта. При этом учитывалось, на основе какого количества мер была извлечена коллокация и в каком количестве словарей она присутствует. Результатом явилось то наблюдение, что информация об устойчивых словосочетаниях в словарных статьях неполная - многие коллокации, выданные мерами ассоциации, отсутствуют в словарях. Те же коллокации были высоко оценены экспертами, что свидетельствует об их "истинности". Представляется возможным включить такие устойчивые словосочетания в словари, таким образом обновляя их.

С помощью вычисления точности мер был проведен анализ эффективности мер ассоциации. Лучшей мерой оказалась мера MI, далее следуют MI.log_f и Min.sensitivity. Это значит, что в подобных исследованиях в первую очередь следует применять именно их.

В целом все меры ассоциации показали высокий уровень работоспособности в сравнении со словарями.

Практическая и научная значимость данной работы связана с возможностью сравнить ее с похожими исследованиями, сопоставить результаты. В последней главе это действие выполнено и наблюдение показывает, что встречаются похожие цифры и результаты.

Список литературы

1. Баранов А.Н., Добровольский Д.О. Основы фразеологии. - М.:Флинта, 2014, с.44-96.
2. Виноградов В.В. Русский язык. – М.: Наука, 1972, с.
3. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005,с 5.
4. Иорданская Л.Н., Мельчук И.А.Смысл и сочетаемость в словаре. - М.: Языки славянских культур, 2007, с.227-228.
5. Хохлова М.В. Экспериментальная проверка методов выделения коллокаций. Slavica Helsingiensia, Хельсинки, 2008, с.354-355.
6. Шанский Н .М. Лексикология современного русского языка. М., 1964, с. 201.
7. Шанский Н.М. Фразеология современного русского языка. – М.: Высшая школа,1985, с.157.
8. Ягунова Е.В., Пивоварова Л.М. Природа коллокаций в русском языке. Опыт автоматического извлечения и классификации на материале новостных текстов. - М.: Всероссийский институт научной и технической информации РАН, 2010, с.14-15.
9. Ярцева В.Н. Лингвистический энциклопедический словарь.— М: Советская энциклопедия, 1990. URL:http://tapemark.narod.ru/les/index.html.
10. Ashmanov I., Grigoryev S., Gusev V., Kharin N., Shabanov V. Using Statistical Method for Intelligent Computer-Based Text Processing/ The Proceedings of the Dialog-’97,1997, pp. 33-37.
11. Benko V. Aranea: Yet Another Family of (Comparable) Web Corpora // Petr Sojka, Aleš Horák, Ivan Kopeček and Karel Pala (Eds.): Text, Speech and Dialog-ue. 17th International Conference, TSD 2014, Brno: Springer International Publishing Switzerland, 2014, pp.248-253.
12. Čermák F. Statistické metody hledání frazémů a idiomů v korpusech // Kolokace, Praha, 2006, pp.223-248.

13. Daille B. Mixed approach for the automatic extraction of terminology: lexical statistics and linguistic filters[Approche mixte pour l'extraction automatique de terminologie: statistiques lexicales et filtres linguistiques],PhD thesis, Université Paris, 1994 pp.70-72.
14. Dunning T.E. Accurate methods for the statistics of surprise and coincidence. Computational Linguistics, 19(1), 1993, pp 61-74..
15. Evert S. The statistics of word cooccurrences: Word pairs and collocations. PhD thesis, University of Stuttgart, 2004, p.35.
16. Evert S., Kermes H. Experiments on candidate data for collocation extraction. In Proceedings of the 10th Conference of The European Chapter of the Association for Computational Linguistics (EACL), 2003, p.17.
17. Fano R.M. Transmission of information; a statistical theory of communications. MIT Press, New York, 1961, pp.5-62.
18. Khokhlova M, Zakharov V, EFFICIENCY OF THE SKETCH GRAMMAR FOR RUSSIAN, St.Petersburg, 2007, pp.4-6.
19. Kilgarriff A., Rychly P., Smrz P., Tugwell D., The NoSketch Engine, Proceedings of EURALEX-2004, 2004, pp.105-116.
20. Kormacheva D., Pivovarova L. & Kopotev M.' Automatic Collocation Extraction and Classification of Automatically Obtained Bigrams' in Proceedings: Workshop on Computational, Cognitive, and Linguistic Approaches to the Analysis of Complex Words and Collocations, 2014, pp.3-4.
21. Křen M. Collocation Measures and the Czech Language: Comparison on the Czech National Corpus data, Praha, 2006, pp.246-247.
22. Mel'ˇcuk I. Collocations and lexical functions. // Cowie AP (ed) Phraseolog-y. Theory, Analysis, and Applications, Claredon Press, Oxford, 1998, pp.23-53.
23. Pecina P. Lexical association measures and collocation extraction. Language Resources and Evaluation 1(44), 2010, pp.27-28, 48.
24. Ramisch C. A generic and open framework for multiword expressions treatment: from acquisition to applications. Computation and Language. Universidade Federal do Rio Grande do Sul, 2012, p.46.

25. Rychlý, P. Manatee/Bonito – A Modular Corpus Manager // 1st Workshop on Recent Advances in Slavonic Natural Language *Processing*. Brno: Masaryk University, 2007, pp. 65-70.
26. Sag I., Baldwin T., Bond F., Copestake A., Flickinger D. Multiword Expressions: A Pain in the Neck for NLP? International Conference on Computational Linguistics and Intelligent Text Processing, Mexico City, Mexico, Springer, 2002, pp.3-7.
27. Seretan V. Syntax-based Collocation extraction. Text, Speech and Language. – Springer Science, 2011, p.43.
28. Sinclair J. Corpus, Concordance, Collocation. Oxford University Press, Oxford, 1991, pp.123-140.
29. Zakharov V.P. AUTOMATIC COLLOCATION EXTRACTION: ASSOCIATION MEASURES EVALUATION AND INTEGRATION// Dialog-2017, 2017 (in print), pp.1-10.

Источники золотого стандарта

1. Ассоциативная база данных УрРАС. URL: *iling-ran.ru/main/publications/urras.*
2. Большой академический словарь русского языка: РАН, Ин-т лингвистич. исследований; Под ред. Л. Кругликовой, А. Шушкова. - М.: Наука, 2004.
3. Денисов П.Н., Морковкин В.В. Словарь сочетаемости слов русского языка - М.:Рус.яз, 1983.
4. Русский ассоциативный словарь.URL:http://www.tesaurus.ru/dict/dict.php.
5. СИБАС (Сибирский ассоциативный словарь русского языка). URL: http://adictru.nsu.ru/.
6. Словарь русского языка: В 4-х т. / РАН, Ин-т лингвистич. исследований; Под ред. *А. П. Евгеньевой*. — 4-е изд., стер. — М.: Рус. яз.; Полиграфресурсы, 1999.
7. Словарь-тезаурус ЕВРАС. URL: *iling-ran.ru/main/publications/evras.*

8. Современный толковый словарь русского языка/ Т. Ф. Ефремова - М.: АСТ, 2006.
9. *Толковый словарь* русского языка / Под ред. Д.Н. *Ушакова*. — М.: Гос. ин-т "Сов. энцикл."; ОГИЗ; Гос. изд-во иностр. и нац. слов., 1935-1940.
10. Толковый словарь русского языка/ Под ред. С.И. Ожегова. - М.: Оникс, 2010.
11. Молотков А.И.Фразеологический словарь русского языка - М.:Советская энциклопедия, 1968.
12. Фелицина В.П, Мокиенко В.М. Русские фразеологизмы: Лингво-страноведческий словарь - М.: Рус.яз, 1990.

Приложение 1. Коллокаты для слова *сердце*

													log-dice		MI3		T-score		Log-likelihood	
	словарь сочетаемости	МАС	Сиба с	БТС	Русский ассоциативный словарь	УрРАС	ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремовой	Словарь Молотка	Словарь Мокиенко	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0
ангела					+															
аритмия														+						
бездонное						+														
безумное			+																	
бешеное			+																	
биение	+		+											+		+		+		+
благородство	+						+													
близко к		+		+		+		+	+		+			+						+
Бога													+		+		+			
болезнь														+		+		+		+
болеть/любить всем сердцем за кого-л		+		+		+			+		+									
болит	+	+	+		+	+	+	+					+		+					
боль в	+		+			+	+							+		+		+		+
больного	+																			
больное	+		+	+	+	+										+				
большое	+	+	+	+	+	+					+			+				+		

боязнь	+																			
Брать за́ сердце		+		+				+	+		+									
Бычье			+	+	+											+				+
бьется	+		+	+	+	+	+	+	+				+		+	+	+		+	
бьющееся			+	+										+		+				+
в глубине сердца		+		+							+			+		+				
в пятки ушло			+		+															
в самое														+						
в сердцах		+		+				+	+	+	+									
веление	+			+										+		+		+		+
великодушное					+															
Верное	+			+	+		+									+				
взыграло		+		+																
волновать	+																			
воспоминания																				+
вредно для	+																			
всем сердцем		+		+							+									
выбор	+																			
вылетело			+																	
Вынуть		+																		
выпрыгнет															+					
вырвать из				+					+		+									
выстрелить в	+			+		+														
героя	+				+															
гибнет					+															
говорит	+																			
говяжье					+											+				+
голос						+										+				
горит		+		+	+						+		+							
города	+		+	+				+					+		+		+			

горячее	+		+		+	+	+							+		+		+		+
давление														+		+				
дама		+		+				+						+		+		+		
дамы			+																	
дать волю сердцу											+									
девичье																				
девушки			+										+		+		+			
действует на	+																			
Держать сердце на кого-либо		+		+					+		+									
держаться за	+																			
детское																				+
деятельность	+																	+		
до глубины		+																		
доброе	+		+	+	+	+	+	+	+		+			+		+		+		+
дочери	+																			
дрогнуло	+			+	+		+						+		+				+	
друг сердца				+																
друга					+															
Европы													+		+		+		+	
екнуло			+	+											+				+	
жгучее					+															
женское														+		+		+		
женщины													+		+		+		+	
жены	+																			
жестокое					+															
живет													+							
живое			+	+	+															
жить в сердце				+																
жить сердцем				+																

забилось													+		+					
заболевание														+		+		+		+
завоевать														+		+		+		+
зажечь	+																			
зажигать				+																
закрыто			+																	
замерло															+				+	
замирает	+			+	+	+	+						+		+		+			
заныло															+					
занято						+	+													
запасть в				+																
затаить в				+																
заячье	+			+																
здоровое	+				+	+	+							+		+		+		
Земли			+																	
злое														+						
зов	+			+										+		+				
золотое	+		+	+	+	+	+	+	+							+				
из камня			+																	
измученное																+				
изношенное	+			+																
изучать	+																			
искусственное	+			+										+		+		+		+
испортить	+																			
исследование																		+		
Кавказа			+																	
Как маслом по сердцу		+									+									
Как ножом по сердцу		+		+							+									
каменное	+		+		+	+	+	+								+				

камень с сердца											+									
клапан	+		+		+				+					+		+		+		
клетки														+				+		
ключик к																+				
колотится	+			+									+		+		+			
кольнуло															+					
красавицы			+		+								+		+		+		+	
красное	+		+			+	+													
кровообращение														+		+			+	
кровью обливается		+		+			+		+		+			+		+		+		+
ледяное																+				
лечить	+																			
льва	+																			
Львиное	+		+	+			+							+		+		+		+
любимого													+				+			
любимое			+		+		+													
любящее	+		+	+	+	+	+							+		+		+		+
людей													+							
людское			+																	
маленькое														+						+
мамы													+							
массаж	+													+		+		+		
массировать	+																			
матери	+		+		+		+						+						+	
материнское														+		+		+		
миллионов													+							
мое			+		+		+							+		+				+
молодое	+				+															
молчит			+																	
мощное			+																	

мудрого человека			+																
мужское																+			
мужчины													+		+		+		+	
музыка														+						
мышцы	+													+		+				
На с. накипело, наболело				+																
На с. тяжело, легко, тоскливо				+																
На сердце кошки скребут		+		+							+									
нагрузка на														+		+		+		+
надрывать				+				+			+									
нарушения														+				+		+
не выдержало					+									+		+		+		+
не камень													+							
не лежит		+		+				+	+		+		+				+		+	
не на месте		+		+	+			+			+									
не обманет					+															
не прикажешь													+		+					
не хватит				+																
недостаточность																+				+
нежное	+			+	+		+									+				+
нет сердца				+					+											
Ничто не шевельнулось в моём с.				+																
новое			+																	
ноет		+			+	+		+												
область	+													+		+		+		+
Обнаженные																+				+

оборвалось		+		+																
обследование	+													+						
огромное			+	+			+											+		
огрубело															+					
одинокие			+			+								+		+		+		
ожесточилось															+					
ожирение	+																			
операция	+			+	+									+		+		+		+
оперировать	+																			
останавливается													+							
остановилось	+		+		+		+								+					
остановка														+		+		+		
от всего		+		+				+	+	+				+						
от глубины		+									+			+						
от полноты		+									+									
от сердца отлегло													+			+				
от чистого														+		+		+		+
отважное	+																			
отдавать	+				+				+		+									
отдано															+					
отзывчивое									+							+				
отклик														+		+		+		+
открытое			+	+	+			+												
открыть	+			+							+			+		+		+		+
отлегло от		+						+	+		+									
отошло		+		+							+									
оттаяло															+					
отца													+							
падает				+				+			+									
память	+			+												+	+			

Пармы													+		+		+		+	
патологии														+						
перебои														+		+		+		+
перевернулось		+																		
пересадка	+			+	+											+				
пересаживать	+																			
Петербурга													+				+		+	
пламенное	+				+		+													
плода													+		+		+		+	
по сердцу		+						+	+	+	+									
победить	+																			
подсказывает	+			+									+		+		+		+	
поет					+										+					
поклонников													+		+		+		+	
покорить	+	+		+				+			+			+		+		+		+
полезно для	+																			
положа руку на сердце	+	+		+				+	+		+							+		+
попасть в	+			+																
пополам			+																	
поражения														+						
порок	+			+				+	+					+		+		+		+
потрясти	+			+																
преданное	+																			
предложение руки и		+		+				+						+		+		+		+
прижать к	+			+				+												
принадлежит	+												+				+			
принимает	+																			
прихватило															+					
проблемы														+		+		+		
проверять	+																			

продолговатое			+	+																
пронзить	+			+					+							+				
просит				+																
простое	+																			
пылкое					+															
работа	+													+		+		+		+
работает	+				+		+						+				+			
радость														+						
радуется	+				+				+						+					
разбито			+		+	+														
разбитое			+		+	+	+							+		+		+		
разбить	+	+	+	+																
разрыв	+			+																
разрывается				+	+						+		+		+		+			
раненое																+				
ранимое					+															
растопить														+		+				+
растравить	+																			
растревожить	+			+																
рвется			+				+													
ребенка	+												+		+		+			
режет			+																	
риск														+						
ритм	+													+		+		+		+
родное					+															
России																	+			
руку и сердце														+		+		+		
русское														+						
с замиранием		+		+							+			+		+		+		+
с легким		+									+					+				
с открытым		+									+			+		+		+		

с тяжелым		+									+					+				
с чистым				+					+		+			+		+				+
С. сердцу весть подаёт				+																
светлое			+		+															
свинец на		+																		
свиное																				
сдает				+																
сердце моё		+		+				+	+				+							
сжалось													+		+				+	
сжимается	+		+	+	+								+		+		+		+	
сильное			+		+															
скрепя		+		+				+			+			+		+		+		+
слабое			+		+		+							+				+		+
смелое	+																			
собачье			+		+		+							+		+		+		+
согревает													+		+				+	
согреть																+				
сокращения														+						+
Солдатское																+				
Сорвать сердце на ком-чем		+		+				+	+		+									
состояние														+				+		
сосуды														+		+		+		+
спокойное					+															
спортсмена	+																			
Стальное														+		+				+
старика	+																			
старое	+																			
столицы	+												+		+		+		+	
страдает	+																+		+	
страны					+															

стук	+		+			+	+							+		+		+		
стучит	+		+	+	+	+	+						+		+		+		+	
схватиться за	+			+				+												
теплое														+						
теплое					+		+													
ткани														+						
Тоны	+															+				
трансплантация														+		+				+
тревожное					+															
трепетное					+															
трепещет															+		+		+	
трогать сердце	+			+												+				
труса	+																			
трусливое	+			+																
удар																+				
Ужалить в самое с				+																
УЗИ														+						+
Уколоть в самое с				+																
укреплять														+		+		+		+
ум				+										+		+				+
успокоилось															+					
функции																		+		
Холодное			+		+		+							+		+		+		+
хорошее	+				+															
Храброе			+				+									+				+
хрупкое						+	+													
человека	+				+	+	+						+		+		+		+	
человеческое	+																			
черствое	+			+					+						+					

честное	+																			
четырехкамерное																+				
чистое	+			+				+			+			+		+		+		+
Читать в		+									+									
чувствительное					+				+											
чувствовать	+			+		+														
чувствует	+		+										+							
чует					+										+					
чуткое																				
шалит	+			+																
широта	+																			
шумы	+													+		+				+
щемит		+		+				+												
щемит		+		+				+					+		+				+	
ЭКГ													+	+						
юноши	+										+									

Приложение 2. Коллокаты для слова *вода*

														log-Dice		MI3		T-score		Log-likelihood	
	словарь сочетаемости	МАС	СИБАС	БТС	Русский ассоциативный словарь	УрРАС	ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремовой	Словарь Молоткова	Словарь Мокиенко	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0	от 0 до 3
анализ															+		+		+		+
артезианская															+		+		+		+
Байкала	+																				
бак															+		+		+		
бассейн			+												+		+		+		+
бежит	+		+		+	+															
бесцветная						+															
большая				+						+											
бочка	+																				
бочка для	+																				
бочка с	+																				
бросить в	+																				
броситься в	+																				
брызги	+																				
бурлит	+																				
Буря в стакане воды		+		+																	

бутилированная															+		+		+		+
бутылка из-под	+																			
бутылка с	+																			
бутылка	+		+												+		+		+		+
бытовая															+					+	
В мутной воде рыбу ловить		+		+					+			+	+								
ванна															+		+		+		+
ведро	+														+		+		+		+
вешние воды									+	+											
вещества															+		+		+		+
Вилами на воде писано		+		+								+									
вкус	+														+		+		+		+
вкусная	+		+			+															
внутренние				+					+	+											
Водой не разлить (не разольешь)*кого*		+		+					+	+											
водопроводная	+									+					+		+		+		+
воды моря														+							
Воды не замутит		+		+					+	+											
воды отошли				+					+												
возить	+																				
Возить воду		+		+						+											
войти в	+																				
Волги	+			+						+											

вольная		+							+												
вскипятить	+																				
Вывести на чистую воду		+		+					+	+			+								
выйти из	+																				
Выйти сухим из воды		+		+					+	+											
выкачать	+																				
вылить	+																				
вылить	+																				
вынырнуть из	+																				
выпить	+					+									+		+		+		
выпустить	+																				
высокая				+					+	+											
вытекла	+																				
вышла из берегов				+					+												
газированная	+	+	+	+	+	+				+	+				+		+		+		+
глотать	+																				
глоток	+																				
голубая			+		+	+															
горькая	+					+															
горячая	+		+		+	+									+		+		+		+
графин	+																				
графин для	+																				
графин с	+																				
грунтовые															+		+		+		+
грязная	+		+		+	+															
давление	+														+		+		+		+

дать	+														+		+		+		+
движение	+														+		+		+		+
дезинфицировать	+																				
дистиллированная	+				+	+									+		+		+		+
для мытья	+																				
для питья	+																				
для поливки	+																				
для стирки	+																				
для технических нужд	+																				
для хозяйственных нужд	+																				
для чая	+																				
Днепра	+																				
добавлять	+														+		+		+		+
добавляют														+		+		+		+	
дождевая	+								+						+		+		+		+
дорогая			+																		
доставка	+														+		+		+		+
емкость															+		+		+		+
жесткая	+		+			+			+						+		+		+		
жесткость															+		+		+		+
живая		+		+	+	+							+		+		+		+		+
жидкая			+																		
жить без	+																				
жить в	+																				
журчит						+															
загрязнения	+														+				+		+

загрязнять	+																			
закипит	+		+			+								+		+				
залить	+														+					
залить водой															+		+		+		+
замерзла	+																			
запас	+														+		+		+		+
запах	+																			
запить	+																			
затопила	+																			
зачерпнуть	+																			
здешняя	+																			
зеленая			+			+														
идти за	+			+						+										
ижевская									+											
из воды															+					
из источника	+													+						
из колодца	+			+																
из родника	+																			
из скважины														+						
избыток	+																			
из-под крана	+		+			+								+						
иметь в составе	+																			
искать	+																			
испарения	+														+		+		+		+
испарилась	+																			
использовать	+														+				+		+

используется														+				+		+	
Кавказские															+		+		+		+
Как (будто, словно) в воду опущенный		+		+					+												
Как в воду глядел		+		+						+											
Как в воду канул		+		+					+												
Как водой смыло		+		+																	
Как две капли воды		+		+					+												
Как рыба в воде		+		+					+			+									
Как с гуся вода		+		+					+			+	+								
камень точит			+			+															
канистра			+																		
капает	+																				
капля	+		+																		
капля															+		+				+
кастрюля															+		+		+		+
качать	+																				
качество															+		+		+		+
кипяченая	+			+		+			+						+		+		+		+
кипячение	+																				
кипящая															+		+				+
кислая	+																				
ключевая	+		+		+	+															
количество	+														+						
колодезная	+					+			+								+		+		

колодец															+					
комнатная																	+		+		+
концы в воду		+		+					+											
кончилась						+														
котел для	+																			
кружка	+		+																	
кувшин для	+																			
кулер с																	+		+		
купаться в			+																	
ледяная	+					+									+		+		+		+
летать над	+																			
лечебная	+			+							+									
лечиться на водах				+					+	+	+									
лимонная									+											
Лить воду		+								+	+									
Лить воду на мельницу		+		+					+	+										
лишить	+																			
ложку															+		+		+		+
льется	+					+														
любить	+																			
малая				+						+										
менять															+		+		+		+
мертвая		+		+																
местная	+																			
минеральная	+		+	+		+			+	+	+				+		+		+		+
Много воды утекло		+	+	+					+	+										
молекулы															+		+		+		

морская	+	+	+	+		+			+	+					+		+		+		+
московская	+																				
Мутить воду		+		+																	
мутная	+		+			+									+		+		+		
мыльная	+														+		+		+		+
мыться	+																				
мягкая	+								+						+		+		+		+
Набрать воды в рот		+		+						+											
наглотаться	+																				
нагрев	+														+				+		+
найти	+																				
наклонитьс я к	+																				
налить	+														+		+		+		+
наличие	+																				
наполнить	+																				
напор	+														+		+		+		+
направиться к	+																				
находиться в	+																				
невкусная						+															
негазированная			+														+		+		+
недостаток	+																				
нейтральные				+						+											
низкая									+												
носить	+																				
носить решетом воду		+		+						+											

обезжелезивание																	+				
обеззараживание															+						
обеспечить	+																				
обитать в	+																				
обливание водой																	+		+		+
облить	+																				
обнаружить в	+																				
обнаружить воду	+																				
обработка															+		+		+		+
обрызгать	+																				
обтираться	+																				
обходиться без	+																				
обыкновенная	+																				
обычная															+						
озера	+		+											+		+		+		+	
океана			+											+		+		+		+	
околоплодные															+		+		+		
окунуть	+																				
ополоснуть															+		+		+		+
опустить в	+			+																	
оставаться в	+																				
остаться без	+																				
остыла	+																				
отделиться от	+																				
отключить	+			+																	

оторваться от	+																			
отравить	+																			
отравленная						+														
отсутствие															+		+		+		+
охладить	+																			
охлаждение															+		+		+		+
очистить	+																			
очистка воды	+														+		+		+		+
очищенная	+														+		+		+		+
паводковые																	+				
перевозка	+																				
перекрыть	+																				
перелить	+																				
переправлять	+																				
пить воду	+														+		+		+		+
питьевая	+		+		+	+			+						+		+		+		+
плескаться в	+																				
плохая	+																				
плыть по/ под	+									+											
плыть против	+								+												
поверхность	+														+						
погружение															+		+		+		
погрузить в	+																				
под водой															+						
подавать	+																				
подача	+														+		+		+		+

подземные	+														+		+		+		+
подкисленная																	+		+		
подогрев	+																				
подогретая	+																				
подогреть	+																				
подойти к	+																				
подсоленная															+		+				+
поить	+																				
показаться из	+																				
полая				+																	
полива															+		+		+		
полная				+																	
положить в	+																				
послать за	+																				
поступает														+							
поток	+														+		+		+		
потребление	+														+		+		+		+
потребность в	+																				
появиться из	+																				
превратилась в пар	+																				
превращение в пар	+																				
предпочитать	+																				
пресная	+		+			+									+		+		+		
прибрежных															+		+		+		+
прибывает	+					+															

приводит в движение	+																			
привыкнуть к	+																			
принести	+																			
пробовать	+																			
прогреваетс я														+						
прозрачная	+		+		+	+									+		+		+		+
Пройти огонь и воду (и медные трубы)		+		+		+			+											
пролить	+																			
промыть водой	+														+		+				+
прописать	+																			
прополоскать	+																			
прополоскать в трёх водах				+						+										
пропускать	+																			
прорвала	+																			
просачивается																				
простая															+		+		+		+
протечка																	+				
проточная	+					+									+		+				+
прохладная						+									+		+		+		+
проходит														+							+
прыгнуть в	+																			
пустить	+																			
путешестви е по	+		+						+	+										

работа под	+																				
работать под	+			+																	
разбавить	+														+		+		+		+
развести водой	+														+		+		+		+
разлить	+																				
размыла	+																				
размыть водой	+																				
растворить в	+														+		+		+		+
расход															+		+		+		+
резервуар															+		+				
реки	+																				
рекомендовать	+																				
речная	+	+		+						+											
ржавая						+															
родниковая	+				+	+				+											
розовая				+						+							+		+		+
С лица не воду пить										+											
сброс																	+		+		+
свежая	+					+									+		+				
свойства															+		+		+		+
святая			+	+		+			+						+		+		+		+
Седьмая вода на киселе		+		+								+	+								
сельтерская									+												
сесть на	+																				
синяя			+																		

скользить по	+																			
скопление	+																			
сладкая	+		+																	
слив															+		+		+		+
сливать															+		+				
слить															+						+
слой	+																				
смешать с	+																				
смотреть на	+																				
смочить																	+				
смывается														+							
смыла	+																				
смыть водой	+																				
снабжать	+																				
снабжение	+																				
снести водой	+																				
содержание в чем-л	+																				
содержит	+													+		+				+	
содержится в	+																				
содовая	+																				
соленая	+		+			+									+		+		+		+
состав															+		+		+		+
состоит														+							+
сосуд с	+																				
спокойная						+															
спустить корабль на	+			+						+											
стакан	+	+	+	+	+				+	+					+		+		+		+

стакан с	+																			
стекает														+		+		+		+	
стоит														+				+			
столовая	+																				
сточные															+		+		+		+
стоячая	+					+															
струится						+															
струя	+														+		+		+		+
студеная		+				+															
счетчик															+		+		+		+
сырая	+		+	+		+															
тазик																	+		+		+
талая	+														+		+		+		+
Темна вода во облацех		+		+					+			+									
температура	+														+		+		+		+
теплая	+		+		+	+									+		+		+		+
термальной															+		+		+		+
территориальные воды	+	+		+						+											
течение	+														+		+		+		+
течет	+				+	+								+		+		+		+	
Тише воды, ниже травы		+		+									+								
Толочь воду (в ступе);		+		+					+	+											
толща															+						+
требуется															+		+		+		+
туалетная		+		+						+					+		+		+		+
Тяжелая вода		+		+		+				+											

удельный вес	+																				
умываться			+																		
унесла	+																				
употреблять															+		+		+		+
уровень	+																				
уровень															+				+		
уронить в	+																				
Утопить в ложке воды		+		+																	
фильтрация	+														+		+				
фильтры для															+				+		+
фруктовая	+	+		+						+											
химический состав	+																				
хлорированная	+				+	+											+		+		
холодная	+		+		+	+									+		+		+		+
Холодной водой окатить (или облить)		+		+																	
хорошая	+					+															
хранение	+																				
цвет	+																				
целебная	+		+																		
циркуляция															+						
цистерна для	+																				
Чающие движения воды		+																			
Черного моря	+																				

чистая	+	+	+		+	+									+		+		+		+
чистейшей		+		+					+	+					+		+		+		+
шумит	+					+															
энергия	+																				

Приложение 3. Коллокаты для слова *рука*

														log-dice		MI3		T-score		Log-likelihood	
	словарь сочетаемости	МАС	СИБАС	БТС	Русский ассоциативный словарь	УрРАС	ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремовой	Словарь Молоткова	Словарь Мокиенко	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0	от 0 до 3	от -3 до 0
ампутировать	+																			
ампутация	+																			
Ани	+																			
балерины	+																			
Без рук!		+		+										+						
белые	+		+		+		+													
божья					+															
больные	+				+															
большие	+		+		+	+	+													
болят		+	+		+		+													
брать														+						
брать инициативу в свои																				+
Бриллиантовая			+		+	+	+									+				+
В руках чьих или у кого		+				+			+											

в четыре руки				+					+					+				+		+
вдоль туловища														+		+		+		+
верная					+															
вести за	+			+		+			+											
вести под				+					+											
взмахнуть																+				
взявшись за																+		+		+
взять в	+	+	+	+			+							+		+		+		+
взять за	+		+	+			+	+												
взять на	+		+	+					+											
взяться за	+								+											
влажные	+																			
властная			+																	
власть в руках														+						
воздеть руки																+				+
волосатые	+		+		+	+	+													
Всё валится из рук				+					+		+				+	+				+
всплеснуть	+															+				+
выбить из	+																			
вывих	+																			
вывихнуть	+							+												
выпрямить																+				
выпустить из				+				+	+		+	+		+		+				
вырвать из	+																			
выронить из	+								+											

вытирать	+																			
вытянуть	+															+		+		+
гибкая					+		+													
Глаза боятся, а руки делают				+																
голыми		+		+			+		+					+		+		+		+
горячая	+				+															
Греть руки		+		+	+				+											
грубые	+																			
грязные	+		+		+		+							+		+		+		+
Давать волю рукам		+		+					+											
Дать по рукам кому-л				+					+			+								
Дать руку на отсечение		+		+																
движение м														+		+				+
движения рук														+		+		+		+
двумя														+		+		+		+
девушки					+								+				+		+	
дело рук														+		+				+
держать в	+	+		+		+								+		+		+		+
держать на	+					+		+	+											
Держать руку чью		+																		
держаться за	+													+						
детские	+																			
длинные	+		+				+							+		+		+		+

До ручки дойти									+										
добрые руки														+				+		+
достать															+					
дотянуться	+																			
дрожат		+	+				+	+							+				+	
дрожащие																+				
друга			+		+	+	+										+			
дружеская					+															
железные					+		+													
женские	+	+	+	+	+	+	+													
женщины	+																+		+	
жесткие	+				+															
заботливые																+				
загорелые	+						+													
загребущая					+	+														
зажать в	+						+													
закона			+																	
замерзли		+																		
заняты															+		+		+	
Запустить руку во что .		+		+																
здоровые	+																			
знать чью-л руку	+	+		+					+	+										
золотые		+	+	+	+		+	+	+					+		+		+		+
и ногами		+		+			+		+				+		+				+	
идти под руку	+	+		+					+											
из первых				+					+		+			+		+				+

Из рук в руки				+		+			+		+								
из рук вон плохо				+					+		+								
изящные	+		+				+												
Иметь (сильную) рукугде		+									+								
Искать чьей руки		+																	
испачкать	+																		
как без рук				+							+								
как рукой сняло									+			+			+			+	
карты в руки				+															
кисть	+													+		+		+	+
кожа	+													+		+		+	+
короткие	+				+														
корявые							+												
коснуться																+			
костлявые	+																		
красивые	+		+		+		+							+				+	
крем																+		+	+
крепкие	+				+		+												
кривые			+		+	+										+			+
ласковая			+				+												
левая	+		+	+	+	+	+	+	+	+				+		+		+	+
легкая рука		+		+	+		+				+								
ледяные	+																		
лежит													+						
лечить	+																		

Лизать руки		+		+																
линии	+																			
лишиться	+																			
Ломать руки		+		+																
маленькие	+		+		+	+	+													
малыша													+				+		+	
Марать рукиоб		+		+																
массажиста															+		+		+	
матери					+	+	+						+				+			
махать														+		+				+
махнуть	+	+		+								+		+		+		+		+
мозоли на	+																			
мозолистые	+																			
мокрые	+																			
мокрыми																+				+
морщинистые	+																			
моторика																+		+		+
мошенников																+				
мужские	+						+													
мужчины	+				+												+			
мыть руки	+	+		+			+							+		+		+		+
мышцы														+				+		+
мягкие	+		+																	
на все руки мастер									+											

на расстоянии вытянутой														+		+		+		+
На руках иметь		+		+		+			+											
На́ рукукому		+		+					+											
на скорую														+		+				+
набить		+		+							+					+		+		+
Нагреть руки		+		+					+											
надежные														+		+		+		+
надеть на	+			+					+											
наложить гипс						+			+											
Наложить руки на себя.		+		+					+											
натянуть на	+																			
не доходят				+					+				+				+		+	
не покладая рук				+					+		+			+		+		+		+
Не рука кому-л (устар.)		+							+		+									
не с руки																+		+		
не хватает рук		+																		
небольшие	+																			
нежные			+		+		+													
нести в	+		+	+																
нечистые			+	+					+							+				+
ноет		+			+															

носить на	+		+	+				+	+				+							
носят								+												
Обагрить руки кровью		+		+																
обе	+													+				+		+
обеими														+		+		+		
Обломать руки		+		+																
обхватив																+				+
одной рукой														+		+		+		+
опускаются													+		+		+		+	
опустить	+	+		+			+	+	+					+		+		+		+
опытные															+	+		+		
от руки				+																
отбиться от				+							+									
отекла		+																		
отморозить	+																			
очумелые ручки							+													
пальцы	+		+			+	+							+		+		+		+
перевязка	+																			
перелом	+		+	+																
писателя		+		+				+												
писать			+		+	+	+													
по обе руки		+																		
По рукам!		+		+					+			+								
повязка на			+																	
погладить	+	+		+			+													

под		+		+					+					+						+
под горячую														+				+		
под рукой		+		+					+							+				
Подать (или протянуть) руку (помощи)		+	+	+	+	+	+	+	+						+		+		+	
подать руку	+							+	+							+				+
подвернуться под				+					+											
подержать														+						+
поднять	+			+				+						+		+		+		+
поднять руки к небу													+	+						
Поднять рукуна		+							+											
подобрать что-л по руке	+	+		+					+											
Подписаться обеими руками под чем-л.				+								+								
пожать	+		+	+					+					+		+		+		+
показывать					+															
полные	+								+											
положа руку на сердце												+			+	+	+	+	+	+
положение														+				+		+
положить на	+			+																
попасть в																				+

поранить	+																			
порезать	+						+													
потирать																+				
правая	+	+	+	+	+	+	+	+		+	+			+		+		+		+
правосудия			+			+	+													
предложить руку и сердце		+	+	+	+		+		+				+		+	+	+		+	
прибрать к									+					+		+				+
придерживая																+		+		
прижать руки к груди														+	+	+				+
прикоснуться			+				+													
приложить		+		+					+			+				+		+		+
провести	+																			
просить руки		+		+					+					+						
протянуть	+		+	+	+			+						+		+		+		+
профессионалов														+				+		+
рабочего	+				+	+		+												
рабочие													+				+		+	
развести руками	+	+		+				+				+		+		+		+		+
развязать		+		+					+							+				+
размахивать	+															+		+		
разогнуть	+																			
рана на	+																			
раненая	+						+													
раскинув																+				+

раскинуть																+					
распухла		+																			
ребенка	+												+		+		+				
розовая			+																	+	
рука мастера													+		+		+		+		
рука на пульсе			+			+							+		+	+	+		+		
рука не дрогнет				+	+				+		+										
рука не поднимается			+	+	+				+		+		+		+				+		
Рука об руку		+	+	+	+	+	+		+		+										
рука руку моет		+	+	+	+		+		+								+				
руки в боки															+						
руки вверх		+		+			+	+						+		+		+			
руки за голову																			+		
руки за спину	+	+		+			+						+		+		+				
Руки коротки!				+			+														
Руки по швам		+		+								+									
Руки прочь от		+					+		+												
рукой подать		+							+				+				+				
с легкой руки														+		+		+		+	
с пустыми				+												+		+		+	
с руками оторвут																	+				

С рукикому		+		+																
сбыть с рук														+				+		
свободной																		+		+
своими														+		+		+		+
связать	+			+																
сделать своими руками	+					+														
сжать руки в кулак																	+			+
сжимать в			+																	
сильные	+		+	+		+	+							+				+		+
синица в																+				+
синяк на	+																			
скрестив	+	+	+				+	+								+		+		+
слабые	+																			
сложа руки		+		+										+		+		+		+
сложенные																+				
сломать руку	+		+			+	+													+
смотреть из-под	+																			
снимок (рентгеновский)	+																			
собственными														+		+		+		+
согнуть	+															+		+		
сойти с рук				+					+		+									
Сон в руку				+	+				+											

специалистов															+		+				
старика	+																				
сунуть	+																				
сунуть руки в карман	+																				
суставы																		+			
сухие	+																				
схватить за	+			+					+							+		+		+	
талантливая				+																	
твердая			+		+	+	+		+												
теплые	+		+		+		+														
товарища					+	+															
тонкие	+		+		+		+														
травма	+																				
трогать								+								+		+		+	
трясутся																+					
трясущиеся																+				+	
тяжелая рука		+	+	+	+		+														
тянется																	+		+		
Тянуть чью руку		+			+																
ударить	+		+		+		+														
ударить по	+											+									
удариться	+																				
укол в руку	+																				
Укоротить руки кому		+		+																	
умелые			+		+		+	+						+		+		+		+	

умывать		+		+												+				
ухоженные														+		+		+		+
холеные	+																			
холодные	+	+			+		+													
хорошие														+				+		+
+						+														
художника													+				+			
худые	+		+				+													
целовать	+	+						+								+				+
частные																+		+		
человека			+		+	+	+						+		+		+		+	
человеческая																		+		+
чешутся				+					+				+		+		+		+	
чистые	+		+		+	+	+		+					+		+		+		+
Что-л. само в руки идёт				+					+											
чужие руки		+							+					+		+				+
Чужими руками жар загребать				+	+	+		+												
штангиста	+																			
щедрая																+				
щедрой рукой		+																		

Приложение 4. Коллокаты для слова *белый*

	словарь сочетаемости	МАС	СИБАС	БТС	Русский ассоциативный словарь	УрРАС	ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремова	Словарь Молотковa	Словарь Мокиенко	log-dice	MI3	T-score	Log-likelihood
билет		+	+					+	+							
вино	+	+						+	+				+		+	+
ворона		+	+			+							+		+	+
железо		+														
изба		+	+													
кость		+						+								
мухи		+														
мясо		+						+	+						+	+
ночи		+			+			+	+				+		+	+
пятна	+	+				+							+		+	+
стихи		+				+		+	+							+
уголь		+						+							+	+
хлеб	+	+	+		+	+	+	+	+	+			+		+	+
ангел			+			+	+						+			
Белыми ниткам и шито		+														
ветер							+						+		+	
воротничок						+							+	+	+	+
гвардия	+	+	+					+	+	+					+	
горячка		+						+	+		+	+			+	+
гриб	+					+	+	+	+	+			+	+	+	+

Дела как сажа бела		+													
Довести до белого каления		+									+	+		+	
дом						+	+	+					+		+	+
духовенство		+						+							+	+
заяц			+		+	+	+									
зима					+	+										
и пушистый			+			+	+									
как снег			+		+											
клык			+		+	+	+						+		+	
кот			+			+	+						+		+	+
лебедь	+		+		+	+	+						+	+	+	+
лист			+		+	+	+			+			+		+	+
магия		+				+							+		+	+
медведь		+	+		+	+	+						+	+	+	
Называть белое черным		+														
облако	+		+										+		+	+
олимпиада		+											+			
пепел			+				+									
пух			+		+	+	+						+			
свет		+	+		+	+	+	+					+		+	

Сказка про белого бычка		+									+	+				
снег	+		+		+	+	+			+			+		+	+
Среди бела дня		+				+	+	+	+	+	+	+	+	+	+	+
танец					+	+	+		+				+			
флаг			+		+	+	+						+		+	+
фон	+												+	+	+	+
халат						+	+						+	+	+	+
цвет	+				+		+						+	+	+	+
человек		+			+	+	+		+	+			+		+	
Черным по белому		+														
шум			+										+	+	+	+
море													+	+	+	+
цветок													+	+		+
камень													+	+		+
платье													+	+		+
песок													+	+	+	+
золото													+	+	+	+
глина													+	+	+	+
конь													+	+	+	+
налет													+	+	+	+
краска													+	+	+	+
роза													+	+	+	+
рубашка													+	+	+	+
полоса													+	+	+	+
список													+		+	+
одежда													+		+	+

стена													+		+	+
мрамор													+		+	+
шоколад													+		+	+
город													+		+	
тигр													+		+	+
зал													+		+	+
порошок													+		+	+
экран													+		+	+
голубь													+		+	+
потолок													+		+	
сахар													+		+	
зависть													+		+	+
акула													+		+	+
зарплата													+		+	
пляж													+		+	
парус													+		+	+
береза													+		+	
лимузин													+		+	
бумага													+			+
чай													+			+
дача													+			+
кролик													+			+
орел													+			+
каление													+			+

Приложение 5. Коллокаты для слова *скакать*

	словарь сочетаемости	МАС	СИБАС	БТС	Русский ассоциативный словарь	УрРАС	ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремова	Словарь Молоткова	Словарь Мокиенко	log-dice	MI3	T-score	Log-likelihood
бегать и						+										
беззаботно													+	+	+	+
белка		+											+		+	+
бешено													+	+	+	+
блоха								+								
бодро													+	+	+	+
быстро					+	+							+		+	+
вверх-вниз													+	+	+	+
верхом									+	+						
весело													+	+	+	+
во весь дух или опор								+			+	+				
вокруг														+	+	+
воробей					+								+		+	
впереди														+	+	+
вприпрыжку													+	+	+	+
всадник													+	+	+	+
галопом													+	+		
давление			+							+			+		+	
девочка			+		+			+								
доллар													+	+	+	+
жеребенок					+											

зайчик													+	+	+	+
заяц					+			+	+				+	+	+	+
кавалерия													+	+	+	+
кенгуру														+	+	+
ковбой													+	+	+	+
козел													+	+	+	+
конь													+	+	+	+
кузнечик													+	+	+	+
лихо													+	+	+	+
ловко													+	+	+	+
лошадь		+	+		+		+	+	+				+		+	+
лягушка								+					+		+	+
мысли		+														
мысль													+	+	+	+
мяч		+														
на одной ноге.								+								
навстречу													+	+	+	+
напряжение													+	+	+	+
настроение													+	+	+	+
неуклюже													+	+	+	+
обезьяна													+	+	+	+
опрометью														+	+	+
перестать														+	+	+
посещаемость													+	+	+	
постоянно														+	+	+
проворно													+	+	+	+
прочь														+	+	+

псина														+	+	+
птица														+	+	+
пульс													+	+	+	+
радостно													+	+	+	+
резво													+	+	+	+
рядом														+	+	+
температура		+							+	+			+		+	+
тень													+	+	+	+
тройка													+	+	+	+
трусцой													+	+	+	+
туда-сюда													+	+	+	+
цены									+							
через веревочку		+						+		+						
через огонь		+														

Приложение 6. Коллокаты для слова *семь*

	словарь сочетаемости	МАС	СИБАС	БТС	Русский ассоциативный словарь	УрРАС	ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремова	Словарь Молоткова	Словарь Мокиенко	log-dice	MI3	T-score	Log-likelihood
дней			+										+	+	+	+
За семь верст киселя хлебать		+														
За семью замками		+						+								
Книга за семью печатями		+														
С. бед - один ответ									+		+	+				
седьмое небо													+		+	
седьмой десяток													+			
седьмой сын															+	
Семи пядей во лбу		+						+					+	+		
семь ангелов													+	+	+	+
семь богатырей													+	+	+	
семь вечеров														+		+

семь гномов													+	+	+	+
семь грехов														+	+	+
семь мудрецов														+		
семь нот													+		+	
Семь потов сошло с кого		+											+			
Семь пятниц на неделе		+						+			+	+		+		+
семь раз отмерь			+								+	+		+	+	+
семь холмов													+			+
семь цветов радуги													+	+	+	
семь чакр													+	+	+	+
семь чудес света													+	+		+
у семи нянек дитя без глазу													+	+		

Приложение 7. Коллокаты для слова *свой*

	словарь сочетаемости	МАС	СИБАС	БТС	Русский ассоциативный словарь	УрРАС	ЕВРАС	словарь Ушакова	словарь Ожегова	словарь Ефремова	Словарь Молоткова	Словарь Мокиенко	log-dice	MI3	T-score	Log-likelihood
Своя голова на плечах		+														
(не)В своем уме		+							+							
(не)На своем месте		+														
(Рассказать) своими словами		+								+					+	
бизнес														+	+	
Брать (взять) свое		+							+	+						
Быть не в своей тарелке		+														+
В свое время		+							+							
В свое удовольствие		+														
В своем роде .		+														
В свою очередь		+											+	+		

выбор						+	+									
дом			+			+	+			+						
Жить своим умом		+						+								
Знать свое место		+								+						
Идти своей дорогой		+														
Идти своим ходом		+														
Идти своим чередом		+														
Мастер своего дела		+														
На свой страх (и риск)		+														
На своих двоих		+														
На свою голову		+														
Называть вещи своими именами		+							+							
Не в свои сани сесть		+														

Не верить своим глазам		+														
Не своим голосом		+							+							
нести свой крест							+									
Остаться при своих		+														
отпуск за свой счет							+									
парень			+		+		+	+	+							
по-своему									+	+						
Поставить на свое место		+						+								
принцип							+									
Принять на свой счет		+					+									
Сам не свой		+														
Свое на уме		+														
Свое я		+														
Своего поля ягода		+									+	+				
Своего рода		+														

Свой брат		+	+			+				+						
Свой в доску		+	+		+			+		+						
свой народ		+			+		+						+	+	+	+
Своим порядком		+														
Своих не узнаешь		+						+	+							
Своя ноша не тянет								+			+				+	+
Своя рубаха ближе к телу.								+								
Своя рука владыка		+														
Сделать своими руками		+						+	+	+			+	+		+
Сказать свое слово		+														
Стоять на своих ногах		+														
Умереть (не) своей смертью		+														
свое дело													+	+	+	+
свое мнение													+	+	+	+

своя квартира													+	+	+	+
страна													+	+	+	+
вкус													+	+	+	+
усмотрение													+	+	+	+
свое отношение													+	+	+	+
решение													+	+	+	+
знать свой организм													+	+	+	+
свое состояние													+	+	+	+
свой компьютер													+	+	+	+
свой участок													+	+	+	+
свое лицо													+	+	+	+
подтвердить свою репутацию													+	+	+	+
свой доход													+	+	+	+
свой адрес													+	+	+	+
свой заказ													+	+	+	+

своя голова на плечах													+	+	+	+
оставить свой отзыв													+	+	+	+
своя фантазия													+	+	+	+
написать свое пожелание													+	+	+	+
дать свое согласие													+	+	+	+
указать свой телефон													+	+	+	+
свое воображение													+	+	+	+

Приложение 8. Экспертная оценка коллокаций, содержащих слово *сердце*

	Количество словарей	Количество мер	Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	среднее арифметическое	стандартное отклонение
в самое	0	1	2	2	2	2	2	2	0
девичье	0	1	2	2	2	2	2	2	0
детское	0	1	1	1	1	1	1	1	0
замерло	0	2	2	2	2	2	2	2	0
заныло	0	1	2	2	2	2	0	1,6	0,894427
злое	0	1	2	2	1	2	1	1,6	0,547723
измученное	0	1	0	0	0	0	0	0	0
исследование	0	1	0	0	0	0	0	0	0
ключик к	0	1	2	2	2	2	2	2	0
ледяное	0	1	2	2	2	2	2	2	0
любимого	0	2	0	1	0	0	0	0,2	0,447214
людей	0	1	0	0	0	1	2	0,6	0,894427
людское	1	0	0	0	0	0	2	0,4	0,894427
маленькое	0	2	0	0	0	0	0	0	0
мамы	0	1	1	1	1	1	1	1	0
мужское	0	1	0	0	0	0	0	0	0
музыка	0	1	0	0	2	0	2	0,8	1,095445
не камень	0	2	2	2	2	2	2	2	0
не прикажешь	0	2	2	2	2	2	2	2	0
недостаточность	0	2	1	1	1	1	2	1,2	0,447214
Обнаженные	0	2	0	1	0	1	2	0,8	0,83666
огрубело	0	1	0	0	1	0	1	0,4	0,547723
ожесточилось	0	1	0	0	0	0	1	0,2	0,447214
останавливается	0	1	1	1	1	1	1	1	0
от сердца отлегло	0	2	2	2	2	2	2	2	0
отдано	0	1	2	2	2	2	2	2	0
оттаяло	0	1	2	2	2	2	2	2	0
отца	0	1	1	1	1	1	1	1	0
патологии	0	1	1	1	1	1	0	0,8	0,447214
поражения	0	1	1	1	1	1	0	0,8	0,447214
прихватило	0	1	2	1	2	2	2	1,8	0,447214

радость	0	1	0	0	1	0	0	0,2	0,447214
раненое	0	1	1	1	1	1	1	1	0
риск	0	1	0	0	0	1	0	0,2	0,447214
России	0	1	1	1	1	1	1	1	0
русское	0	1	1	1	1	1	1	1	0
согреть	0	1	1	1	1	1	1	1	0
сокращения	0	2	1	1	1	1	2	1,2	0,447214
Солдатское	0	1	1	1	1	1	2	1,2	0,447214
состояние	0	2	1	2	1	1	0	1	0,707107
теплое	0	1	1	1	1	1	1	1	0
ткани	0	1	1	1	2	1	0	1	0,707107
удар	0	1	1	1	1	0	0	0,6	0,547723
УЗИ	0	2	1	1	1	1	0	0,8	0,447214
успокоилось	0	1	1	1	1	0	0	0,6	0,547723
функции	0	1	1	1	1	0	0	0,6	0,547723
четырехкамерное	0	1	1	2	1	0	2	1,2	0,83666
ЭКГ	0	1	1	2	1	1	2	1,4	0,547723
аритмия	0	3	1	1	0	0	2	0,8	0,83666
Бога	0	3	1	1	1	1	1	1	0
болезнь	0	4	1	1	1	1	0	0,8	0,447214
болеть/любить всем сердцем за кого-л	5	0	2	2	2	2	2	2	0
Брать за́ сердце	5	0	2	2	2	2	2	2	0
выпрыгнет	0	3	1	1	0	1	2	1	0,707107
давление	0	3	1	1	1	1	2	1,2	0,447214
Европы	0	4	1	1	1	1	1	1	0
женское	0	3	1	0	1	0	1	0,6	0,547723
женщины	0	4	1	1	1	1	1	1	0
забилось	0	3	2	2	1	2	1	1,6	0,547723
заболевание	0	4	1	0	1	1	0	0,6	0,547723
завоевать	0	4	2	2	2	2	2	2	0
клетки	0	3	1	0	1	0	0	0,4	0,547723
кольнуло	0	3	1	0	1	1	1	0,8	0,447214

кровообращение	0	3	1	1	1	1	2	1,2	0,447214
материнское	0	3	2	2	2	2	2	2	0
миллионов	0	3	1	1	1	1	2	1,2	0,447214
мужчины	0	4	0	0	0	0	0	0	0
нагрузка на	0	4	1	1	1	0	0	0,6	0,547723
нарушения	0	3	1	1	1	1	0	0,8	0,447214
остановка	0	3	1	1	0	1	0	0,6	0,547723
от чистого	0	4	2	2	2	2	2	2	0
отклик	0	4	0	0	0	0	0	0	0
Пармы	0	4	1	1	1	1	1	1	0
перебои	0	4	0	0	0	0	0	0	0
Петербурга	0	3	1	1	1	2	1	1,2	0,447214
плода	0	4	0	0	0	2	0	0,4	0,894427
поклонников	0	4	0	0	0	0	2	0,4	0,894427
проблемы	0	3	0	0	0	0	0	0	0
растопить	0	3	2	2	2	2	2	2	0
руку и сердце	0	3	2	2	2	2	2	2	0
сжалось	0	3	1	2	1	1	2	1,4	0,547723
согревает	0	3	1	1	1	1	1	1	0
сосуды	0	4	1	1	1	1	0	0,8	0,447214
Стальное	0	3	1	1	1	1	2	1,2	0,447214
трансплантация	0	3	0	1	0	0	0	0,2	0,447214
трепещет	0	3	1	1	0	1	2	1	0,707107
укреплять	0	4	0	0	0	0	0	0	0
в глубине сердца	3	2	1	1	1	1	2	1,2	0,447214
воспоминания	3	1	0	0	0	0	0	0	0
говяжье	1	2	0	1	0	0	0	0,2	0,447214
голос	1	1	0	1	0	0	1	0,4	0,547723
деятельность	1	1	0	0	0	0	0	0	0
екнуло	2	2	1	1	0	1	2	1	0,707107
зов	2	2	1	1	1	1	2	1,2	0,447214
мышцы	1	2	0	0	0	0	0	0	0
обследование	1	1	0	0	0	0	0	0	0
огромное	3	1	0	0	0	0	0	0	0

от глубины	2	1	0	1	0	0	2	0,6	0,894427
отзывчивое	1	1	0	1	0	0	1	0,4	0,547723
память	2	2	0	0	0	0	2	0,4	0,894427
пересадка	3	1	0	0	0	0	0	0	0
поет	1	1	1	1	1	1	2	1,2	0,447214
принадлежит	1	2	1	1	0	1	2	1	0,707107
работает	3	1	0	1	0	0	0	0,2	0,447214
радуется	3	1	2	2	2	2	2	2	0
разрывается	3	2	2	2	2	2	2	2	0
ребенка	1	2	0	0	0	0	1	0,2	0,447214
с легким	2	1	2	2	2	2	2	2	0
с тяжелым	2	1	2	2	2	2	2	2	0
страдает	1	2	1	1	1	1	1	1	0
Тоны	1	1	0	0	0	1	0	0,2	0,447214
трогать сердце	2	1	0	0	1	0	2	0,6	0,894427
Храброе	2	2	2	2	2	2	2	2	0
черствое	3	1	2	2	2	2	2	2	0
чувствует	2	1	0	0	0	1	2	0,6	0,894427
чует	1	1	0	0	0	0	2	0,4	0,894427
биение	2	4	0	2	0	0	2	0,8	1,095445
Бычье	3	3	0	0	1	0	2	0,6	0,894427
бьющееся	2	3	0	0	0	1	0	0,2	0,447214
веление	2	4	2	2	2	2	2	2	0
дама	3	3	2	0	2	2	2	1,6	0,894427
девушки	1	3	0	0	0	0	0	0	0
искусственное	2	4	1	1	1	1	0	0,8	0,447214
колотится	2	3	1	1	1	1	2	1,2	0,447214
красавицы	2	4	1	2	1	1	2	1,4	0,547723
массаж	1	3	1	1	1	1	0	0,8	0,447214
мое	3	3	0	0	0	0	0	0	0
не выдержало	1	4	1	1	2	1	0	1	0,707107
область	1	4	0	1	0	0	0	0,2	0,447214
одинокие	2	3	0	0	0	0	1	0,2	0,447214
операция	3	4	0	0	0	1	1	0,4	0,547723
открыть	3	4	1	1	1	1	2	1,2	0,447214
подсказывает	2	4	1	1	1	1	1	1	0

предложение руки и	3	4	2	2	2	2	2	2	0
работа	1	4	0	0	1	0	0	0,2	0,447214
ритм	1	4	0	0	0	0	0	0	0
с замиранием	3	4	2	2	2	2	2	2	0
с открытым	2	3	2	2	2	2	2	2	0
с чистым	3	3	2	2	2	2	2	2	0
слабое	3	3	2	2	2	2	2	2	0
собачье	3	4	2	2	2	1	2	1,8	0,447214
столицы	1	4	1	2	1	1	1	1,2	0,447214
ум	1	3	0	0	1	0	0	0,2	0,447214
Холодное	3	4	2	2	2	2	2	2	0
чуткое	1	3	2	2	2	2	2	2	0
шумы	1	3	1	1	1	1	1	1	0
щемит	3	3	1	1	1	1	2	1,2	0,447214
близко к	6	2	2	2	2	2	2	2	0
болит	7	2					2	2	#ДЕЛ/0!
больное	5	1	1	1	1	0	0	0,6	0,547723
большое	7	2	1	2	1	1	2	1,4	0,547723
горит	4	1	1	0	1	1	2	1	0,707107
золотое	8	1	2	2	2	2	2	2	0
каменное	6	1	2	2	1	2	2	1,8	0,447214
матери	4	2	1	1	1	0	1	0,8	0,447214
нежное	4	2	1	1	1	1	1	1	0
остановилось	4	1	2	2	2	2	0	1,6	0,894427
от всего сердца	5	1	2	2	2	2	2	2	0
положа руку на сердце	6	2	2	2	2	2	2	2	0
сердце моё	4	1	0	0	0	0	2	0,4	0,894427
боль в	4	4	0	0	0	1	0	0,2	0,447214
бьется	8	4	0	0	1	0	0	0,2	0,447214
Верное	4	3	2	2	2	2	2	2	0
города	4	3	0	0	0	0	0	0	0
горячее	5	4	2	2	2	2	2	2	0
доброе	9	4	2	2	2	2	2	2	0
дрогнуло	4	3	2	2	2	2	2	2	0
замирает	5	3	2	2	2	2	2	2	0

здоровое	4	3	0	0	1	0	0	0,2	0,447214
клапан	4	3	0	0	0	0	0	0	0
кровью обливается	5	4	2	2	2	2	2	2	0
Львиное	4	4	2	2	2	2	2	2	0
любящее	6	4	0	0	0	1	0	0,2	0,447214
не лежит	5	3	1	1	1	1	2	1,2	0,447214
покорить	5	4	2	2	2	2	2	2	0
порок	4	4	0	0	0	1	0	0,2	0,447214
разбитое	4	3	2	2	2	2	2	2	0
сжимается	4	4	2	2	2	2	2	2	0
скрепя	4	4	2	2	2	2	2	2	0
стук	4	3	0	0	0	0	0	0	0
стучит	6	4	0	0	1	0	0	0,2	0,447214
человека	4	4	0	0	1	0	0	0,2	0,447214
человеческое	4	4	0	0	0	0	0	0	0
чистое	4	4	1	0	1	1	2	1	0,707107
ангела	1	0	0	0	0	0	0	0	0
бездонное	1	0	0	0	1	0	2	0,6	0,894427
безумное	1	0	0	0	0	1	1	0,4	0,547723
бешеное	1	0	0	0	0	1	2	0,6	0,894427
благородство	2	0	0	0	1	1	1	0,6	0,547723
больного	1	0	0	0	0	1	0	0,2	0,447214
боязнь	1	0	0	0	0	0	0	0	0
в пятки ушло	2	0	2	2	2	2	2	2	0
в сердцах	3	0	2	2	2	2	2	2	0
великодушное	1	0	1	0	1	1	1	0,8	0,447214
взыграло	2	0	1	0	1	1	2	1	0,707107
волновать	1	0	0	0	0	0	0	0	0
вредно для	1	0	0	0	0	0	0	0	0
всем сердцем	3	0	2	2	2	2	2	2	0
выбор	1	0	0	0	0	0	0	0	0
вылетело	1	0	0	0	0	0	2	0,4	0,894427
Вынуть	1	0	1	1	1	1	0	0,8	0,447214
вырвать из	3	0	1	1	1	1	2	1,2	0,447214
выстрелить в	3	0	0	0	0	0	0	0	0

героя	2	0	1	0	1	1	0	0,6	0,547723
гибнет	1	0	1	1	1	1	1	1	0
говорит	1	0	1	1	1	1	1	1	0
дамы	1	0	0	0	0	0	2	0,4	0,894427
дать волю сердцу	1	0	2	2	1	2	2	1,8	0,447214
действует на	1	0	2	2	2	2	2	2	0
держаться за	1	0	2	2	2	2	2	2	0
до глубины	1	0	2	2	2	2	2	2	0
дочери	1	0	0	0	0	0	0	0	0
друг сердца	1	0	2	2	2	2	2	2	0
друга	1	0	1	1	1	1	1	1	0
жгучее	1	0	1	1	2	1	1	1,2	0,447214
жены	1	0	1	1	1	1	0	0,8	0,447214
жестокое	1	0	2	1	2	2	2	1,8	0,447214
живет	0	1	0	0	0	0	0	0	0
живое	3	0	0	0	0	0	0	0	0
жить в сердце	1	0	1	1	1	1	2	1,2	0,447214
жить сердцем	1	0	1	1	1	1	2	1,2	0,447214
зажечь	1	0	0	1	0	1	2	0,8	0,83666
зажигать	1	0	0	1	0	0	2	0,6	0,894427
закрыто	1	0	0	1	0	0	1	0,4	0,547723
занято	2	0	2	2	2	2	2	2	0
запасть в	1	0	2	2	2	2	2	2	0
затаить в	1	0	1	1	1	1	2	1,2	0,447214
заячье	2	0	1	1	1	1	2	1,2	0,447214
Земли	1	0	1	1	1	1	1	1	0
из камня	1	0	1	1	1	1	1	1	0
изношенное	2	0	0	0	0	0	0	0	0
изучать	1	0	0	0	1	0	0	0,2	0,447214
испортить	1	0	0	0	0	0	0	0	0
Кавказа	1	0	1	1	1	1	1	1	0
Как маслом по сердцу	2	0	1	1	1	1	2	1,2	0,447214
Как ножом по сердцу	3	0	2	2	2	2	2	2	0

камень с сердца	1	0	2	2	2	2	2	2	0
лечить	1	0	0	1	0	1	0	0,4	0,547723
льва	1	0	0	1	0	0	0	0,2	0,447214
любимое	3	0	0	0	1	0	0	0,2	0,447214
массировать	1	0	0	0	0	1	0	0,2	0,447214
молодое	2	0	0	0	1	1	0	0,4	0,547723
молчит	1	0	1	1	1	1	1	1	0
мощное	1	0	0	0	0	0	0	0	0
мудрого человека	1	0	0	0	0	0	0	0	0
На с. накипело, наболело	1	0	2	2	2	2	2	2	0
На с. тяжело, легко, тоскливо	1	0	2	2	2	2	2	2	0
На сердце кошки скребут	3	0	2	2	2	2	2	2	0
надрывать	3	0	0	0	0	0	2	0,4	0,894427
не обманет	1	0	2	2	2	2	2	2	0
не хватит	1	0	0	0	1	0	1	0,4	0,547723
нет сердца	2	0	2	2	2	2	2	2	0
Ничто не шевельнулось в моём с.	1	0	2	2	2	2	2	2	0
новое	1	0	0	0	0	0	0	0	0
оборвалось	2	0	1	1	1	1	1	1	0
ожирение	1	0	0	0	0	0	0	0	0
оперировать	1	0	0	0	0	1	0	0,2	0,447214
от полноты	2	0	1	1	1	1	1	1	0
отважное	1	0	1	1	1	1	1	1	0
отошло	3	0	1	1	1	1	2	1,2	0,447214
падает	3	0	2	2	2	2	2	2	0
перевернулось	1	0	2	2	2	2	2	2	0
пересаживать	1	0	0	1	0	0	0	0,2	0,447214
пламенное	3	0	0	1	0	0	2	0,6	0,894427
победить	1	0	1	2	1	1	1	1,2	0,447214
полезно для	1	0	0	0	0	0	0	0	0
попасть в	2	0	0	0	0	1	0	0,2	0,447214

пополам	1	0	1	1	1	1	1	1	0
потрясти	2	0	1	1	1	1	1	1	0
преданное	1	0	1	1	1	1	2	1,2	0,447214
прижать к	3	0	2	2	2	2	2	2	0
принимает к	1	0	1	1	1	1	1	1	0
проверять	1	0	0	1	0	1	1	0,6	0,547723
продолговатое	2	0	0	0	0	0	0	0	0
пронзить	3	1	0	0	0	0	1	0,2	0,447214
просит	1	0	1	1	1	1	1	1	0
простое	1	0	1	1	2	1	1	1,2	0,447214
пылкое	1	0	1	1	1	1	1	1	0
разбито	3	0	2	2	2	2	2	2	0
разрыв	2	0	1	1	2	1	2	1,4	0,547723
ранимое	1	0	1	1	1	1	1	1	0
растравить	1	0	1	1	1	1	1	1	0
растревожить	2	0	1	2	2	1	2	1,6	0,547723
рвется	2	0	0	0	0	1	1	0,4	0,547723
режет	1	0	0	0	0	0	2	0,4	0,894427
родное	1	0	1	1	1	1	1	1	0
С. сердцу весть подаёт	1	0	1	1	1	1	1	1	0
светлое	2	0	1	2	1	1	1	1,2	0,447214
свинец на	1	0	1	1	1	1	1	1	0
свиное			0	0	0	0	0	0	0
сдает	1	0	0	0	0	0	2	0,4	0,894427
сильное	2	0	0	0	0	0	0	0	0
смелое	1	0	1	1	2	1	1	1,2	0,447214
спокойное	1	0	1	2	1	1	1	1,2	0,447214
спортсмена	1	0	0	0	0	1	0	0,2	0,447214
старика	1	0	0	0	0	0	0	0	0
старое	1	0	0	0	0	2	0	0,4	0,894427
страны	1	0	1	1	1	1	1	1	0
схватиться за	3	0	2	1	2	2	2	1,8	0,447214
теплое	2	0	0	0	0	0	1	0,2	0,447214
тревожное	1	0	0	0	0	0	0	0	0
трепетное	1	0	1	0	1	1	1	0,8	0,447214

труса	1	0	1	1	1	1	1	1	0
трусливое	2	0	1	1	1	1	1	1	0
Ужалить в самое с	1	0	1	1	2	1	2	1,4	0,547723
Уколоть в самое с	1	0	1	0	2	1	2	1,2	0,83666
хорошее	2	0	0	0	0	0	0	0	0
хрупкое	2	0	1	1	1	1	1	1	0
честное	1	0	1	1	1	1	1	1	0
Читать в	2	0	1	1	1	1	1	1	0
чувствительное	2	0	0	2	0	0	1	0,6	0,894427
чувствовать	3	0	0	0	0	2	2	0,8	1,095445
шалит	2	0	1	1	1	1	2	1,2	0,447214
широта	1	0	1	1	1	1	2	1,2	0,447214
щемит	3	0	1	1	1	1	1	1	0
юноши	2	0	0	0	0	2	0	0,4	0,894427
Держать сердце на кого-либо	4	0	2	1	2	2	2	1,8	0,447214
красное	4	0	0	0	0	0	0	0	0
не на месте	5	0	2	2	2	2	2	2	0
ноет	4	0	1	1	1	1	1	1	0
отдавать	4	0	2	2	2	2	2	2	0
открытое	4	0	0	0	1	0	1	0,4	0,547723
отлегло от	4	0	2	2	2	2	2	2	0
по сердцу	5	0	2	2	2	1	2	1,8	0,447214
разбить	4	0	2	2	2	1	2	1,8	0,447214
Сорвать сердце на ком-чем	5	0	2	2	2	2	2	2	0

Приложение 9. Экспертная оценка коллокаций, содержащих слово *вода*

	Количество словарей	Количество мер	Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	среднее арифметическое	стандартное отклонение
бытовая	0	2	0	1	0	0	0	0,2	0,447214
используется	0	2	0	0	0	0	0	0	0
колодец	0	1	0	0	0	0	0	0	0
кулер с	0	2	1	1	1	1	0	0,8	0,447214
обезжелезивание	0	1	0	1	0	1	2	0,8	0,83666
обеззараживание	0	1	0	0	0	0	0	0	0
обычная	0	1	1	1	1	1	0	0,8	0,447214
паводковые	0	1	1	1	1	1	2	1,2	0,447214
под водой	0	1	0	0	1	0	0	0,2	0,447214
поступает	0	1	0	0	0	0	0	0	0
прогревается	0	1	0	0	0	0	0	0	0
протечка	0	1	0	0	0	0	0	0	0
проходит	0	2					0	0	#ДЕЛ/0!
резервуар	0	2	0	0	0	0	0	0	0
сливать	0	2	0	0	0	0	0	0	0
слить	0	2	0	0	1	0	0	0,2	0,447214
смочить	0	1	0	0	0	0	0	0	0
смывается	0	1	0	0	0	0	0	0	0
состоит	0	2	0	0	0	1	0	0,2	0,447214
стоит	0	2	1	1	1	1	1	1	0
толща	0	2	0	0	0	0	1	0,2	0,447214
уровень	0	2	1	1	1	1	1	1	0
циркуляция	0	1	0	1	0	0	0	0,2	0,447214
анализ	0	4	0	0	0	0	0	0	0

артезианская	0	3	1	1	1	1	2	1,2	0,447214
бак	0	3	0	0	0	0	0	0	0
бутилированная	0	4	0	0	0	0	2	0,4	0,894427
ванна	0	4	0	0	0	1	0	0,2	0,447214
вещества	0	4	0	0	1	1	0	0,4	0,547723
воды моря	0	3	1	1	1	1	0	0,8	0,447214
грунтовые	0	4	1	1	1	1	2	1,2	0,447214
добавить	0	4	0	0	0	0	0	0	0
емкость	0	4	0	0	0	0	0	0	0
жесткость	0	4	1	1	1	1	1	1	0
залить водой	0	4	0	0	0	0	0	0	0
из воды	0	3	0	2	0	0	0	0,4	0,894427
из скважины	0	3	2	2	2	2	1	1,8	0,447214
Кавказские	0	4	2	2	2	2	2	2	0
капля	0	3	0	0	0	0	0	0	0
кастрюля	0	4	0	0	0	0	0	0	0
качество	0	4	0	0	0	0	0	0	0
кипяченая	0	4	1	1	1	1	2	1,2	0,447214
кипящая	0	3	0	0	2	0	1	0,6	0,894427
комнатная	0	3	1	1	1	1	0	0,8	0,447214
ложка	0	4	0	0	0	0	0	0	0
менять	0	4	0	1	0	2	0	0,6	0,894427
молекулы	0	3	0	2	1	0	0	0,6	0,894427
обливание водой	0	3	2	2	2	2	1	1,8	0,447214
обработка	0	4	0	0	2	1	0	0,6	0,894427
околоплодные	0	3	2	2	2	2	2	2	0

ополоснуть	0	4	1	1	1	1	1	1	0
отсутствие	0	4	1	1	1	1	0	0,8	0,447214
охлаждение	0	4	0	2	1	1	0	0,8	0,83666
погружение	0	3	0	0	0	0	0	0	0
подземные		4	1	1	1	1	2	1,2	0,447214
подкисленная	0	3	1	0	0	0	2	0,6	0,894427
подсоленная	0	3	1	1	1	1	1	1	0
полив	0	3	1	1	1	1	1	1	0
прибрежных	0	4	1	1	1	1	2	1,2	0,447214
простая	0	4	1	1	1	1	0	0,8	0,447214
расход	0	4	1	1	1	1	0	0,8	0,447214
сброс	0	3	1	1	1	1	0	0,8	0,447214
свойства	0	4	0	0	0	0	0	0	0
слив	0	4	0	0	0	0	0	0	0
состав		4	0	0	0	0	0	0	0
стекает	0	4	0	0	0	0	0	0	0
сточные	0	4	2	2	2	2	2	2	0
счетчик	0	4	1	1	1	1	0	0,8	0,447214
тазик	0	3	0	0	0	0		0	0
термальная	0	4	2	1	2	2	2	1,8	0,447214
требуется	0	4	0	0	0	0	0	0	0
употреблять	0	4	0	0	1	0	0	0,2	0,447214
фильтры для	0	3	0	0	1	0	1	0,4	0,547723
закипит	3	2	1	1	1	1	1	1	0
залить	1	1	0	0	0	0	0	0	0
из-под крана	3	2	1	1	1	1	1	1	0
количество	1	1	0	0	2	0	0	0,4	0,894427

поверхность	1	1	0	1	0	0	0	0,2	0,447214
фильтрация	1	2	0	0	1	0	0	0,2	0,447214
хлорированная	3	2	1	1	1	1	2	1,2	0,447214
бассейн	1	4	0	0	0	0	0	0	0
бутылка	2	4	0	0	0	0	0	0	0
ведро	1	4	0	0	0	0	0	0	0
вкус	1	4	0	0	0	0	0	0	0
водопроводная	2	4	2	2	2	2	2	2	0
выпить	2	3	0	0	1	0	0	0,2	0,447214
давление	1	4	0	0	0	0	0	0	0
дать	1	4	1	1	1	1	0	0,8	0,447214
движение	1	4	0	0	0	1	0	0,2	0,447214
дистиллированная	3	4	1	1	1	1	2	1,2	0,447214
дождевая	2	4	1	1	1	1	2	1,2	0,447214
доставка	1	4	0	0	0	0	0	0	0
загрязнения	1	4	1	1	1	1	0	0,8	0,447214
запас	1	4	1	1	1	1	0	0,8	0,447214
из источника	1	3	1	1	1	1	0	0,8	0,447214
испарения	1	4	0	0	1	0	1	0,4	0,547723
использовать	1	3	0	0	0	0	0	0	0
колодезная	3	3	1	1	1	1	2	1,2	0,447214
ледяная	2	4	1	1	0	1	1	0,8	0,447214
мутная	3	3	1	0	1	1	0	0,6	0,547723
мыльная	1	4	1	2	1	1	1	1,2	0,447214

мягкая	2	4	1	1	1	0	2	1	0,707107
нагрев	1	3	0	0	0	0	0	0	0
налить	1	4	0	0	0	0	0	0	0
напор	1	4	0	0	0	0	0	0	0
негазированная	1	3	1	1	1	1	1	1	0
озера	2	4	1	1	0	1	0	0,6	0,547723
океана	1	4	1	1	1	1	0	0,8	0,447214
очистка воды	1	4	0	0	1	0	0	0,2	0,447214
очищенная	1	4	1	1	1	1	0	0,8	0,447214
пить воду	1	4	0	0	0	0	0	0	0
подача	1	4	0	0	0	0	0	0	0
поток	1	3	0	0	0	0	0	0	0
потребление	1	4	1	0	1	0	0	0,4	0,547723
пресная	3	3	2	2	2	2	2	2	0
промыть водой	1	3	0	1	1	1	0	0,6	0,547723
проточная	2	3	1	1	1	1	2	1,2	0,447214
прохладная	1	4	0	0	2	0	0	0,4	0,894427
разбавить	1	4	1	1	1	1	0	0,8	0,447214
развести водой	1	4	1	1	1	1	0	0,8	0,447214
растворить в	1	4	1	1	1	1	0	0,8	0,447214
розовая	2	3	2	1	0	2	2	1,4	0,894427
свежая	2	2	0	0	0	0	0	0	0
содержит	1	3	0	0	0	0	0	0	0
соленая	3	4	1	1	1	1	0	0,8	0,447214
струя	1	4	0	0	0	0	0	0	0
талая	1	4	1	1	1	1	2	1,2	0,447214

температура	1	4	0	0	0	0	0	0	0
течение	1	4	0	0	0	0	0	0	0
течет	3	4	0	0	0	0	0	0	0
туалетная	3	4	2	2	2	2	2	2	0
родниковая	4	1	2	2	2	2	2	2	0
газированная	8	4	2	2	2	2	2	2	0
горячая	4	4	1	2	1	2	0	1,2	0,83666
жесткая	4	3	1	1	1	1	2	1,2	0,447214
живая	5	4	2	2	2	2	2	2	0
минеральная	7	4	1	1	1	1	2	1,2	0,447214
морская	7	4	1	1	1	1	2	1,2	0,447214
питьевая	5	4	1	2	1	2	2	1,6	0,547723
прозрачная	4	4	0	0	1	0	0	0,2	0,447214
святая	4	4	2	2	2	2	2	2	0
стакан	7	4	2	2	2	2	0	1,6	0,894427
теплая	4	4	1	1	1	1	0	0,8	0,447214
холодная	4	4	1	1	1	1	0	0,8	0,447214
чистая	5	4	0	1	0	0	2	0,6	0,894427
чистейшей	4	4	2	2	2	2	2	2	0
Байкала	1	0	1	1	1	1	0	0,8	0,447214
бесцветная	1	0	0	0	0	1	0	0,2	0,447214
большая	2	0	1	1	1	1	2	1,2	0,447214
бочка	1	0	0	0	1	0	0	0,2	0,447214
бочка для	2	0	0	0	0	0	0	0	0
бочка с	1	0	0	1	0	0	0	0,2	0,447214
бросить в	1	0	0	0	0	0	0	0	0
броситься в	1	0	0	0	0	0	0	0	0

брызги	1	0	0	0	1	0	0	0,2	0,447214
бурлит	1	0	0	0	0	1	0	0,2	0,447214
Буря в стакане воды	2	0	2	0	2	2	2	1,6	0,894427
бутылка из-под	1	0	0	0	0	0	0	0	0
бутылка с	1	0	0	1	0	1	0	0,4	0,547723
вешние воды	2	0	2	2	2	2	2	2	0
Вилами на воде писано	3	0	2	2	2	2	2	2	0
вкусная	3	0	0	1	0	0	0	0,2	0,447214
внутренние	3	0	2	2	2	2	0	1,6	0,894427
воды отошли	2	0	2	2	2	2	2	2	0
Возить воду	3	0	0	0	0	0	0	0	0
войти в	1	0	0	0	1	0	0	0,2	0,447214
Волги	3	0	1	1	1	1	0	0,8	0,447214
вольная	2	0	1	1	1	1	0	0,8	0,447214
вскипятить	1	0	1	1	1	1	1	1	0
выйти из	1	0	0	0	0	0	0	0	0
выкачать	1	0	0	0	0	0	0	0	0
вылить	1	0	0	0	0	0	0	0	0
вынырнуть из	1	0	1	1	1	1	0	0,8	0,447214
выпустить	1	0	0	0	0	0	0	0	0
высокая	3	0	1	1	1	1	2	1,2	0,447214
вытекла	1	0	0	0	0	0	0	0	0
вышла из берегов	2	0	1	1	1	1	2	1,2	0,447214
глотать	1	0	0	1	0	0	0	0,2	0,447214

глоток	1	0	0	0	0	1	0	0,2	0,447214
голубая	3	0	0	0	1	0	0	0,2	0,447214
горькая	2	0	0	1	0	0	0	0,2	0,447214
графин	1	0	1	1	1	1	0	0,8	0,447214
графин для	1	0	1	1	1	1	0	0,8	0,447214
графин с	1	0	1	1	1	1	0	0,8	0,447214
дезинфицировать	1	0	0	1	0	1	0	0,4	0,547723
для мытья	1	0	0	0	1	1	0	0,4	0,547723
для питья	1	0	0	0	1	0	0	0,2	0,447214
для поливки	1	0	0	0	0	1	0	0,2	0,447214
для стирки	1	0	0	0	0	0	0	0	0
для технических нужд	1	0	0	1	0	0	2	0,6	0,894427
для хозяйственных нужд	1	0	0	1	0	0	1	0,4	0,547723
для чая	1	0	0	0	0	0	0	0	0
Днепра	1	0	1	1	1	1	0	0,8	0,447214
добавлять	1		0	0	0	0	0	0	0
дорогая	1	0	0	0	0	0	0	0	0
жидкая	1	0	0	0	0	0	0	0	0
жить без	1	0	0	0	1	0	0	0,2	0,447214
жить в	1	0	0	0	0	0	0	0	0
журчит	1	0	0	0	0	1	1	0,4	0,547723
загрязнять	1	0	1	1	1	1	0	0,8	0,447214
замерзла	1	0	0	0	1	0	1	0,4	0,547723
запах	1	0	0	0	0	1	0	0,2	0,447214

запить	1	0	1	1	1	1	0	0,8	0,447214
затопила	1	0	1	0	0	0	0	0,2	0,447214
зачерпнуть	1	0	0	0	1	0	0	0,2	0,447214
здешняя	1	0	0	0	0	0	0	0	0
зеленая	2	0	0	0	0	0	0	0	0
идти за	3	0	0	0	1	0	0	0,2	0,447214
ижевская	1	0	1	1	1	1	0	0,8	0,447214
из колодца	2	0	1	1	0	1	0	0,6	0,547723
из родника	1	0	1	1	1	1	0	0,8	0,447214
избыток	1	0	0	0	0	0	0	0	0
иметь в составе	1	0	0	1	0	0	0	0,2	0,447214
искать	1	0	0	0	0	1	0	0,2	0,447214
испарилась	1	0	0	0	0	0	0	0	0
Как (будто, словно) в воду опущенный	3	0	2	2	2	2	2	2	0
Как в воду глядел	3	0	2	2	2	2	2	2	0
Как в воду канул	3	0	2	2	2	2	2	2	0
Как водой смыло	2	0	2	2	2	2	2	2	0
Как две капли воды	3	0	2	2	1	2	2	1,8	0,447214
камень точит	2	0	2	2	2	2	2	2	0
канистра	1	0	0	2	0	0	0	0,4	0,894427
капает	1	0	0	0	1	0	0	0,2	0,447214
капля	2	0	0	0	0	1	0	0,2	0,447214
качать	1	0	0	1	0	0	0	0,2	0,447214

кипячение	1	0	0	0	0	0	1	0,2	0,447214
кислая	1	0	0	0	1	0	0	0,2	0,447214
концы в воду	3	0	2	2	2	2	2	2	0
кончилась	1	0	0	0	0	0	0	0	0
котел для	1	0	0	1	0	0	0	0,2	0,447214
кружка	2	0	1	1	1	1	0	0,8	0,447214
кувшин для	1	0	1	1	1	1	0	0,8	0,447214
купаться в	1	0	0	0	0	1	0	0,2	0,447214
летать над	1	0	0	0	0	0	0	0	0
лечебная	3	0	1	1	1	1	0	0,8	0,447214
лимонная	1	0	1	0	1	1	0	0,6	0,547723
Лить воду	3	0	1	1	1	1	1	1	0
лишить	1	0	0	0	0	0	0	0	0
льется	2	0	0	0	0	0	0	0	0
любить	1	0	0	0	0	0	0	0	0
малая	2	0	0	0	0	0	0	0	0
мертвая	2	0	2	2	2	2	2	2	0
местная	1	0	0	0	1	0	0	0,2	0,447214
московская	1	0	0	1	0	0	0	0,2	0,447214
Мутить воду	2	0	1	1	0	1	2	1	0,707107
мыться	1	0	0	0	0	0	0	0	0
Набрать воды в рот	3	0	2	2	2	2	2	2	0
наглотаться	1	0	1	0	1	1	0	0,6	0,547723
найти	1	0	0	0	0	1	0	0,2	0,447214
наклонитьс я к	1	0	0	0	1	0	0	0,2	0,447214
наличие	1	0	0	0	0	0	0	0	0

наполнить	1	0	0	1	0	0	0	0,2	0,447214
направиться к	1	0	0	0	1	0	0	0,2	0,447214
находиться в	1	0	0	1	0	0	0	0,2	0,447214
невкусная	1	0	0	0	1	0	0	0,2	0,447214
недостаток	1	0	0	0	0	1	0	0,2	0,447214
нейтральные	2	0	0	1	1	0	2	0,8	0,83666
низкая	1	0	0	0	0	0	1	0,2	0,447214
носить	1	0	0	1	1	0	0	0,4	0,547723
носить решетом воду	3	0	1	1	1	1	2	1,2	0,447214
обеспечить	1	0	0	0	0	0	0	0	0
обитать в	1	0	0	1	0	1	0	0,4	0,547723
облить	1	0	0	0	1	0	0	0,2	0,447214
обнаружить в	1	0	0	1	1	1	0	0,6	0,547723
обнаружить воду	1		0	0	0	1	0	0,2	0,447214
обрызгать	1	0	0	0	0	0	0	0	0
обтираться	1	0	1	1	1	1	0	0,8	0,447214
обходиться без	1	0	0	0	1	0	0	0,2	0,447214
обыкновенная	1	0	0	0	0	0	0	0	0
окунуть	1	0	0	0	0	0	0	0	0
опустить в	2	0	0	0	0	0	0	0	0
оставаться в	1	0	0	0	0	0	0	0	0
остаться без	1	0	0	0	0	0	0	0	0
остыла	1	0	0	0	0	0	0	0	0
отделиться от	1	0	0	0	0	0	0	0	0
отключить	2	0	0	0	0	0	0	0	0

оторваться от	1	0	0	0	0	0	0	0	0
отравить	1	0	0	0	0	0	0	0	0
отравленная	1	0	0	0	0	0	0	0	0
охладить	1	0	0	0	0	0	0	0	0
очистить	1	0	0	0	0	0	0	0	0
перевозка	1	0	0	0	1	0	0	0,2	0,447214
перекрыть	1	0	1	1	1	1	0	0,8	0,447214
перелить	1	0	0	0	0	0	0	0	0
переправлять	1	0	0	0	0	0	0	0	0
плескаться в	1	0	0	0	0	0	1	0,2	0,447214
плохая	1	0	0	0	0	0	0	0	0
плыть по/ под	2	0	0	0	0	0	0	0	0
плыть против	2	0	0	0	0	0	0	0	0
погрузить в	1	0	0	0	1	0	0	0,2	0,447214
подавать	1	0	0	0	0	0	0	0	0
подогрев	1	0	0	0	0	0	0	0	0
подогретая	1	0	0	1	0	0	0	0,2	0,447214
подогреть	1	0	0	0	0	0	0	0	0
подойти к	1	0	0	0	0	0	0	0	0
поить	1	0	0	0	0	0	0	0	0
показаться из	1	0	0	0	0	1	0	0,2	0,447214
полая	1	0	1	1	1	1	0	0,8	0,447214
полная	1	0	2	2	2	2	0	1,6	0,894427
положить в	1	0	0	0	1	0	0	0,2	0,447214
послать за	1	0	0	0	0	0	0	0	0
потребность в	1	0	0	0	0	0	0	0	0
появиться из	1	0	0	0	11	0	0	2,2	4,91935

превратилась в пар	1	0	1	1	1	1	0	0,8	0,447214
превращение в пар	1	0	0	1	0	0	0	0,2	0,447214
предпочитать	1	0	0	0	0	0	0	0	0
прибывает	2	0	1	1	1	1	1	1	0
приводит в движение	1	0	0	0	0	0	0	0	0
привыкнуть к	1	0	0	0	0	0	0	0	0
принести	1	0	0	0	0	0	0	0	0
пробовать	1	0	0	0	0	0	0	0	0
пролить	1	0	0	0	0	0	0	0	0
прописать	1	0	1	1	1	1	0	0,8	0,447214
прополоскать	1	0	1	1	1	1	1	1	0
прополоскать в трёх водах	2	0	2	2	2	2	1	1,8	0,447214
пропускать	1	0	0	0	0	0	0	0	0
прорвала	1	0	1	1	1	1	0	0,8	0,447214
просачивается	1	0	0	0	0	0	0	0	0
прыгнуть в	1	0	0	0	0	0	0	0	0
пустить	1	0	0	0	0	0	0	0	0
работа под	1	0	0	0	0	0	0	0	0
работать под	2	0	0	0	0	0	0	0	0
разлить	1	0	0	0	0	0	0	0	0
размыла	1	0	0	0	0	0	0	0	0
размыть водой	1	0	1	1	1	1	0	0,8	0,447214
реки	1	0	1	1	1	1	0	0,8	0,447214
рекомендовать	1	0	0	0	1	0	0	0,2	0,447214
ржавая	1	0	0	0	0	0	0	0	0
С лица не воду пить	1	0	2	2	2	2	2	2	0

сельтерская	1	0	2	2	2	2	2	2	0
сесть на	1	0	0	0	0	0	2	0,4	0,894427
синяя	1	0	0	0	0	0	0	0	0
скользить по	1	0	0	0	1	0	0	0,2	0,447214
скопление	1	0	0	0	0	0	0	0	0
сладкая	2	0	0	1	0	0	0	0,2	0,447214
слой	1	0	0	0	0	0	0	0	0
смешать с	1	0	0	0	0	0	0	0	0
смотреть на	1	0	0	0	0	1	0	0,2	0,447214
смыла	1	0	0	0	0	0	0	0	0
смыть водой	1	0	0	1	0	0	0	0,2	0,447214
снабжать	1	0	0	0	1	0	0	0,2	0,447214
снабжение	1	0	0	0	0	1	0	0,2	0,447214
снести водой	1	0	0	1	0	0	0	0,2	0,447214
содержание в чем-л	1	0	0	1	0	0	0	0,2	0,447214
содержится в	1	0	0	0	0	0	0	0	0
содовая	1	0	1	1	1	1	0	0,8	0,447214
сосуд с	1	0	0	0	0	0	0	0	0
спокойная	1	0	0	0	0	0	0	0	0
спустить корабль на	3	0	1	1	1	1	2	1,2	0,447214
стакан с	1	0	1	1	1	1	0	0,8	0,447214
столовая	1	0	0	0	0	0	0	0	0
стоячая	2	0	1	1	1	1	2	1,2	0,447214
струится	1	0	0	0	0	0	0	0	0
студеная	2	0	1	1	1	1	1	1	0
Тише воды, ниже травы	3	0	1	0	1	1	2	1	0,707107

удельный вес	1	0	0	0	0	0	0	0	0
умываться	1	0	0	0	1	0	0	0,2	0,447214
унесла	1	0	0	0	0	0	0	0	0
уровень	1	0	0	0	0	0	0	0	0
уронить в	1	0	0	0	0	0	0	0	0
Утопить в ложке воды	2	0	1	1	1	1	0	0,8	0,447214
химический состав	1	0	0	0	1	0	0	0,2	0,447214
Холодной водой окатить (или облить)	2	0	2	2	2	2	2	2	0
хорошая	2	0	0	1	0	0	0	0,2	0,447214
хранение	1	0	0	0	0	0	0	0	0
цвет	1	0	0	0	0	1	0	0,2	0,447214
целебная	2	0	0	0	0	0	0	0	0
цистерна для	1	0	0	0	0	0	0	0	0
Чающие движения воды	1	0	2	2	2	2	0	1,6	0,894427
Черного моря	1	0	1	1	1	1	0	0,8	0,447214
шумит	2	0	0	0	0	0	0	0	0
энергия	1	0	0	0	0	0	0	0	0
бежит	4	0	1	1	1	1	1	1	0
В мутной воде рыбу ловить	5	0	2	2	2	2	2	2	0
Водой не разлить (не разольешь) кого	4	0	2	2	2	2	2	2	0
Воды не замутит	4	0	2	2	2	2	2	2	0
возить	4	0	1	1	0	1	0	0,6	0,547723

Вывести на чистую воду	5	0	2	2	2	2	2	2	0
Выйти сухим из воды	4	0	2	2	2	2	2	2	0
вылить	4	0	0	0	0	0	0	0	0
грязная	4	0	0	0	0	0	0	0	0
Как рыба в воде	4	0	2	2	2	2	2	2	0
Как с гуся вода	5	0	2	2	0	2	2	1,6	0,894427
ключевая	4	0	1	0	1	1	2	1	0,707107
лечиться на водах	4	0	2	2	2	2	2	2	0
Лить воду на мельницу	4	0	2	2	2	2	2	2	0
Много воды утекло	5	0	2	2	2	2	2	2	0
Пройти огонь и воду (и медные трубы)	4	0	2	2	2	2	2	2	0
путешествие по	4	0	0	0	0	0	0	0	0
речная	4	0	0	0	0	0	0	0	0
Седьмая вода на киселе	4	0	2	2	2	2	2	2	0
сырая	4	0	1	1	1	1	1	1	0
Темна вода во облацех	4	0	2	2	2	2	2	2	0
территориальные воды	4	0	1	1	0	1	2	1	0,707107
Толочь воду (в ступе);	4	0	2	1	2	2	2	1,8	0,447214
Тяжелая вода	4	0	1	0	1	1	2	1	0,707107
фруктовая	4	0	2	2	1	1	1	1,4	0,547723

Приложение 10. Экспертная оценка коллокаций, содержащих слово *рука*

	Количество словарей	Количество мер	Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	среднее арифметическое	стандартное отклонение
брать за	0	1	1	1	1	1	0	0,8	0,447214
брать инициативу в свои	0	1	2	2	2	2	2	2	0
взмахнуть	0	1	1	1	0	1	1	0,8	0,447214
власть в руках	0	1	2	2	2	2	2	2	0
воздеть руки	0	2	2	2	2	2	2	2	0
выпрямить	0	1	0	0	0	0	0	0	0
достать	0	1	0	0	1	0	0	0,2	0,447214
дрожащие	0	1	1	1	1	1	0	0,8	0,447214
заботливые	0	1	1	1	1	1	2	1,2	0,447214
коснуться	0	1	0	1	0	0	0	0,2	0,447214
лежит	0	1	0	0	0	1	0	0,2	0,447214
мокрыми	0	2	2	2	2	2	0	1,6	0,894427
мошенников	0	1	1	1	1	0	0	0,6	0,547723
не с руки	0	2	2	2	2	2	0	1,6	0,894427
обхватив	0	2	1	1	1	1	0	0,8	0,447214
опытные	0	2	1	1	0	1	2	1	0,707107
под горячую	0	2	2	2	2	2	2	2	0
подержать	0	2	1	1	1	1	0	0,8	0,447214
поднять руки к небу	0	2	1	1	1	1	0	0,8	0,447214
потирать	0	2	0	0	0	0	1	0,2	0,447214

придержива я	0	2	0	0	0	0	0	0	0
раскинув	0	2	1	0	1	1	1	0,8	0,44721 4
раскинуть	0	1	1	1	1	1	0	0,8	0,44721 4
руки в боки	0	1	1	1	0	1	2	1	0,70710 7
руки за голову	0	1	2	2	2	2	1	1,8	0,44721 4
с руками оторвут	0	1	2	2	2	2	2	2	0
сбыть с рук	0	2	1	1	1	1	2	1,2	0,44721 4
свободной	0	2	1	1	1	0	0	0,6	0,54772 3
сжать руки в кулак	0	2	2	2	2	2		2	0
синица в	0	2	2	2	2	2	2	2	0
сложенные	0	2	1	1	1	0	0	0,6	0,54772 3
специалист ов	0	2	1	1	1	1	0	0,8	0,44721 4
суставы	0	1	0	0	0	0	0	0	0
трясутся	0	1	0	1	0	0	1	0,4	0,54772 3
трясущиеся	0	2	0	0	0	0	0	0	0
тянется	0	2	0	0	1	0	0	0,2	0,44721 4
частные	0	2	1	1	1	1	2	1,2	0,44721 4
человеческа я	0	2	0	0	0	0	0	0	0
щедрая	0	1	0	0	0	0	2	0,4	0,89442 7
вдоль туловища	0	4	0	1	0	0	1	0,4	0,54772 3
взявшись за	0	3	0	0	0	0	0	0	0
движением	0	3	1	1	1	1	2	1,2	0,44721 4
движения рук	0	4	0	0	0	0	0	0	0
двумя	0	4	1	1	1	1	0	0,8	0,44721 4

дело рук	0	3	1	1	0	1	2	1	0,707107
добрые руки	0	3	1	1	1	1	2	1,2	0,447214
заняты	0	3	0	0	0	0	0	0	0
крем для	0	3	0	0	0	0	0	0	0
малыша	0	3	0	0	0	0	0	0	0
массажиста	0	3	0	0	0	0	0	0	0
махать	0	3	0	0	1	0	0	0,2	0,447214
моторика	0	3	0	0	0	0	0	0	0
мышцы	0	3	0	0	0	0	0	0	0
на расстоянии вытянутой	0	4	1	1	1	1	1	1	0
на скорую	0	3	1	1	1	1	2	1,2	0,447214
надежные	0	4	1	1	1	1	2	1,2	0,447214
обеими	0	3	0	0	1	0	0	0,2	0,447214
одной рукой	0	4	0	0	0	0	0	0	0
опускаются	0	4	1	1	1	1	2	1,2	0,447214
положа руку на сердце	0	3	2	2	2	2	2	2	0
положение	0	3	0	0	0	0	0	0	0
прижать руки к груди	0	3	0	0	1	0	0	0,2	0,447214
профессионалов	0	3	1	1	1	1	1	1	0
рабочие	0	3	0	0	0	0	2	0,4	0,894427
рука мастера	0	4	1	0	1	1	2	1	0,707107
с легкой руки	0	4	2	2	2	2	2	2	0
своими	0	4	1	1	0	1	1	0,8	0,447214
собственными	0	4	1	1	1	1	1	1	0

ухоженные	0	4	0	0	0	0	0	0	0
хорошие	0	3	0	0	0	0	1	0,2	0,447214
всплеснуть	1	2	0	1	0	1	1	0,6	0,547723
держаться за	1	1	0	0	0	0	0	0	0
женщины	1	2	0	0	0	0	0	0	0
размахивать	1	2	0	1	0	1	0	0,4	0,547723
розовая	1	1	0	0	0	1	0	0,2	0,447214
согнуть	1	2	0	0	0	0	0	0	0
Без рук!	2	1	2	2	2	2	2	2	0
как рукой сняло	2	2	2	2	2	2	2	2	0
мужчины	2	1	0	0	0	0	0	0	0
рукой подать	2	2	2	1	2	2	2	1,8	0,447214
умывать руки	2	1	2	2	2	2	2	2	0
Всё валится из рук	*3*	*2*	2	2	2	2	2	2	0
кривые	3	2	1	1	1	1	2	1,2	0,447214
матери	3	2	0	0	0	0	1	0,2	0,447214
нечистые	3	2	0	0	0	0	2	0,4	0,894427
под	3	2	0	0	1	0	2	0,6	0,894427
под рукой	3	1	1	1	1	1	2	1,2	0,447214
подать руку	3	2	0	0	0	0	1	0,2	0,447214
просить руки	3	1	1	1	1	1	2	1,2	0,447214
целовать	3	2	0	0	1	0	0	0,2	0,447214
вытянуть	1	3	0	0	0	0	0	0	0
девушки	1	3	0	0	0	1	0	0,2	0,447214
кисть	1	4	0	0	0	0	0	0	0

кожа	1	4	0	0	0	0	0	0	0
обе	1	3	0	0	1	0	0	0,2	0,447214
прибрать к	1	3	1	1	1	1	2	1,2	0,447214
ребенка	1	3	0	0	0	0	0	0	0
с пустыми	1	3	1	1	1	1	2	1,2	0,447214
трогать	1	3	0	0	0	0	0	0	0
чешутся	1	4	1	0	1	1	2	1	0,707107
в четыре руки	*2*	*3*	2	2	2	2	2	2	0
не доходят	2	3	2	2	2	2	2	2	0
рука на пульсе	2	4	2	2	1	2	2	1,8	0,447214
сложа руки	2	4	2	2	2	2	2	2	0
чужие руки	2	3	0	0	0	0	2	0,4	0,894427
длинные	3	3	0	0	0	0	0	0	0
из первых	3	3	1	1	1	1	2	1,2	0,447214
набить	3	3	1	1	1	1	2	1,2	0,447214
не покладая рук	3	4	2	2	2	2	2	2	0
поднять	3	4	0	0	0	0	1	0,2	0,447214
развязать	3	3	0	0	0	0	2	0,4	0,894427
схватить за	3	3	0	0	1	0	2	0,6	0,894427
Бриллиантовая	4	2	2	2	2	2	1	1,8	0,447214
дрожат	4	2	0	0	0	0	0	0	0
друга	4	1	0	0	1	0	2	0,6	0,894427
красивые	4	2	0	0	0	0	0	0	0
сломать руку	4	1	0	0	0	0	0	0	0
выпустить из	5	2	0	0	0	0	2	0,4	0,894427

носить на	5	1	1	1	1	1	2	1,2	0,447214
рука руку моет	6	1	1	1	1	1	2	1,2	0,447214
голыми	4	4	1	1	1	1	2	1,2	0,447214
грязные	4	4	0	0	0	0	2	0,4	0,894427
держать в	4	4	0	0	1	1	2	0,8	0,83666
и ногами	4	3	2	2	0	2	2	1,6	0,894427
махнуть	4	4	0	0	0	0	2	0,4	0,894427
мыть руки	4	4	0	0	0	0	0	0	0
пальцы	4	4	0	0	0	0	2	0,4	0,894427
пожать	4	4	1	1	1	1	1	1	0
приложить	4	3	1	1	1	1	2	1,2	0,447214
руки вверх	4	3	1	1	1	1	2	1,2	0,447214
руки за спину	4	3	1	1	1	1	1	1	0
умелые	4	4	1	1	1	1	2	1,2	0,447214
человека	4	4	0	0	0	0	0	0	0
взять в	5	4	1	1	1	1	0	0,8	0,447214
протянуть	5	3	1	1	1	1	0	0,8	0,447214
развести руками	5	4	1	1	1	1	2	1,2	0,447214
рука не поднимается	5	3	2	2	2	2	2	2	0
сильные	5	3	0	0	0	1	0	0,2	0,447214
скрестив	5	3	0	0	0	0	0	0	0
опустить	6	4	1	1	1	1	2	1,2	0,447214
предложить руку и сердце	6	4	2	2	2	2	2	2	0

чистые	6	4	0	0	0	0	2	0,4	0,894427
золотые	7	4	2	2	2	2	2	2	0
Подать (или протянуть) руку (помощи)	8	3	2	2	2	2	2	2	0
левая	9	4	1	1	1	1	0	0,8	0,447214
правая	10	4	2	1	2	2	2	1,8	0,447214
ампутировать	1	0	0	1	0	0	0	0,2	0,447214
ампутация	1	0	0	0	0	0	0	0	0
Ани	1	0	0	0	0	0	0	0	0
балерины	1	0	0	0	0	0	0	0	0
божья	1	0	1	1	1	1	2	1,2	0,447214
верная	1	0	1	1	1	1	1	1	0
влажные	1	0	0	0	1	0	0	0,2	0,447214
властная	1	0	1	1	1	1	2	1,2	0,447214
выбить из	1	0	0	0	0	0	2	0,4	0,894427
вывих	1	0	0	0	0	1	0	0,2	0,447214
вырвать из	1	0	0	0	0	0	2	0,4	0,894427
вытирать	1	0	0	0	1	0	0	0,2	0,447214
Глаза боятся, а руки делают	*1*	*0*	2	2	2	2	2	2	0
грубые	1	0	0	0	0	0	0	0	0
Держать руку чью	1	0	0	0	0	1	0	0,2	0,447214
детские	1	0	0	0	0	0	0	0	0
До ручки дойти	1	0	1	1	1	1	2	1,2	0,447214
дотянуться	1	0	0	1	0	0	0	0,2	0,447214

дружеская	1	0	1	1	1	1	1	1	0
рука закона	1	0	1	1	1	1	2	1,2	0,447214
замерзли	1	0	0	0	0	0	0	0	0
здоровые	1	0	1	1	1	1	0	0,8	0,447214
Искать чьей руки	1	0	2	2	2	2	2	2	0
испачкать	1	0	0	0	1	0	0	0,2	0,447214
карты в руки	1	0	2	2	2	2	2	2	0
корявые	1	0	1	1	1	1	2	1,2	0,447214
костлявые	1	0	0	0	0	0	0	0	0
ледяные	1	0	0	0	0	0	1	0,2	0,447214
лечить	1	0	0	0	0	1	0	0,2	0,447214
линии	1	0	0	0	0	0	2	0,4	0,894427
лишиться	1	0	0	0	0	0	0	0	0
мозоли на	1	0	0	0	0	1	0	0,2	0,447214
мозолистые	1	0	0	0	0	0	0	0	0
мокрые	1	0	0	1	0	0	0	0,2	0,447214
морщинистые	1	0	0	0	0	0	0	0	0
на все руки мастер	1	0	1	1	1	1	2	1,2	0,447214
натянуть на	1	0	0	0	0	0	0	0	0
не хватает рук	1	0	2	2	2	2	2	2	0
небольшие	1	0	0	0	1	0	0	0,2	0,447214
носят	1	0	1	1	1	1	2	1,2	0,447214
от руки	1	0	1	1	1	1	2	1,2	0,447214
отекла	1	0	0	0	0	1	0	0,2	0,447214
отморозить	1	0	0	0	0	0	0	0	0

очумелые ручки	1	0	2	2	2	2	2	2	0
перевязка	1	0	0	0	0	0	0	0	0
по обе руки	1	0	0	0	0	0	1	0,2	0,447214
повязка на	1	0	0	0	0	0	0	0	0
показывать	1	0	0	0	1	0	0	0,2	0,447214
попасть в	1	0	1	1	1	1	2	1,2	0,447214
поранить	1	0	0	1	0	0	0	0,2	0,447214
провести	1	0	0	0	0	0	0	0	0
разогнуть	1	0	0	0	0	0	0	0	0
рана на	1	0	0	0	0	0	0	0	0
распухла	1	0	0	0	0	1	0	0,2	0,447214
С руки кому	1	0	1	1	1	1	2	1,2	0,447214
сжимать в	1	0	0	0	0	0	0	0	0
синяк на	1	0	0	0	0	0	0	0	0
слабые	1	0	0	0	0	0	1	0,2	0,447214
смотреть из-под	1	0	1	1	1	1	0	0,8	0,447214
снимок (рентгеновский)	1	0	0	0	1	0	0	0,2	0,447214
старика	1	0	0	0	0	0	0	0	0
сунуть	1	0	0	0	0	0	0	0	0
сунуть руки в карман	1	0	0	0	0	0	0	0	0
сухие	1	0	0	0	0	1	0	0,2	0,447214
талантливая	1	0	1	1	1	1	1	1	0
травма	1	0	0	0	0	0	0	0	0
удариться	1	0	0	0	0	0	0	0	0
укол в руку	1	0	0	0	0	0	0	0	0
холеные	1	0	0	0	0	1	0	0,2	0,447214

штангиста	1	0	0	1	0	0	0	0,2	0,447214
щедрой рукой	1	0	1	1	1	1	2	1,2	0,447214
больные	2	0	1	1	1	1	0	0,8	0,447214
вести под	2	0	1	1	1	1	1	1	0
взяться за	2	0	1	1	1	1	1	1	0
вывихнуть	2	0	0	0	0	0	0	0	0
выронить из	2	0	0	0	0	0	0	0	0
гибкая	2	0	0	0	0	0	0	0	0
горячая	2	0	0	0	0	0	2	0,4	0,894427
Дать руку на отсечение	2	0	2	1	2	2	2	1,8	0,447214
железные	2	0	1	1	1	1	2	1,2	0,447214
жесткие	2	0	0	0	0	1	2	0,6	0,894427
загорелые	2	0	0	0	0	0	0	0	0
загребущая	2	0	2	2	2	2	2	2	0
зажать в	2	0	0	0	0	0	0	0	0
Запустить руку во что .	2	0	0	0	0	0	2	0,4	0,894427
Иметь (сильную) руку где	2	0	2	2	2	2	0	1,6	0,894427
как без рук	*2*	0	2	1	2	1	2	1,6	0,547723
короткие	2	0	1	1	1	1	0	0,8	0,447214
ласковая	2	0	0	0	0	0		0	0
Лизать руки	2	0	0	0	0	0	2	0,4	0,894427
Ломать руки	2	0	0	0	0	0	2	0,4	0,894427
Марать руки об	2	0	1	1	1	1	2	1,2	0,447214
мужские	2	0	0	0	0	1	2	0,6	0,894427

мягкие	2	0	0	0	0	0	0	0	0
наложить гипс	2	0	0	0	1	0	0	0,2	0,447214
ноет	2	0	0	0	0	0	0	0	0
Обагрить руки кровью	2	0	1	1	1	1	2	1,2	0,447214
Обломать руки	2	0	1	1	1	1	2	1,2	0,447214
отбиться от	2	0	2	2	2	2	2	2	0
подвернуться под	2	0	0	0	0	0	2	0,4	0,894427
Поднять руку на	2	0	1	1	1	1	2	1,2	0,447214
Подписатьс я обеими руками под чем-л.	*2*	0	2	2	2	2	2	2	0
полные	2	0	0	0	0	0	2	0,4	0,894427
положить на	2	0	0	1	0	0	0	0,2	0,447214
порезать	2	0	0	0	0	0	0	0	0
прикоснуться	2	0	0	0	0	0	0	0	0
раненая	2	0	0	0	0	0	2	0,4	0,894427
Руки коротки!	*2*	0	2	2	2	2	2	2	0
связать	2	0	0	0	1	0	2	0,6	0,894427
сделать своими руками	2	0	1	1	1	1	2	1,2	0,447214
товарища	2	0	1	1	1	1	2	1,2	0,447214
Тянуть чью руку	2	0	0	0	0	0	0	0	0
ударить по	2	0	1	1	1	1	2	1,2	0,447214
Укоротить руки кому	2	0	2	2	2	2	2	2	0
художника	2	0	1	1	1	1	2	1,2	0,447214

Что-л. само в руки идёт	*2*	0	2	2	2	2	2	2	0
В рукахчьих или у кого	3	0	2	2	2	2	2	2	0
из рук вон плохо	3	0	2	2	2	2	2	2	0
изящные	3	0	0	0	0	1	0	0,2	0,447214
крепкие	3	0	0	0	1	0	1	0,4	0,547723
На́ руку кому	3	0	2	2	2	2	2	2	0
Нагреть руки	3	0	1	1	1	1	2	1,2	0,447214
надеть на	3	0	0	0	0	0	0	0	0
Наложить руки на себя.	3	0	2	2	2	2	2	2	0
Не рука кому (устар.)	3	0	2	2	2	2	2	2	0
нежные	3	0	0	0	0	0	0	0	0
нести в	3	0	0	1	0	0	0	0,2	0,447214
перелом	3	0	0	0	0	0	0	0	0
писателя	3	0	1	1	1	1	2	1,2	0,447214
подобрать что-л по руке	3	0	0	0	0	0	1	0,2	0,447214
правосудия	3	0	0	0	1	0	2	0,6	0,894427
Руки по швам	3	0	1	1	1	1	2	1,2	0,447214
Руки прочь от	3	0	1	1	1	1	2	1,2	0,447214
сойти с рук	3	0	2	2	2	2	2	2	0
Сон в руку	3	0	2	2	2	2	2	2	0
худые	3	0	1	1	1	1	2	1,2	0,447214
Давать волю рукам	3	0	2	2	2	2	2	2	0

Дать по рукам кому-л	*3*	*0*	2	2	2	2	2	2	0
белые	4	0	0	0	0	0	2	0,4	0,894427
болят	4	0	0	1	0	0	0	0,2	0,447214
вести за	4	0	0	0	0	0	0	0	0
взять на	4	0	0	0	1	0	0	0,2	0,447214
Греть руки	4	0	0	0	0	0	2	0,4	0,894427
держать на	4	0	0	0	0	0	2	0,4	0,894427
идти под руку	4	0	1	1	1	1	2	1,2	0,447214
Из рук в руки	4	0	1	1	1	1	2	1,2	0,447214
На руках иметь	4	0	1	1	1	1	2	1,2	0,447214
писать	4	0	0	0	1	0	2	0,6	0,894427
По рукам!	4	0	2	2	2	2	2	2	0
погладить	4	0	0	0	0	0	0	0	0
рабочего	4	0	1	1	1	1	2	1,2	0,447214
рука не дрогнет	4	0	2	2	2	2	2	2	0
теплые	4	0	0	0	0	0	0	0	0
тонкие	4	0	0	1	0	1	0	0,4	0,547723
ударить	4	0	0	0	0	0	2	0,4	0,894427
холодные	4	0	0	0	0	0	0	0	0
Чужими руками жар загребать	4	0	2	2	2	2	2	2	0
большие	5	0	0	0	0	0	0	0	0
взять за	5	0	0	0	0	0	0	0	0
волосатые	5	0	0	0	0	0	2	0,4	0,894427
знать чью-л руку	5	0	1	1	1	1	2	1,2	0,447214

легкая рука	5	0	1	1	1	1	2	1,2	0,447214
маленькие	5	0	0	1	0	1	0	0,4	0,547723
твердая	5	0	1	1	1	1	2	1,2	0,447214
тяжелая рука	5	0	1	1	1	1	2	1,2	0,447214
женские	7	0	0	0	0	0	2	0,4	0,894427
Рука об руку	8	0	1	1	1	1	2	1,2	0,447214

Приложение 11. Экспертная оценка коллокаций, содержащих слово *белый*

	Количество словарей	Количество мер	Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	среднее арифметическое	стандартное отклонение
город	0	2	0	0	0	0	0	0	0
потолок	0	2	0	0	0	0	0	0	0
сахар	0	2	0	0	0	0	0	0	0
зарплата	0	2	2	2	2	2	2	2	0
пляж	0	2	2	2	2	2	1	1,8	0,447214
береза	0	2	0	0	0	0	0	0	0
лимузин	0	2	0	0	0	0	0	0	0
бумага	0	2	0	1	0	0	0	0,2	0,447214
чай	0	2	0	0	0	0	2	0,4	0,894427
дача	0	2	0	0	0	0	0	0	0
кролик	0	2	0	0	1	0	0	0,2	0,447214
орел	0	2	0	0	0	0	0	0	0
море	0	4	2	2	2	2	2	2	0
цветок	0	3	0	0	1	0	0	0,2	0,447214
камень	0	3	0	1	0	0	0	0,2	0,447214
платье	0	3	0	0	0	0	0	0	0
песок	0	4	0	0	0	0	0	0	0
золото	0	4	1	1	1	1	2	1,2	0,447214
глина	0	4	1	1	1	1	2	1,2	0,447214
конь	0	4	0	0	0	0	0	0	0
налет	0	4	0	0	0	0	0	0	0
краска	0	4	0	0	0	0	0	0	0
роза	0	4	0	1	0	0	0	0,2	0,447214
рубашка	0	4	0	0	0	0	0	0	0
полоса	0	4	0	0	0	0	0	0	0
список	0	3	2	2	2	2	2	2	0
одежда	0	3	0	0	0	0	0	0	0
стена	0	3	0	0	1	0	0	0,2	0,447214
мрамор	0	3	0	0	0	1	0	0,2	0,447214
шоколад	0	3	2	2	2	2	2	2	0
тигр	0	3	2	2	2	2	2	2	0
зал	0	3	2	2	2	2	2	2	0
порошок	0	3	0	1	0	0	0	0,2	0,447214

экран	0	3	1	1	1	1	2	1,2	0,447214
голубь	0	3	0	0	0	0	0	0	0
зависть	0	3	2	2	2	2	2	2	0
акула	0	3	0	0	0	1	2	0,6	0,894427
парус	0	3	2	2	2	2	2	2	0
мясо	3	2	2	2	2	2	2	2	0
уголь	2	2	2	2	2	2	2	2	0
ангел	3	1	0	0	1	2	2	1	1
ветер	1	2	2	2	2	2	2	2	0
воротничок	1	4	2	2	2	2	2	2	0
Довести до белого каления	3	3	2	2	2	2	2	2	0
духовенство	2	2	2	2	2	2	2	2	0
олимпиада	1	1	2	2	2	2	2	2	0
ворона	3	3	2	2	2	2	2	2	0
пятна	3	3	2	2	0	0	2	1,2	1,095445
дом	3	3	2	2	2	2	2	2	0
кот	3	3	0	0	0	0	0	0	0
магия	2	3	2	2	2	2	2	2	0
облако	2	3	0	0	0	0	0	0	0
фон	1	4	0	0	0	0	0	0	0
халат	2	4	2	2	2	2	1	1,8	0,447214
цвет	3	4	0	0	1	0	0	0,2	0,447214
шум	1	4	2	2	2	2	2	2	0
стихи	4	1	2	2	2	2	2	2	0
гвардия	6	1	2	2	2	2	2	2	0
горячка	5	2	2	2	2	2	2	2	0
клык	4	2	2	2	2	2	2	2	0
пух	4	1	0	2	0	0	0	0,4	0,894427
свет	6	2	2	2	2	2	2	2	0
танец	4	1	2	2	2	2	1	1,8	0,447214
человек	6	2	2	2	2	2	2	2	0
вино	4	3	2	2	2	2	2	2	0
ночи	4	3	2	2	2	2	2	2	0

хлеб	9	3	2	2	2	2	2	2	0
гриб	6	4	2	2	2	2	2	2	0
лебедь	5	4	2	2	2	2	1	1,8	0,447214
лист	5	3	0	0	0	1	2	0,6	0,894427
медведь	5	3	2	2	2	2	2	2	0
снег	6	3	2	2	2	2	2	2	0
Среди бела дня	8	4	2	2	2	2	2	2	0
флаг	4	3	2	2	2	2	2	2	0
железо	1	0	0	0	0	1	2	0,6	0,894427
изба	2	0	2	2	2	2	2	2	0
кость	2	0	2	2	2	2	2	2	0
мухи	1	0	2	2	2	2	2	2	0
Белыми нитками шито	1	0	2	2	2	2	2	2	0
Дела как сажа бела	1	0	2	2	2	2	2	2	0
зима	2	0	2	2	2	2	1	1,8	0,447214
и пушистый	3	0	2	2	2	2	2	2	0
как снег	2	0	2	2	2	2	0	1,6	0,894427
Называть белое черным	1	0	2	2	2	2	2	2	0
пепел	2	0	2	2	2	2	2	2	0
Сказка про белого бычка	3	0	2	2	2	2	2	2	0
Черным по белому	1	0	2	2	2	2	2	2	0
заяц	4	0	1	1	1	1	1	1	0
билет	4	0	2	2	2	2	2	2	0

Приложение 12. Экспертная оценка коллокаций, содержащих слово *скакать*

			Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	среднее арифметическое	стандартное отклонение
галопом	0	2	2	1	1	2	2	1,5	0,547723
беззаботно	0	4	0	0	0	0	0	0	0
бешено	0	4	1	1	1	1	0	1	0,447214
бодро	0	4	1	1	1	1	0	1	0,447214
вверх-вниз	0	4	0	0	0	0	0	0	0
весело	0	4	0	0	0	0	0	0	0
впереди	0	3	0	0	0	0	0	0	0
вприпрыжку	0	4	1	1	1	1	1	1	0
всадник	0	4	2	2	2	2	1	2	0,447214
доллар	0	4	2	2	2	2	2	2	0
зайчик	0	4	0	0	0	0	0	0	0
кавалерия	0	4	1	1	1	1	0	1	0,447214
кенгуру	0	3	0	0	0	0	0	0	0
ковбой	0	4	0	0	0	0	0	0	0
козел	0	4	0	0	0	0	0	0	0
конь	0	4	1	1	1	1	0	1	0,447214
кузнечик	0	4	0	0	0	0	0	0	0
лихо	0	4	0	0	0	0	0	0	0
ловко	0	4	0	0	0	0	0	0	0
мысль	0	4	2	2	2	2	2	2	0
навстречу	0	4	0	0	0	0	0	0	0
напряжение	0	4	2	2	2	2	2	2	0
настроение	0	4	2	2	2	2	2	2	0
неуклюже	0	4	0	0	0	0	0	0	0
обезьяна	0	4	0	0	0	0	0	0	0
опрометью	0	3	0	0	0	0	0	0	0
перестать	0	3	0	0	0	0	0	0	0
посещаемость	0	3	2	2	2	2	2	2	0
постоянно	0	3	0	0	0	0	0	0	0
проворно	0	4	0	0	0	0	0	0	0
прочь	0	3	0	0	0	0	0	0	0
псина	0	3	0	0	0	0	0	0	0

птица	0	3	0	0	0	0	0	0	0
пульс	0	4	2	2	2	2	2	2	0
радостно	0	4	0	0	0	0	0	0	0
резво	0	4	0	0	0	0	0	0	0
рядом	0	3	0	0	0	0	0	0	0
тень	0	4	1	1	1	1	0	1	0,447214
тройка	0	4	2	2	2	2	2	2	0
трусцой	0	4	2	2	2	2	2	2	0
туда-сюда	0	4	0	0	0	0	0	0	0
воробей	1	2	0	0	0	0	0	0	0
давление	2	2	2	2	2	2	2	2	0
белка	1	3	0	0	0	0	0	0	0
быстро	2	3	0	0	0	0	0	0	0
заяц	3	4	0	0	0	0	0	0	0
лягушка	1	3	0	0	0	0	0	0	0
температура	3	3	2	2	2	2	2	2	0
лошадь	6	3	0	0	0	0	0	0	0
блоха	1	0	1	1	1	1	0	1	0,447214
верхом	2	0	2	2	2	2	2	2	0
во весь дух или опор	3	0	2	2	2	2	2	2	0
вокруг	0	3	0	0	0	0	0	0	0
девочка	3	0	0	0	0	0	0	0	0
жеребенок	1	0	0	0	0	0	0	0	0
мяч	1	0	2	2	2	2	0	2	0,894427
на одной ноге.	1	0	2	2	2	2	0	2	0,894427
цены	1	0	2	2	2	2	2	2	0
через веревочку	3	0	2	2	2	2	1	2	0,447214
через огонь	1	0	2	2	2	2	0	2	0,894427
бегать и	1	0	0	0	0	0	0	0	0

Приложение 13. Экспертная оценка коллокаций, содержащих слово *семь*

	Количество словарей	Количество мер	Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	среднее арифметическое	стандартное отклонение
седьмое небо	0	2	2	2	2	2	2	2	0
седьмой десяток	0	1	2	1	1	1	0	1	0,707107
седьмой сын	0	1	0	1	2	1	0	0,8	0,83666
семь вечеров	0	2	1	2	0	1	2	1,2	0,83666
семь мудрецов	0	1	2	2	2	2	2	2	0
семь нот	0	2	0	1	0	0	1	0,4	0,547723
семь холмов	0	2	0	1	0	0	0	0,2	0,447214
у семи нянек дитя без глазу	0	2	2	2	1	2	2	1,8	0,447214
семь ангелов	0	4	0	0	0	0	2	0,4	0,894427
семь богатырей	0	3	0	1	1	1	2	1	0,707107
семь гномов	0	4	1	1	1	1	1	1	0
семь грехов	0	3	2	1	1	0	2	1,2	0,83666
семь цветов радуги	0	3	0	0	0	0	1	0,2	0,447214
семь чакр	0	4	0	1	1	1	1	0,8	0,447214
семь чудес света	0	3	2	2	2	2	2	2	0
Семи пядей во лбу	2	2	2	2	2	2	2	2	0
Семь потов сошло с кого	1	1	2	2	2	2	2	2	0

семь раз отмерь	3	3	2	2	2	2	2	2	0
дней	1	4	0	1	1	0	0	0,4	0,547723
За семь верст киселя хлебать	1	0	2	2	2	2	2	2	0
За семью замками	2	0	2	2	2	2	2	2	0
Книга за семью печатями	1	0	2	2	2	2	2	2	0
С. бед - один ответ	3	0	2	2	2	2	2	2	0

Приложение 14. Экспертная оценка коллокаций, содержащих слово *свой*

	Количество словарей	Количество мер	Эксперт 0	Эксперт 1	Эксперт 2	Эксперт 3	Эксперт 4	среднее арифметическое	стандартное отклонение
бизнес	0	2	1	1	1	1	1	1	0
свое дело	0	4	1	1	1	1	1	1	0
свое мнение	0	4	1	1	1	1	0	0,8	0,447214
своя квартира	0	4	1	1	1	2	1	1,2	0,447214
страна	0	4	1	1	1	1	1	1	0
вкус	0	4	1	1	1	1	1	1	0
усмотрение	0	4	1	1	1	2	1	1,2	0,447214
свое отношение	0	4	1	1	1	2	0	1	0,707107
решение	0	4	1	1	1	1	0	0,8	0,447214
знать свой организм	0	4	0	1	0	0	0	0,2	0,447214
свое состояние	0	4	0	1	0	0	0	0,2	0,447214
свой компьютер	0	4	0	1	0	0	0	0,2	0,447214
свой участок	0	4	0	1	0	0	0	0,2	0,447214
свое лицо	0	4	0	1	0	0	0	0,2	0,447214
подтвердить свою репутацию	0	4	0	1	0	0	0	0,2	0,447214
свой доход	0	4	0	1	0	0	0	0,2	0,447214
свой адрес	0	4	0	1	0	0	0	0,2	0,447214
свой заказ	0	4	0	1	0	0	0	0,2	0,447214
своя голова на плечах	0	4	2	1	2	2	2	1,8	0,447214
оставить свой отзыв	0	4	0	1	0	0	0	0,2	0,447214
своя фантазия	0	4	0	1	0	0	0	0,2	0,447214

написать свое пожелание	0	4	0	1	0	0	1	0,4	0,547723
дать свое согласие	0	4	2	1	2	2	2	1,8	0,447214
указать свой телефон	0	4	0	1	0	0	0	0,2	0,447214
свое воображение	0	4	0	1	0	0	0	0,2	0,447214
(Рассказать) своими словами	2	1	2	1	2	2	2	1,8	0,447214
Быть не в своей тарелке	1	1	2	1	2	2	2	1,8	0,447214
В свою очередь	1	2	2	1	2	2	2	1,8	0,447214
Своя ноша не тянет	2	2	2	1	2	2	2	1,8	0,447214
свой народ	3	4	1	1	1	2	1	1,2	0,447214
Сделать своими руками	4	3	2	1	2	2	2	1,8	0,447214
Своя голова на плечах	1	0	2	1	2	2	2	1,8	0,447214
(не)В своем уме	2	0	2	1	2	2	2	1,8	0,447214
(не)На своем месте	1	0	2	1	2	2	2	1,8	0,447214
Брать (взять) свое	3	0	2	1	2	2	2	1,8	0,447214
В свое время	2	0	2	1	2	2	2	1,8	0,447214
В свое удовольствие	1	0	2	1	2	2	2	1,8	0,447214
В своем роде	1	0	2	1	2	2	2	1,8	0,447214
выбор	2	0	0	1	0	0	0	0,2	0,447214

Жить своим умом	2	0	2	1	2	2	2	1,8	0,447214
Знать свое место	2	0	2	1	2	2	2	1,8	0,447214
Идти своей дорогой	1	0	2	1	2	2	2	1,8	0,447214
Идти своим ходом	1	0	2	1	2	2	2	1,8	0,447214
Идти своим чередом	1	0	2	1	2	2	2	1,8	0,447214
Мастер своего дела	1	0	2	1	2	2	2	1,8	0,447214
На свой страх (и риск)	1	0	2	1	2	2	2	1,8	0,447214
На своих двоих	1	0	2	1	2	2	2	1,8	0,447214
На свою голову	1	0	2	1	2	2	2	1,8	0,447214
Называть вещи своими именами	2	0	2	1	2	2	2	1,8	0,447214
Не в свои сани сесть	1	0	2	1	2	2	2	1,8	0,447214
Не верить своим глазам	1	0	2	1	2	2	2	1,8	0,447214
Не своим голосом	2	0	2	1	2	2	2	1,8	0,447214
нести свой крест	1	0	2	1	2	2	2	1,8	0,447214
Остаться при своих	1	0	2	1	2	2	2	1,8	0,447214
отпуск за свой счет	1	0	0	1	0	0	2	0,6	0,894427
по-своему	2	0	0	1	0	0	2	0,6	0,894427
Поставить на свое место	2	0	0	1	0	0	2	0,6	0,894427

принцип	1	0	0	1	0	0	0	0,2	0,447214
Принять на свой счет	2	0	2	1	2	2	2	1,8	0,447214
Сам не свой	1	0	2	1	2	2	2	1,8	0,447214
Свое на уме	1	0	2	1	2	2	2	1,8	0,447214
Свое я	1	0	2	1	2	2	1	1,6	0,547723
Своего поля ягода	3	0	2	1	2	2	2	1,8	0,447214
Своего рода	1	0	2	1	2	2	2	1,8	0,447214
Своим порядком	1	0	2	1	2	2	2	1,8	0,447214
Своих не узнаешь	3	0	2	1	2	2	2	1,8	0,447214
Своя рубаха ближе к телу.	1	0	2	1	2	2	2	1,8	0,447214
Своя рука владыка	1	0	2	1	2	2	2	1,8	0,447214
Сказать свое слово	1	0	1	1	1	1	1	1	0
Стоять на своих ногах	1	0	1	1	1	1	2	1,2	0,447214
Умереть (не) своей смертью	1	0	2	1	2	2	2	1,8	0,447214
дом	4	0	1	1	1	1	1	1	0
парень	5	0	2	2	2	2	2	2	0
Свой брат	4	0	2	1	2	2	2	1,8	0,447214
Свой в доску	5	0	2	1	2	2	2	1,8	0,447214

Приложение 15. Коэффициент корреляции Спирмена между различными мерами для слова *сердце*

Коэффициент корреляции Спирмена	Коэффициэнт корреляции меры с основным рангом	T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,4589	X	-0,0709	0,4760546	0,341431	0,1117652	0,1731556	0,1372623
MI	-0,0482	-0,0709	X	0,4344339	0,5469534	0,8208915	0,5712681	0,2671196
MI3	0,6018	0,4761	0,4344339	X	0,7760243	0,6062761	0,5280576	0,2125833
log-likelihood	0,7379	0,3414	0,5469534	0,7760243	X	0,7459267	0,4080657	0,1368813
min. sensitivity	0,2930	0,1118	0,8208915	0,6062761	0,7459267	X	0,5348134	0,2207986
log-Dice	0,0608	0,1732	0,5712681	0,5280576	0,4080657	0,5348134	X	0,2207986
MI.log-_f	-0,0421	0,1373	0,2671196	0,2125833	0,1368813	0,2207986	0,3842942	X

Приложение 16.Коэффициент корреляции Спирмена между различными мерами для слова *вода*

Коэффициент корреляции Спирмена	Коэффициэнт корреляции меры с основным рангом	T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,9999	X	-0,5554	0,874153	0,979242	0,955725	0,999797	0,501194
MI	-0,5620	-0,5554	X	-0,15201	-0,4082	-0,38475	-0,54672	0,321911
MI3	0,8703	0,8742	-0,15201	X	0,950133	0,94105	0,877483	0,821187
log-likelihood	0,9775	0,9792	-0,4082	0,950133	X	0,975809	0,980731	0,645801
min. sensitivity	0,9544	0,9557	-0,38475	0,94105	0,975809	X	0,956088	0,632585
log-Dice	0,9995	0,9998	-0,54672	0,877483	0,980731	0,956088	X	0,511093
MI.log-_f	0,4932	0,5012	0,321911	0,821187	0,645801	0,632585	0,511093	X

Приложение 17.Коэффициент корреляции Спирмена между различными мерами для слова *рука*

Коэффициент корреляции Спирмена	Коэффициэнт корреляции меры с основным рангом	T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,9424	X	0,0867	0,50526	0,651345	0,346511	0,266253	0,29538
MI	0,1455	0,0867	X	0,141543	0,338573	0,775252	0,285882	0,903406
MI3	0,5363	0,5053	0,141543	X	0,778135	0,177671	0,764865	0,298079
log-likelihood	0,7089	0,6513	0,338573	0,778135	X	0,422769	0,582932	0,523636
min. sensitivity	0,4055	0,3465	0,775252	0,177671	0,422769	X	0,339341	0,812443
log-Dice	0,3064	0,2663	0,285882	0,764865	0,582932	0,339341	X	0,812443
MI.log-_f	0,3693	0,2954	0,903406	0,298079	0,523636	0,812443	0,371697	X

Приложение 18.Коэффициент корреляции Спирмена между различными мерами для слова *белый*

Коэффициент корреляции Спирмена	Коэффициэнт корреляции меры с основным рангом	T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,7017	X	0,5520	0,738832	0,695228	0,423133	0,634428	0,38312
MI	0,0828	0,5520	X	0,706832	0,674183	0,427334	0,717464	0,646232
MI3	0,5014	0,7388	0,706832	X	0,838692	0,427906	0,749369	0,584984
log-likelihood	0,4083	0,6952	0,674183	0,838692	X	0,514905	0,759191	0,600732
min. sensitivity	0,1007	0,4231	0,427334	0,427906	0,514905	X	0,714177	0,743815
log-Dice	0,2343	0,6344	0,717464	0,749369	0,759191	0,714177	X	0,743815
MI.log-_f	0,0178	0,3831	0,646232	0,584984	0,600732	0,743815	0,778864	X

Приложение 19.Коэффициент корреляции Спирмена между различными мерами для слова *скакать*

Коэффициент корреляции Спирмена	Коэффициэнт корреляции меры с основным рангом	T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,8006	X	0,0944	0,01943	0,691161	0,430243	0,432955	0,505095
MI	-0,1216	0,0944	X	-0,18224	0,101878	0,385095	0,384734	0,098326
MI3	0,1567	0,0194	-0,18224	X	-0,02838	-0,17949	-0,17977	-0,12698
log-likelihood	0,3521	0,6912	0,101878	-0,02838	X	0,637432	0,638836	0,828755
min. sensitivity	-0,0777	0,4302	0,385095	-0,17949	0,637432	X	0,999832	0,713651
log-Dice	-0,0742	0,4330	0,384734	-0,17977	0,638836	0,999832	X	0,713651
MI.log-_f	0,1489	0,5051	0,098326	-0,12698	0,828755	0,713651	0,714599	X

Приложение 20.Коэффициент корреляции Спирмена между различными мерами для слова *семь*

Коэффициент корреляции Спирмена	Коэффициэнт корреляции меры с основным рангом	T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,9329	X	0,0110	0,830827	0,829626	0,254219	0,130025	0,011031
MI	-0,0485	0,0110	X	0,229298	0,220011	0,367268	0,676743	1
MI3	0,7111	0,8308	0,229298	X	0,983438	0,604061	0,402332	0,229298
log-likelihood	0,6872	0,8296	0,220011	0,983438	X	0,611784	0,426588	0,220011
min. sensitivity	0,0676	0,2542	0,367268	0,604061	0,611784	X	0,538635	0,367268
log-Dice	-0,0455	0,1300	0,676743	0,402332	0,426588	0,538635	X	0,367268
MI.log-_f	-0,0485	0,0110	1	0,229298	0,220011	0,367268	0,676743	X

Приложение 21.Коэффициент корреляции Спирмена между различными мерами для слова *свой*

Коэффициент корреляции Спирмена	Коэффициэнт корреляции меры с основным рангом	T-score	MI	MI3	log-likelihood	min. sensitivity	log-Dice	MI.log-_f
T-score	0,7164	X	0,2952	0,084493	0,378185	0,778714	0,514372	0,27245
MI	-0,0538	0,2952	X	0,38582	0,586988	0,043882	0,257099	0,856146
MI3	0,0318	0,0845	0,38582	X	0,570406	0,087336	0,460264	0,550608
log-likelihood	0,2310	0,3782	0,586988	0,570406	X	0,249829	0,633033	0,720082
min. sensitivity	0,8907	0,7787	0,043882	0,087336	0,249829	X	0,559991	0,096692
log-Dice	0,4415	0,5144	0,257099	0,460264	0,633033	0,559991	X	0,096692
MI.log-_f	-0,0097	0,2724	0,856146	0,550608	0,720082	0,096692	0,335548	X