**Правительство Российской Федерации**

**Федеральное государственное автономное образовательное учреждение высшего образования**

**Национальный исследовательский университет**

**«Высшая школа экономики»**

**Факультет гуманитарных наук**

**Образовательная программа**

**«Цифровые методы в гуманитарных науках»**

**КУРСОВАЯ РАБОТА**

На тему: Изучение эмоциональной культуры на материале дневников

проекта «Прожито».

*Название темы на английском: Emotional culture in “Prozhito” diaries.*

Студентка 1 курса группы МЦМГН201

Воробьева Виктория Александровна\_\_\_\_\_\_\_\_\_\_\_\_

(Ф.И.О.)

Научный руководитель

Бонч-Осмоловская Анастасия Александровна\_\_\_\_\_\_\_\_\_\_\_\_

(Ф.И.О)

кандидат филологических наук\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(должность, звание)

Москва, 2021 г.

Оглавление

[1. Введение 3](#_Toc75089062)

[2. Обзор литературы 6](#_Toc75089063)

[3. Обработка материала 8](#_Toc75089064)

[3.1 Описание корпуса 9](#_Toc75089065)

[3.2 Предобработка материала 13](#_Toc75089066)

[4. Анализ корпуса 14](#_Toc75089067)

[5. Заключение 28](#_Toc75089068)

[Литература 30](#_Toc75089069)

[Приложение 31](#_Toc75089070)

# Введение

Еще в конце XX века автобиографический жанр могли назвать «маргинальным» из-за его пограничного положения между литературой и нелитературой. Исследователи ограничивались классическими большими мемуарами и известными дневниками, оправдывая свой интерес историческим содержанием автобиографических произведений. В современных исследованиях такие оговорки встречаются редко, потому что ученых больше интересуют дневники и мемуары «обычных» людей. Вопрос «*как* написано» постепенно заменяет вопрос «*что* написано».

Автобиографические документы все чаще попадают в поле внимания исследователей. Отметим, что в издательстве «Новое литературное обозрение» только в этом году вышло несколько книг, посвященных автобиографическому жанру. Например, работы Ирины Паперно «Советская эпоха в мемуарах, дневниках, снах» и Йохена Хелльбека «Революция от первого лица: Дневники сталинской эпохи».

Заметим, что исследователей и издателей особенно привлекают эго-документы советской эпохи. На этот факт обращает внимание Ирина Паперно в статье «What Can Be Done with Diaries?» (Paperno 2004: 567). Паперно замечает, что дневники советского периода часто становятся «лабораторией советского человека» — пространством, в котором автор дневника сравнивает себя с идеалом советского гражданина и пытается выстроить свою жизнь по этому образцу. Благодаря такому соединению индивидуального и общественного советские дневники становятся интересным объектом для исследования.

Интерес к дневникам и автобиографиям появляется не только у исследователей, но и у рядовых читателей. Этим отчасти и объясняется популярность таких проектов как «Прожито». Базу «Прожито» может использовать исследователь в качестве источника материала, пользователь — для чтения или сохранения своей семейной истории.

Но несмотря на возрастающий интерес к эго-документам, русскоязычных работ по теории автобиографического повествования не так много. Из наиболее крупных следует выделить работы Л. Я. Гинзбург (Гинзбург 2009), Ю. П. Зарецкого (Зарецкий 2011) и Ирины Паперно (Паперно 2018).

Еще одной заметной работой, посвященной автобиографическому жанру, стала книга А. Л. Зорина «Появление героя. Из истории русской эмоциональной культуры конца XVIII – начала XIX вв.» (Зорин 2016). Зорин показывает, как на эмоциях русского дворянина отразилось влияние важных общественных и культурных институтов того времени: масонского общества, театра и литературы. В центре исследования – дневник русского «романтика до романтизма» Андрея Ивановича Тургенева. Индивидуальное переживание, выраженное в дневнике, изучается в рамках коллективной тенденции. Такая научная парадигма важна и для настоящего исследования.

**Предметом** исследования являются записи из базы личных дневников «Прожито». В эту базу загружены дневники разных временных периодов: от XVIII века до наших дней. Лучше всего представлен XX век, к тому же авторский состав этого периода более демократичен. На первом этапе работы были использованы все записи базы. Среди них по ключевым словам были отобраны записи любовного характера.

В основе исследования лежит следующая **гипотеза**: описывая любовное переживание, авторы используют определенные паттерны. Ими могут быть готовые конструкции, устойчивые сочетания, средства выразительности. Такие паттерны и схемы описания анализировались в качестве **объектов** исследования.

В процессе работы попробуем проанализировать:

1. какие устойчивые сочетания и средства выразительности используют авторы дневников при описании эмоций;
2. изменяются ли паттерны с течением времени;
3. используют ли авторы дневников отсылки к литературным произведениям или героям при описании чувств.

**Цель** исследования – выявить и сравнить паттерны и тенденции, которые встречаются в записях любовного характера в разные временные периоды.

В соответствии с этой целью были определены **задачи**:

1. составить корпус из записей любовного характера и разделить записи на временные периоды;
2. с помощью тематического моделирования выделить темы для каждого периода и сравнить;
3. проанализировать языковые паттерны и повторяющиеся конструкции в записях;
4. сделать выводы о тенденциях, наблюдаемых в записях любовного характера.

**Новизна** исследования состоит в том, что на русском языке не предпринималось исследований, подобных работе Зорина, на большом материале. Проанализировать культурную обусловленность эмоций на материале тысячи дневников гораздо сложнее. Однако для цели настоящего исследования — сравнения паттернов разных периодов — привлечь большой объем материала кажется обоснованным.

Работа состоит из введения, трех глав, заключения и списка литературы. Первая глава посвящена обзору теоретической литературы и методологии исследования. Во второй главе описан материал исследования и процесс обработки текстов. В третьей главе представлен анализ дневниковых записей и сравнение паттернов разных временных периодов. В заключении сделаны выводы о перспективах настоящего исследования.

# Обзор литературы

Как было отмечено во введении, в анализе автобиографических документов долгое время преобладал исторический подход. Дневники воспринимали как источник бытовых подробностей разных исторических эпох. Только женские дневники могли изучать с другой точки зрения – через способы конструирования своего «я» под взглядом мужского «другого» (Савкина 2007).

С другой стороны, еще в XVII-XVIII вв. возрастает популярность интимного дневника. В отличие от мемуаров такой дневник содержал больше самоанализа и саморефлексии. Разница этих двух эго-документов в том, что дневник обращен в будущее, а не в прошлое, как мемуары (Lejeune 2009: 191). Даже интимный дневник, не предназначенный для чтения, ориентирован на подразумеваемого читателя, например, на себя в будущем. Поэтому на страницах дневника чаще разбирают свои эмоции и поступки. С этой точки зрения, дневник представляется подходящим материалом для изучения эмоциональной культуры.

Подобный подход к автобиографическим документам применил А. Л. Зорин в уже упоминавшейся книге «Появление героя». В изучении эмоций Зорин опирается на методологию голландских психологов Николаса Фрайды и Батья Месквито (Frijda, Mesquita 1994). Согласно их теории, эмоциональный процесс можно разделить на несколько частей. Выделение события, которое вызывает эмоцию, называется *кодированием*. Отношение к этому событию выражается с помощью *оценки* — это может быть страх, радость или гнев. Как описывают Фрайда и Месквито, кодирование и оценка обусловлены культурными нормами, правилами и табу. Это представление об эмоциях является важной предпосылкой для гипотезы настоящего исследования.

С точки зрения методологии, другая важная работа для анализа — «История эмоций» Яна Плампера (Plamper 2015). Это междисциплинарное исследование включает как психологический, так и культурный аспект эмоций. В этой работе также поднимается вопрос о том, насколько эмоции универсальны и культурно обусловлены.

Таким образом, в последние годы появилось несколько работ, в которых эмоции рассматриваются как проявление коллективного, а не индивидуального. Однако в большинстве исследований культурное конструирование эмоции анализируется на примере текстов людей высших сословий, например русских дворян. К тому же паттерны поведения не изучались на большом материале и не сравнивались между собой в зависимости от временного периода. В этом заключается новизна настоящего исследования, но также с этим связан ряд проблем, о которых будет рассказано во второй главе.

Несмотря на то, что исторический подход к дневникам и мемуарам еще остается актуальным, количественные методы не раз применяли к автобиографическим документам. В работе «Topic Modeling Martha Ballard’s Diary» (Blevins 2010) исследователь Кэмерон Блевинс с помощью количественных методов выделяет темы и закономерности, которые появляются в дневниках американки Марты Бэллард, жившей в Массачусетсе в XVIII веке.

Блевинс замечает, что тематическое моделирование оказалось продуктивным для анализа дневника Бэллард. В этом дневнике затрагивается ограниченный ряд тем, например, «церковь», «садоводство» или «болезнь». Благодаря количественным методам удалось выделить незаметные на первый взгляд паттерны. Закономерности проявлялись в том, как использовались слова, связанные с определенной темой.

В другом исследовании «Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis» (Albalawi et. al. 2020) ученые сравнивают разные методы тематического моделирования, применяя их к текстам из микроблогов и новостных страниц. Исследователи отмечают, что в небольших текстах из социальных сетей больше ошибок и лишних слов, которые могут помешать корректному анализу. Поэтому целью работы было выявить наиболее эффективные методы для подобных текстов. Ими оказались LDA и NMF — латентное размещение Дирихле и неотрицательная матричная факторизация. Эти выводы важны и для настоящего исследования, ведь материал базы «Прожито» не лишен тех же недостатков: ошибок и опечаток.

Таким образом, выбор базы «Прожито», содержащей более полумиллиона дневниковых записей, представляется оправданным для исследования эмоциональной культуры. А один из используемых методов — тематическое моделирование — не раз оказался эффективным в применении к текстам небольшого объема и дневниковой форме.

# Обработка материала

В качестве материала в исследовании использовались данные базы дневников «Прожито». Данные были получены в виде дампа базы (https://www.dropbox.com/sh/8vfjjt8107sv9r3/AADOVR795MxFovpuGN9PT\_JZa?dl=0). Общее количество дневниковых записей в дампе составляло 384587.

Данные в дампе оформлены в виде csv-файлов. Для их обработки команда «Прожито» разработала Python-библиотеку prozhito-tools (https://github.com/kilomeow/prozhito-tools). В этой библиотеке есть специальный метод, который объединяет csv-файлы в формат prozhitotools.dump.Wrapper. С помощью специальных атрибутов можно получить нужную информацию из объединенных таблиц. Например, атрибут “.ID” выведет ID дневниковой записи, а атрибут “.date” – дату создания. Текст записи можно получить с помощью атрибута “.text”.

На этапе отбора материал не разделялся на временные периоды. Записи отбирались по ключевым словам и просматривались вручную. Для поиска использовали слова «любить», «ревновать», «разлюбить» и т.д. В полученных записях искали другие слова, которые могли стать ключевыми для сбора корпуса. Таким образом, с помощью метода “bootstrapping” был собран список ключевых слов и выражений для сбора корпуса.

Как было замечено, на первом этапе записи не распределялись на временные группы. По этой причине возникла сложность: как отобрать ключевые выражения, если слова меняют значение или выходят из употребления. Например, поиск по слову «свидание» в XIX веке выдавал мало записей на любовную тему. Это слово чаще использовали для обозначения обычной встречи.

По этой причине записи в XIX веке дополнительно анализировали в инструментах AntConc и VoyantTools. Это позволило найти специфические для этих периодов слова, связанные с любовной темой. Например, поиск по слову «волочиться» был продуктивным для записей начала XIX века.

Стоит отметить, что такой метод поиска все равно нельзя назвать исчерпывающим. В корпус могли попасть записи, не связанные с любовной темой. По этой причине от поиска по слову «любить» пришлось отказаться, потому результат содержал много ложноположительных записей.

Таким образом, выбранный метод отбора нельзя назвать абсолютно точным. Однако на этом этапе сложно предоставить формальные критерии для описания любовного переживания. Поэтому корпус отбирался с помощью ключевых слов и регулярных выражений. Посмотреть ход работы более подробно можно в репозитории на GitHub (https://github.com/vicintina/prozhito\_coursework).

## 3.1 Описание корпуса

Основной корпус представлял собой словарь, ключом которого был ID записи, а значением — текст записи. Метаданнные каждой записи — имя автора, дата записи и т. д. — сохранялись в список словарей. Получившийся «любовный» корпус составляет 6283 записи.

Одна из задач исследования – сравнение паттернов разных временных периодов. Поэтому получившийся корпус был разделен на пять периодов. Отдельно был выделен XIX век и период Великой Отечественной войны. В третий период были также включены 1946 и 1947 годы, так как слово «война» оказывается довольно частотным в это время. В 1941-45 гг. у слова «война» 62-й ранг по частотности, в 1946-47 гг. — 61-й ранг. А в 1948-49 гг. слово «война» по частотности на 157-м месте. Хронологические рамки всех периодов представлены в таблице.

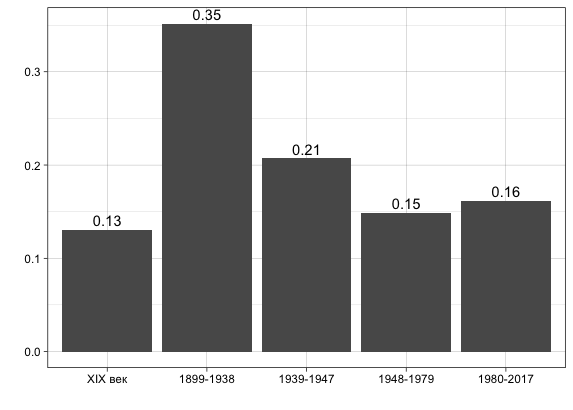
Таблица 1. Временные рамки выделенных периодов

|  |  |  |
| --- | --- | --- |
| Период | Начало | Конец |
| 1 период | 01.01.1799 | 31.12.1898 |
| 2 период | 01.01.1899 | 31.12.1938 |
| 3 период | 01.01.1939 | 31.12.1947 |
| 4 период | 01.01.1948 | 31.12.1979 |
| 5 период | 01.01.1980 | 31.12.2017 |

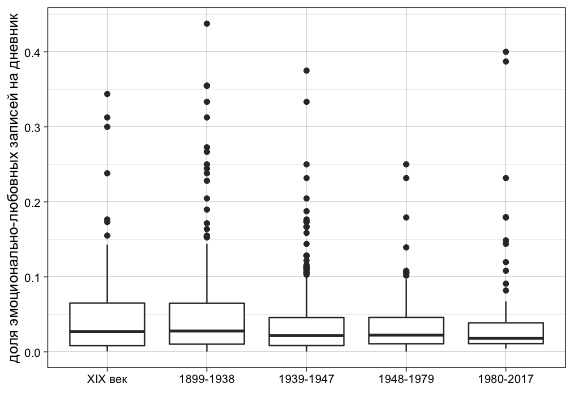
Следующая таблица отображает, как распределились записи получившегося корпуса по временным периодам. Напомним что в базе «Прожито» лучше всего представлен XX век.

Таблица 2. Распределение дневниковых записей

|  |  |  |
| --- | --- | --- |
| Период | Записи в базе | Записи в корпусе |
| 1 период | 50467 | 818 |
| 2 период | 130594 | 2207 |
| 3 период | 73111 | 1301 |
| 4 период | 73045 | 936 |
| 5 период | 53151 | 1017 |

Доли любовных записей по периодам можно описать таким графиком. 

Как уже было замечено, периоды в базе «Прожито» представлены неравномерно. Например, в четвертом периоде было обнаружено несколько крупных дневников, например дневник палеонтолога Олега Амитрова (9399 записей). Больше всего записей в этом дневнике посвящены экспедициям учёного. Другой дневник-рекордсмен принадлежит Владимиру Швецу (9127 записей) — композитору и музыкальному педагогу. Почти весь этот дневник посвящен преподаванию. Можно предположить, что из-за таких специфических дневников доля любовных записей в четвертом периоде так мала.

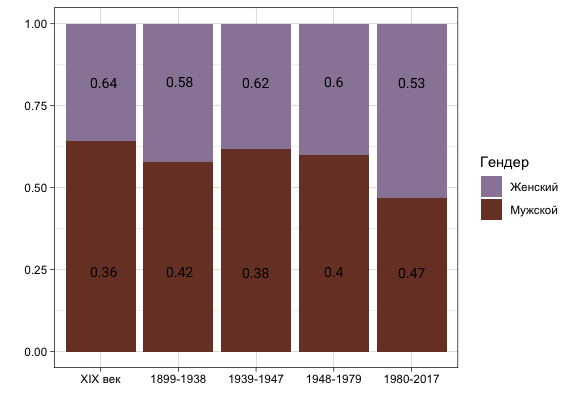
Для сравнения периодов была введена другая мера — на уровне дневников. На графике не отражены дневники, в которых менее 10 записей.

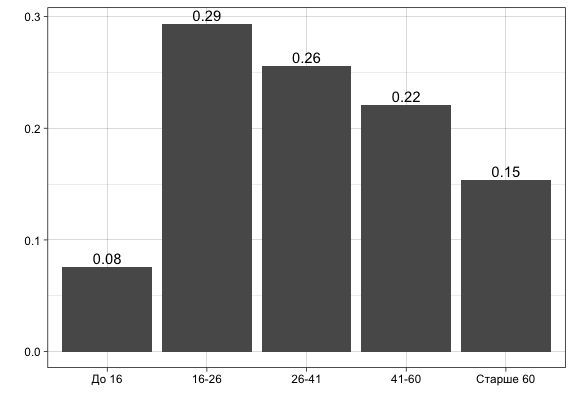
Гендерный состав дневников также неоднороден. Его можно представить в такой таблице.

Таблица 3. Соотношение женщин и мужчин в данных

|  |  |  |
| --- | --- | --- |
|  | Авторы-женщины | Авторы-мужчины |
| В базе | 810 | 2882 |
| В корпусе | 243 | 444 |

Как распределены женские и мужские записи по периодам, можно увидеть на следующем графике.



Охарактеризовать возрастные группы в корпусе можно следующим графиком. Сравнить получившиеся группы корпуса с возрастными группами базы «Прожито» оказалось технически сложным. Однако стоит отметить, что в «любовном» корпусе не попадали дневники преимущественно молодых людей: возрастной состав корпуса довольно разнообразен.

## 3.2 Предобработка материала

Данные в получившемся корпусе было необходимо токенизировать и лемматизировать. Для этого использовались библиотеки nltk и pymorphy2. Данные обрабатывались с помощью функции, в результате которой получались лемматизированные токены с удаленной пунктуацией.

В процессе обработки обнаружились особенности данных базы «Прожито». В некоторых записях встречались орфографические ошибки. Выявлять подобные токены с ошибками оказалось сложным, поэтому их не исключали из анализа.

Также в данных были выявлены биграммы типа «кот орый», «ив анович». Они появились на месте авторских сокращений. Редакторы «Прожито» восстанавливали полную форму, указывая продолжение слова в квадратных скобках, которые удалялись как пунктуационные знаки.

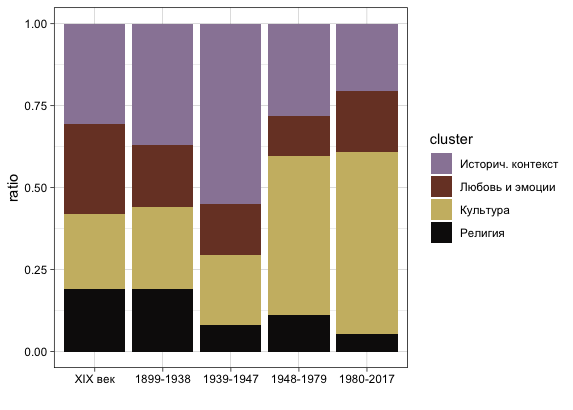
Для тематического моделирования применяли метод LDA из библиотеки gensim. Из корпуса удалялись слова, которые встречались более чем в 10% документов и менее чем в 20 документах. Темы искали для всего корпуса в целом и для каждого периода в отдельности. В результате получилось 20 тем по 10 слов в каждой. Процесс анализа будет подробно описан в следующей главе.

# Анализ корпуса

На первом этапе были проанализированы слова, оказавшиеся в темах. Нужно заметить, что в темах превалировали слова исторического и общественно-политического контекста, например слова «государь», «солдат», «революция», «россия», «война». Это можно объяснить настройкой алгоритмов тематического моделирования, которые отсекали самые частотные слова, которые в этом корпусе были связаны с эмоциями.

Чтобы проанализировать темы, слова разделили на кластеры. В исторический кластер попали слова «война», «революция» и прочие. К эмоциональному кластеру отнесли слова «дружба», «несчастие», «боль», «грустно» и т. д. В культурных кластер вошли слова «стих», «музыка», «зритель», «пьеса» и пр. Слова «митрополит», «храм», «духовный», «христос» отнесли к религиозному кластеру.

Долю кластеров в разные периоды можно увидеть на графике:



Стоит отметить, что в третьем периоде возрастает лексика исторического контекста. В темах этого периода чаще возникают слова «война», «немецкий», «сталин», «ленинград».

Также нужно прокомментировать тот факт, что в четвертом и пятом периодах встречаются дневники театральных деятелей, например дневник Татьяны Дорониной. Так можно объяснить возрастание культурного кластера в этих периодах.

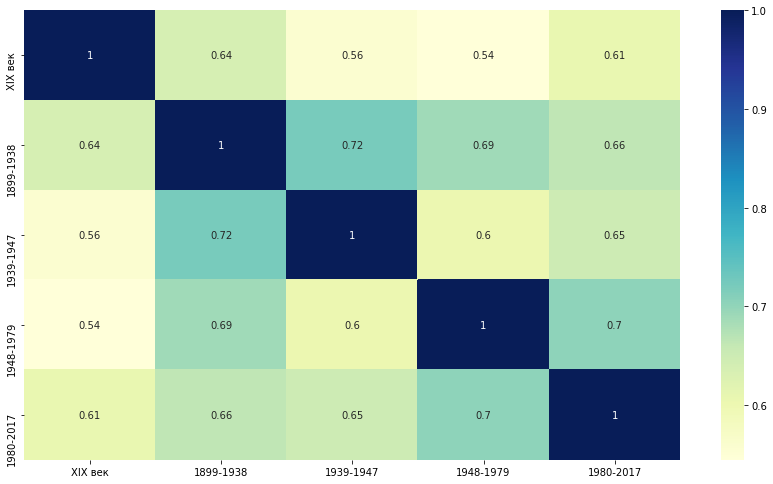
Отметим, что тематическое моделирование не оказалось продуктивным для поиска паттернов: с его помощью удалось только отобразить особенности дневникового материала.

Так как тематическое моделирование не помогло в поиске паттернов, на следующем этапе анализировали частотность слов. Для поиска слов написали функцию, которая удаляла служебные части речи и знаки препинания. Из самых частотных слов составили таблицу, добавив ранг каждого слова. В таблицу вошли первые 200 частотных слов, которые, соответственно, были проранжированы числами от 1 до 200. Получившуюся таблицу можно увидеть в приложении, часть таблицы приведена ниже:

Таблица 4. Пример таблицы с частотными словами

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | XIX век | 1899-1938 | 1939-1947 | 1948-1979 | 1980-2017 |
| мой | 1 | 1 | 2 | 2 | 1 |
| любить | 2 | 2 | 3 | 1 | 2 |
| любовь | 3 | 3 | 1 | 4 | 3 |
| человек | 11 | 7 | 5 | 5 | 5 |
| дело | 14 | 20 | 21 | 20 | 43 |
| друг | 15 | 16 | 9 | 11 | 8 |
| бог | 16 | 24 | 41 | 39 | 18 |

Для оценки зависимостей был выбран коэффициент корреляции Кендалла — так как этот коэффициент учитывает ранги. Для каждой пары периодов был посчитан коэффициент корреляции, и на основе этих данных был нарисован следующий график:



С одной стороны, результаты выглядят довольно ожидаемо: самыми близкими периодами оказались второй (1899-1938 гг.) и третий (1939-1979 гг.), а также четвертый (1948-1979 гг.) и пятый (1980-2017 гг.). Неудивительно, что XIX век стоит особняком.

С другой стороны, нужно обратить внимание на несколько моментов. Во-первых, третий период не так близок к четвертому. Можно сделать вывод о том, что отличия в выборе слов, которые авторы используют для выражения и описания любовных чувств, наблюдаются и в хронологически близких периодах. Таким образом, третий период оказывается ближе ко второму периоду, чем к четвертому. В этом смысле XX век разделяется на две части.

Во-вторых, корреляция XIX века с остальными периодами постепенно снижается от второго периода к четвертому, но неожиданно возрастает в пятом периоде. Близость XIX века с пятым периодом можно объяснить тем, что в пятом периоде обнаружилось большое количество записей поэтессы Галины Ларской (280 записей из 1017 записей в корпусе). Эти записи часто включают стихотворения поэтессы:

*Пускай надеждам никогда не сбыться,*

*лелеять перестал свою мечту.*

*И, как на эшафот самоубийца,*

*я на мучения опять иду.*

*Иду не по дорогам, по утратам,*

*растерзанные дни в один порыв!*

*А нелюбовь твоя моя расплата:*

*другими был любим я, не любив.* (Г. Г. Ларская, 01.04.2015)

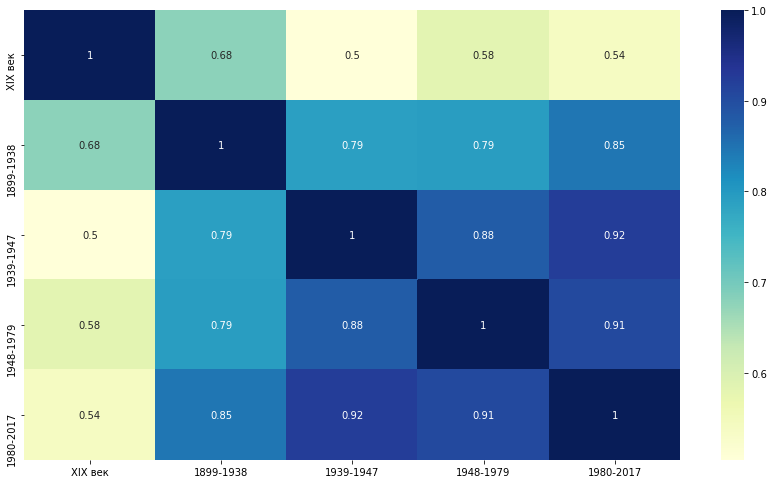
Можно предположить, что в этих записях чаще встречается лексика высокого стиля, поэтому пятый период может сближаться с XIX веком.

На следующем этапе были отобраны часто встречающиеся биграммы, связанные с любовной темой. Биграммы считали с помощью инструмента AntConc и нормализовали, разделив на общее количество биграмм в периоде. Ниже приведен образец получившейся таблицы (таблицу целиком можно найти в Приложении).

Таблица 5. Пример таблицы с долями биграмм

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | XIX век | 1899-1938 | 1939-1947 | 1948-1979 | 1980-2017 |
| страстно любить | 3.149631e-05 | 1.684613e-05 | 1.528402e-05 | 9.321836e-06 | 1.08720e-05 |
| горячо любить | 3.149631e-05 | 2.526920e-05 | 2.445444e-05 | 2.330459e-05 | 0.00000e+00 |
| сердце бьется | 3.937039e-05 | 2.695381e-05 | 1.681243e-05 | 1.165230e-05 | 6.52320e-06 |
| настоящая любовь | 3.149631e-05 | 2.358459e-05 | 3.821006e-05 | 2.796551e-05 | 1.52208e-05 |
| горькие слезы | 1.377964e-05 | 5.053840e-06 | 4.585207e-06 | 0.000000e+00 | 0.000000e+00 |

Корреляцию между периодами посчитали с помощью коэффициента Пирсона. Получившийся график можно увидеть ниже:

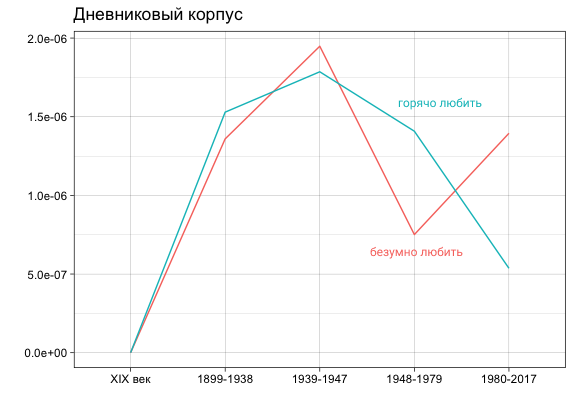
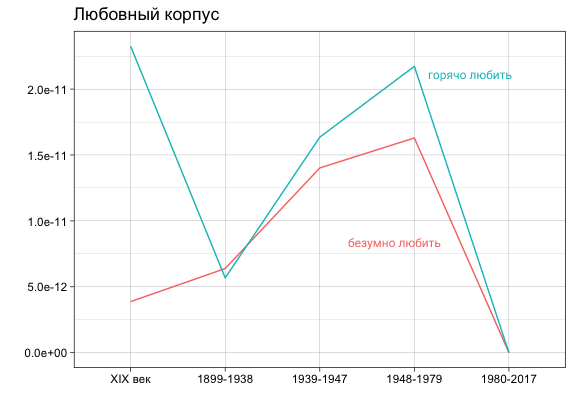


Можно заметить, что XIX век вновь меньше всего коррелирует с остальными периодами, но корреляция уменьшается со временем незначительно. На графике можно также видеть, что у хронологически близких периодов корреляция со временем только повышается: пятый период сильно коррелирует с третьим и четвертым периодами.

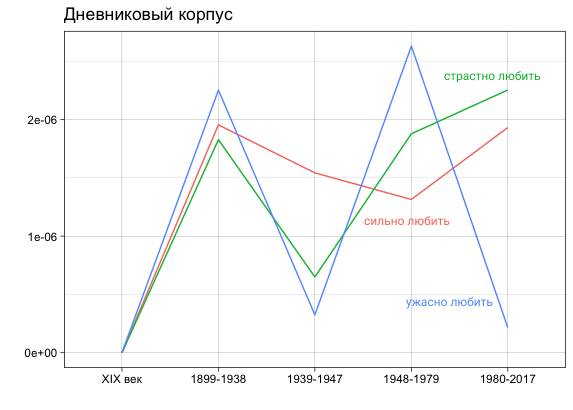
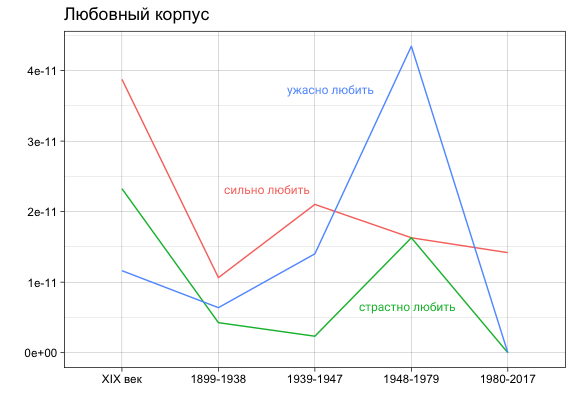
Распределение биграмм можно сравнить на графиках: один визуализирует распределение в отобранном любовном корпусе, второй — в общем дневниковом корпусе «Прожито».

Рассмотрим получившиеся результаты внимательнее. Частотность биграмм изменяется по-разному. В собранном корпусе многие отобранные биграммы встречаются реже во втором периоде (1899-1938 гг.), несмотря на то что в корпусе этот период представлен лучше всего (29% записей). Однако анализ дневникового корпуса показал другую картину: многие отобранные биграммы встречаются гораздо чаще в XX веке, чем в XIX.

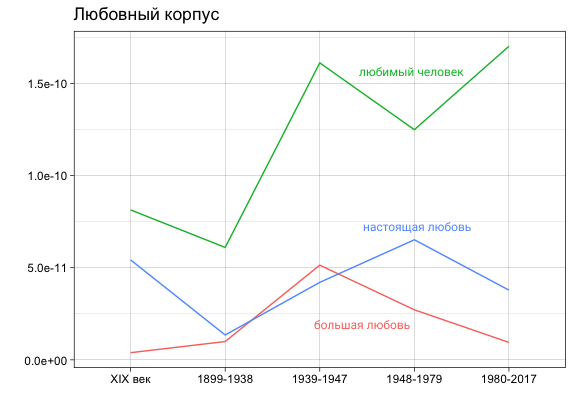
Некоторые результаты выглядят противоположно: биграмма «безумно любить» возрастает в любовном корпусе в четвертом периоде, а в корпусе «Прожито», наоборот, падает. Это можно объяснить тем, что корпус собирали, не используя для поиска слова «любить», «любовь». Таким образом, некоторые записи, посвященные любовной теме, могли не попасть в корпус. Однако такие «популярные» слова не были использованы по такой причине: для задачи была важнее точность, чем полнота.

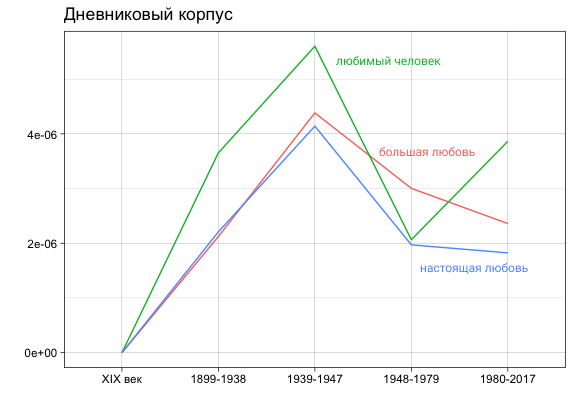


Попробуем проанализировать распределение биграмм. Некоторые данные коррелируют друг с другом, например, в четвертом периоде и в любовном корпусе, и в дневниковом возрастают биграммы «ужасно любить», «страстно любить». Количество этих биграмм падает в третьем периоде, возможно, это объясняется историческими событиями: слово «ужасно» может чаще встречаться в другом контексте.



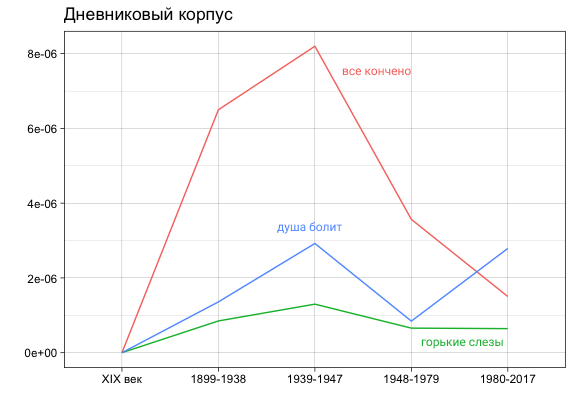
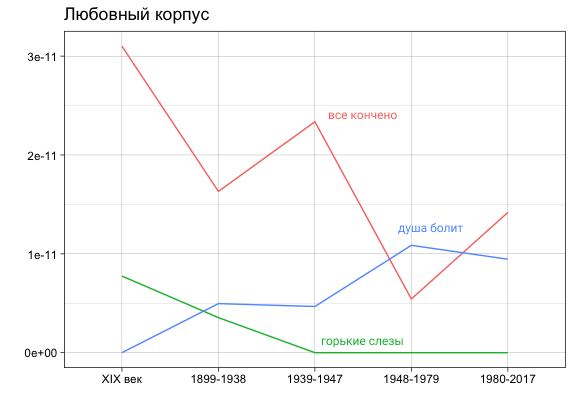
Однако в третьем периоде возрастают другие биграммы: «любимый человек», «большая любовь». Возможное объяснение этому факту приведем немного позже.





В третьем периоде также возрастает биграмма «все кончено», вероятно, этот результат также можно объяснить историческими событиями. Другой примечательный факт — преобладание биграммы «горькие слезы» преимущественно в XIX веке в любовном корпусе, что не соответствует результатам в дневниковом корпусе. Такой результат можно объяснить следующим образом: возможно, биграмма «горькие слезы» чаще использовалась при описании любовных чувств XIX веке, потому что в XX веке также встречается более ироничное употребление этой биграммы «плакать горькими слезами»:

*После чего пьяный Рыбаков сидел под дверью «Литгазеты», плакал горькими слезами и всем жаловался.* (Г. А. Елин, 20.12.1994)



Анализ биграмм выявил определенные диахронические изменения. Некоторые сочетания начинают реже использоваться для описания любовных чувств, например биграмма «горькие слезы». Некоторые биграммы, наоборот, возрастают, например «любимый человек».

Прокомментируем разницу между периодами. Несмотря на исторический контекст, третий период хорошо представлен в корпусе (21% записей) и в этом время возрастает количество многих биграмм. Этот факт можно прокомментировать цитатой одного из дневников:

*Только болтуны могут утверждать, что на фронте не может быть настоящей любви. Поэзию этой любви еще воспоют поэты тонко и глубоко.* (Т. Атабек, 18.03.1945)

Интересную картину представляет также четвертый период, который представлен в корпусе всего 15% записей. Эмоциональность этого периода довольно высока: в это время возрастают биграммы «ужасно любить», «страстно любить», «настоящая любовь».

Чтобы подробнее проанализировать записи, на следующем этапе были использованы методы «близкого» чтения: по ключевым словам отбирались записи разных периодов и размечались вручную. Итого было размечено 160 записей разных периодов. Особое внимание привлекали языковые средства выразительности: устойчивые речевые конструкции (пример: «мучиться ревностью»), эпитеты («черствое сердце»), метафоры и сравнения («сердце сжалось», «как птица в силках»), гиперболы («никогда не смогу забыть»). Также выделялись стилистические средства, например антитеза.

Записи были разделены на темы: «первая любовь», «страдания от любви», «охлаждение чувств», «неразделенная любовь» и другие. Помимо средств выразительности выделялись также цитаты из произведений или отсылки к ним и описания природы.

Количество записей по периодам не было нормализовано, поэтому делать выводы о распространенности той или иной темы в зависимости от периода не стоит. Однако можно заметить, что в третьем периоде (1939-1947 гг.) часто встречаются записи, посвященные темам «желание любви» и «разлука». Эти особенности могут быть объяснить частоту биграммы «любимый человек»:

*Мой первый тост был за самого близкого и* ***любимого человека****, чье чувство ко мне никогда не сможет изменить ни расстояние, ни время, кто не досыпает ночей, думая обо мне, кто часто роняет из-за меня дорогие слезы, кто живет моими письмами, маленькими весточками из этих диких гор и лесов.* (В. И. Стеженский, 01.01.1943)

С помощью близкого чтения обнаружили следующие паттерны. При описании любовной эмоции авторы дневников часто использовали антитезу или сравнивали свое ощущение с состоянием природы.

Например, в следующих цитатах авторы дневников приводят описание природы, рассказывая о своих чувствах.

***Во всем лесу, полном тоже любви и весны****, только в нас было божество! Какое-то общение невидимо происходило между нами… <...> Она тоже чувствовала* ***красоту и поэзию весны***. (В. М. Голицын, 22.04.1922)

*Ненадолго* ***небо становится похожим на нежный серый бархат****. Под ним так уютно, так сладко-томительно. Я встала, отошла… Шура за мной. Он мой сейчас — весь, целиком*. (М. Н. Воробьева, 11.06.1944).

В следующих записях используется противопоставление: радость любви сочетается со страданиями или мрачная действительность раскрашивается светлым любовным чувством.

*Я только сейчас заглянула в глаза истинной любви и поняла, что это такое, поняла, почему она играет такую громадную роль в жизни человека, почему из-за нее,* ***прекрасной****, люди* ***решаются на преступления****, почему* ***из-за нее можно умереть****, почему* ***за радость мига даже любви*** *люди соглашаются* ***терпеть годы страдания****.* (Е. Л. Аргировская, 08.12.1916).

*В душе было* ***мрачно и холодно****, все окружающее казалось* ***серым и обесцвеченным****, и только где-то, далеко-далеко в глубине сердца,* ***теплилась слабая надежда****. Сегодняшняя встреча* ***воскресила мои надежды****,* ***наполнила меня подъемом и бодростью****, и* ***все засияло в блеске яркого солнечного весеннего дня***. (Н. М. Дружинин, 26.03.1921).

Близкое чтение также позволило обнаружить паттерны, которые сложно вычленить автоматически из-за их «размытости». Например, паттерн «не замечать времени», чьим возможным первоисточником может быть цитата из пьесы А. С. Грибоедова «Горе от ума» — «счастливые часов не наблюдают»:

*Мы* ***не замечали, как проходило время до рассвета****: я никогда не забуду эти ночи!* (Владимир К., 20.08.1918)

Триграмму «не замечать времени» проверили по НКРЯ — она вошла в 0.0014 документов в корпусе. Вот один из примеров корпуса, отсылающий к уже упомянутому произведению Грибоедова:

*Дальше события помчались с головокружительной быстротой. Влюбленные не замечают времени.* (Борис Примочкин. Трамбовщик // «Техника — молодежи», 1990)

Нужно отметить, что такой паттерн может встречаться не только в любовном контексте:

*Долго стоял я, не очень замечая, как бежит время, поневоле думая о вершащихся на моем веку переменах…* (О. В. Волков. Из воспоминаний старого тенишевца, 1988)

Из всех вхождений в корпус только 0.073 триграмм связаны с любовной тематикой. Но тем не менее нельзя сказать, что такая триграмма не может быть паттерном для описания любовной встречи.

Проанализируем другой встречающийся паттерн — «быть неспособным любить»:

*До этого* ***я не знал любви*** *и думал, что* ***сердце мое черство****, что* ***я не способен любить****, и вот доказательство как раз противоположного.* (В. П. Трушкин, 19.02.1938)

Триграмму «не способен/на любить» также проверили по НКРЯ — она вошла в 0.00078 документов в корпусе. Приведем цитату из корпуса:

*А я, к сожалению, не способна любить — как писали в старых романах — «всем существом».* (Дина Рубина. На солнечной стороне улицы, 1980-2006)

Нужно заметить, что в этом паттерне может проявляться определенная культурная матрица. Описать ее можно таким примером:

*Час тому назад вы были убеждены, что вы не способны любить, что душа ваша умерла навсегда, навеки, что вам смешон этот лепет, смешон этот вальс… и что же?* (А. П. Чехов. Идиллия, 1884-1885)

Во многих примерах «неспособность» любить оказывается мнимой при встрече с «настоящим» чувством. Возможные культурные источники этого паттерна еще предстоит установить.

Использование «близкого» чтения позволило обнаружить в записях большое количество средств художественной выразительности, употребление высокой лексики. Стоит отметить, что авторы дневников часто прибегают к лексике высокого стиля и в более поздних периодах, а не только в XIX веке.

*Лицо горело я* ***трепетал*** *счастье я хорошо умел скрыть свое волнение.* (К. А. Березкин, 10.09.1849)

*Что это было в моем сердце?! Какой-то необъяснимый* ***трепет*** *и, с тем же, ужас.* (И. К. Пескова, 21.04.1941)

*С какой застенчивостью и с каким* ***трепетом*** *в сердце я смотрел на нее, на свой идеал.* (Е. И. Иванов, 01.02.1948)

Можно обнаружить разницу в употреблении конкретных биграмм, но тем не менее общим местом является описание чувств словами высокого стиля, даже если иногда такие слова сочетаются с обычной лексикой:

*Как пламя в сердце угасить?*

*его залить водою трудно,*

*а от любви оно горит.* (Л. И. Яшина, 08.01.1939)

# Заключение

С помощью инструментов «дальнего» и «близкого» чтения удалось проанализировать способы выражения любовных чувств и эмоций и сравнить их использование в разные периоды.

Сравнение частотности биграмм показало, что средства выражения чувств меняются с течением времени: некоторые биграммы теряют популярность или начинают использоваться в другом контексте, как биграммы «горькие слезы» или «слезы восторга». Такие изменения можно объяснить культурными и языковыми процессами.

Другая причина — исторические события. Возрастание в третьем периоде биграмм «любимый человек» и «настоящая любовь» можно объяснить историческим контекстом: автор дневника, находясь в разлуке с возлюбленной или возлюбленным, мог чаще предаваться воспоминаниям:

*Мне самому очень приятно, когда получаешь известие от* ***любимого человека****. Сразу вспоминается многое, настоящее на миг отходит в забытье, мир становится еще прекрасней и полней. И после всего этого сердце наполняется большой обидой за разлуку, обидой на коварного и бесчеловечного врага.* (В. А. Савченко, 01.07.1942)

Благодаря «близкому» чтению удалось выявить паттерны, которые сложно выделить автоматически. Например, к таким паттернам можно отнести использование антитезы или сравнения психологического состояния с явлениями природы.

Многие паттерны и их культурную обусловленность еще предстоит изучить. Одной из перспектив исследования может быть анализ культурных матриц, стоящих за описываемыми чувствами. К таким культурным матрицам можно отнести отсылки к литературным произведениям и сравнениям с персонажами литературы:

*Я так же,* ***как и Печорин****, не могу кого-нибудь одного любить.* (Е. И. Иванов, 01.02.1948)

Такие матрицы связаны не только с культурными источниками, но и с конкретной лексикой: «черствое сердце», «прежняя наивная любовь». Поэтому анализ таких паттернов должен сочетать как точные методы, так и методы «близкого» чтения.

Особый интерес также представляют случаи, когда культурные матрицы сталкиваются с действительностью и влияют на оценку событий:

*Рыцарей не существует. И по-королевски люди не умеют любить тоже. Все прекрасные позы и прекрасные слова оказались приспособленными лишь к мирному времени, если даже смотреть на эти позы и слова только как на театр. <…> Как смешно и дико вспоминать теперь — вот сегодня! — такие торжественные и красивые слова!* (С. К. Островская, 29.11.1941)

Авторы дневников часто указывают на литературные источники (*Если сделать так, что Дон Кихота — Рыцаря не было, а его выдумала влюбленная Дульцинея.* С. К. Островская, 29.11.1941), поэтому в перспективе нужно анализировать такие паттерны вместе с литературным материалом-первоисточником.

# Литература

Гинзбург 1999 — Гинзбург Л. Я. *О психологической прозе*. М.: INTRADA, 1999. 415 с.

Зарецкий 2011 — Зарецкий Ю. П. *Стратегии понимания прошлого: Теория, история, историография*. М.: Новое литературное обозрение, 2011. 384 с.

Зорин 2016 — Зорин А. Л. *Появление героя. Из истории русской эмоциональной культуры конца XVIII – начала XIX вв*. М.: Новое литературное обозрение, 2016. 568 с.

Паперно 2018 — Паперно И. *«Кто, что я?» Толстой в своих дневниках, письмах, воспоминаниях, трактатах*. М.: Новое литературное обозрение, 2018. 232 с.

Савкина 2007 — Савкина, И. Л. *Разговор с зеркалом и Зазеркальем: Автодокументальные женские тексты в русской литературе первой половины XIX века*. М.: Новое литературное обозрение, 2007. 416 с.

Albalawi et. al. 2020 — Albalawi, R., Yeap, T. H., Benyoucef, M. *Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis*. Front. Artif. Intell, 3:42, 2020.

Blevins 2010 — Blevins, C. *Topic modeling Martha Ballard’s diary.* 2010.

Frijda, Mesquita — Frijda N. H., Mesquita B. The Social Roles and Functions of Emotions. // *Emotion and culture: Empirical studies of mutual influence.* American Psychological Association, 1994. P. 51-87.

Lejeune 2009 — Lejeune, Ph. *On diary*. Honolulu: University of Hawaii Press, 2009. 365 p.

Moretti 2013 — Moretti, I. *Distant Reading.* London, New York: Verso, 2013. P. 224.

Paperno 2004 — Paperno, I.What Can Be Done with Diaries? // *The Russian Review* Vol. 63, No. 4 (Oct., 2004). P. 561-573.

Plamper 2015 — Plamper, J. *The History of Emotions: An Introduction*. Oxford: Oxford Press, 2015. 352 pp.

# Приложение

Таблица с рангами самых частотных слов в корпусе

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | XIX век | 1899-1938 | 1939-1947 | 1948-1979 | 1980-2017 |
| мой | 1 | 1 | 2 | 2 | 1 |
| любить | 2 | 2 | 3 | 1 | 2 |
| любовь | 3 | 3 | 1 | 4 | 3 |
| человек | 11 | 7 | 5 | 5 | 5 |
| дело | 14 | 20 | 21 | 20 | 43 |
| друг | 15 | 16 | 9 | 11 | 8 |
| бог | 16 | 24 | 41 | 39 | 18 |
| душа | 17 | 15 | 26 | 22 | 13 |
| чувство | 22 | 34 | 32 | 67 | 54 |
| жизнь | 23 | 6 | 7 | 10 | 6 |
| письмо | 25 | 19 | 33 | 32 | 52 |
| сердце | 27 | 58 | 47 | 66 | 74 |
| мысль | 30 | 47 | 43 | 60 | 57 |
| час | 32 | 31 | 34 | 45 | 63 |
| рука | 33 | 27 | 28 | 35 | 28 |
| первый | 34 | 50 | 59 | 46 | None |
| хороший | 24 | 18 | 19 | 17 | 25 |
| чувствовать | 42 | 43 | 50 | 50 | 78 |
| лицо | 45 | 53 | 66 | 38 | 46 |
| последний | 44 | 36 | 51 | 49 | 56 |
| минута | 46 | 102 | 151 | 168 | 109 |
| ребёнок | 47 | 44 | 36 | 48 | 69 |
| голова | 52 | 69 | 100 | 74 | 76 |
| церковь | 55 | 189 | None | None | None |
| сила | 56 | 49 | 30 | 61 | 61 |
| глаз | 57 | 38 | 27 | 31 | 38 |
| отец | 59 | 159 | 181 | 105 | 122 |
| счастие | 61 | 105 | 88 | 150 | 110 |
| случай | 65 | None | None | None | None |
| вечер | 67 | 60 | 63 | 95 | 88 |
| женщина | 73 | 52 | 29 | 122 | 66 |
| жена | 75 | 88 | 94 | 106 | 95 |
| свет | 76 | None | 112 | 103 | 86 |
| молодой | 78 | None | None | None | None |
| великий | 77 | None | None | None | None |
| вместе | 79 | None | None | 94 | 111 |
| мать | 80 | None | 56 | 97 | None |
| отношение | 81 | 63 | 58 | 76 | 89 |
| добрый | 86 | None | None | None | None |
| брат | 87 | None | None | None | None |
| счастливый | 93 | 163 | 166 | None | 179 |
| вид | 94 | None |  | None | None |
| дорога | 95 | None | 80 | 120 | 129 |
| твой | 97 | 104 | 91 | 142 | 103 |
| вчера | 98 | 87 | None | None | None |
| разговор | 99 | 100 | None | 80 | 92 |
| прекрасный | 100 | 197 | 190 | 128 | 199 |
| один | None | 9 | 13 | 12 | 10 |
| день | 19 | 13 | 8 | 18 | 17 |
| слово | None | 14 | 16 | 15 | 12 |
| дом | 36 | 25 | 15 | 14 | 15 |
| конец | 160 | 55 | 45 | 42 | 58 |
| комната | 143 | 72 | 96 | 89 | 195 |
| работа | None | 74 | 52 | 21 | 32 |
| мама | 184 | 75 | 62 | 40 | 99 |
| милый | 141 | 90 | 174 | None | None |
| книга | 156 | 91 | 123 | 52 | 37 |
| вопрос | 153 | 94 | None | 88 | 112 |
| ночь | 134 | 107 | 70 | 149 | 128 |
| смерть | 129 | 109 | 76 | 118 | 115 |
| театр | None | 111 | 196 | 73 | 26 |
| радость | 194 | 113 | 85 | 152 | 131 |
| мир | 142 | 114 | 77 | 62 | 50 |
| война | None | None | 17 | 170 | None |
| город | None | 199 | 46 | 111 | 96 |
| немец | None | None | 49 | None | None |
| любимый | None | 161 | 53 | 100 | 68 |
| девушка | None | None | 90 | None | None |
| земля | 168 | 129 | 93 | 121 | None |
| фронт | None | None | 95 | None | 123 |
| путь | None | None | 103 | 188 | None |
| живой | None | 180 | 104 | None | None |
| настоящий | None | 141 | 106 | None | 145 |
| правда | 113 | None | 107 | 83 | 114 |
| встреча | None | None | 153 | None | 101 |
| сон | None | None | 188 | None | 98 |
| муж | None | 173 | 170 | 182 | None |
| стих | None | 167 | None | 47 | 19 |
| год | None | None | None | 116 | None |
| вера | 105 | 165 | 157 | 126 | None |
| голос | 196 | 170 | None | 127 | 87 |
| семья | None | None | None | 129 | 193 |
| спектакль | None | None | None | 135 | 40 |
| музыка | None | None | None | 139 | 132 |
| улица | None | 191 | 127 | 155 | None |
| искусство | None | None | None | 172 | 180 |
| слеза | None | None | None | 187 | None |
| сцена | None | None | None | None | 71 |
| судьба | None | None | None | None | 77 |
| поэт | None | None | None | None | 83 |
| роль | None | None | None | 194 | 97 |
| боль | None | None | None | None | 119 |
| петь | None | None | None | None | 135 |
| народ | 117 | 145 | 175 | None | 138 |
| россия | None | 177 | None | None | 140 |
| взгляд | None | None | None | None | 143 |
| год | None | None | None | None | 147 |
| образ | 111 | 155 | 146 | None | 152 |
| страна | None | None | None | None | 154 |
| близкий | None | None | 158 | None | 156 |
| красивый | 197 | None | None | 195 | 159 |
| фильм | None | None | None | None | 163 |
| история | None | None | None | 199 | 164 |
| актёр | None | None | None | None | 167 |
| роман | None | None | None | None | 168 |
| умереть | None | None | None | None | 172 |
| плохой | None | None | None | None | 175 |
| месяц | 169 | 168 | 136 | None | 184 |
| смысл | None | None | None | None | 194 |
| нравиться | 147 | None | None | None | 198 |

Таблица с частотностью биграмм

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | XIX век | 1899-1938 | 1939-1947 | 1948-1979 | 1980-2017 |
| страстно любить | 1,00E-05 | 1,00E-05 | 0 | 1,00E-05 | 0 |
| горячо любить | 1,00E-05 | 1,00E-05 | 1,00E-05 | 1,00E-05 | 0 |
| сердце бьётся | 4,00E-05 | 3,00E-05 | 2,00E-05 | 1,00E-05 | 1,00E-05 |
| настоящая любовь | 3,00E-05 | 2,00E-05 | 3,00E-05 | 3,00E-05 | 2,00E-05 |
| поэзия любви | 0 | 1,00E-05 | 2,00E-05 | 1,00E-05 | 1,00E-05 |
| слова любви | 5,00E-05 | 3,00E-05 | 3,00E-05 | 1,00E-05 | 2,00E-05 |
| мой ангел | 4,00E-05 | 1,00E-05 | 0 | 1,00E-05 | 0 |
| горькие слезы | 0 | 0 | 0 | 0 | 0 |
| любить по-настоящему | 0 | 0 | 3,00E-05 | 2,00E-05 | 2,00E-05 |
| душа болит | 0 | 1,00E-05 | 0 | 0 | 0 |
| сильно любить | 2,00E-05 | 1,00E-05 | 1,00E-05 | 1,00E-05 | 1,00E-05 |
| ужасно любить | 1,00E-05 | 1,00E-05 | 1,00E-05 | 2,00E-05 | 0 |
| безумно любить | 0 | 1,00E-05 | 1,00E-05 | 1,00E-05 | 0 |
| большая любовь | 0 | 1,00E-05 | 3,00E-05 | 1,00E-05 | 0 |
| потерять голову | 0 | 1,00E-05 | 1,00E-05 | 0 | 0 |
| душа изнемогает | 0 | 1,00E-05 | 0 | 0 | 0 |
| свет жизни | 2,00E-05 | 2,00E-05 | 2,00E-05 | 2,00E-05 | 1,00E-05 |
| страх разлуки | 0 | 0 | 0 | 0 | 0 |
| слезы умиления | 0 | 0 | 0 | 0 | 1,00E-05 |
| слезы катились | 1,00E-05 | 0 | 1,00E-05 | 0 | 0 |
| слезы восторга | 1,00E-05 | 0 | 0 | 0 | 0 |
| мой любимый | 0 | 1,00E-05 | 3,00E-05 | 3,00E-05 | 2,00E-05 |
| моя милая | 0 | 1,00E-05 | 1,00E-05 | 0 | 0 |
| моя любимая | 1,00E-05 | 2,00E-05 | 1,00E-04 | 5,00E-05 | 6,00E-05 |
| любовь моя | 4,00E-05 | 2,00E-05 | 6,00E-05 | 6,00E-05 | 5,00E-05 |
| моя душа | 1,00E-05 | 3,00E-05 | 1,00E-05 | 2,00E-05 | 4,00E-05 |
| сильное впечатление | 1,00E-05 | 2,00E-05 | 1,00E-05 | 0 | 2,00E-05 |
| все кончено | 2,00E-05 | 2,00E-05 | 2,00E-05 | 0 | 1,00E-05 |
| это любовь | 2,00E-05 | 2,00E-05 | 4,00E-05 | 0 | 2,00E-05 |
| любимый человек | 4,00E-05 | 7,00E-05 | 0.00011 | 5,00E-05 | 8,00E-05 |
| любви нет | 1,00E-05 | 1,00E-05 | 1,00E-05 | 0 | 1,00E-05 |
| радость жизни | 0 | 1,00E-05 | 2,00E-05 | 0 | 1,00E-05 |
| моя любовь | 5,00E-05 | 9,00E-05 | 1,00E-04 | 8,00E-05 | 8,00E-05 |