Воробьева Виктория Александровна

КОЛИЧЕСТВЕННЫЙ ПОДХОД К ИЗУЧЕНИЮ ДНЕВНИКОВ НА МАТЕРИАЛЕ ПРОЕКТА «ПРОЖИТО»

Выпускная квалификационная работа — МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

по направлению подготовки Фундаментальная и прикладная лингвистика

образовательная программа «Цифровые методы в гуманитарных науках»

Рецензент

к.ф.н.

Д.А. Скоринкин

Руководитель

к.ф.н., доцент Школы лингвистики

НИУ ВШЭ

А.А. Бонч-Осмоловская

Москва 2022

Contents

1. Introduction

The diary as a genre has always been ambiguous and complex, balancing "between literary and historical writing, between the spontaneity of reportage and reflectiveness of the crafted text, between selfhood and events, between subjectivity and objectivity, between the private and the public" (Langford and West 1999: 8).

For a long time, researchers have been struggling to define its specifics, and the hitherto common historical framework has been only recently replaced by the intertextual approach. It has become realizable mostly due to the works of French structuralist Philippe Lejeune (Lejeune 2009).

Due to its genre ambiguity and comprehensiveness, the diary gains a lot of interest both from researchers and common readers. Nevertheless, there are not so many Russian-language works on the theory of autobiographical narrative. Among the most major ones are the works by Lidiya Ginzburg (Ginzburg 2009), Yuriy Zaretsky (Zaretsky 2011) and Irina Paperno (Paperno 2014).

Another remarkable research was conducted by Andrey Zorin in his book *The Emergence of a Hero: From the History of Russian Emotional Culture in the Late 18th – Early 19th Centuries* (Zorin 2016). Zorin shows how the emotions of a Russian nobleman were influenced by social and cultural phenomena of that time like Masonic society, theater, and literature. The research focuses on the diary written by Russian *romantic before Romanticism* Andrey Turgenev. In this work, an individual experience is examined as an expression of a collective tendency. Therefore, this research paradigm is important for our study.

In addition to various researches, growing interest in diaries in the post-Soviet space was marked by the launch of the Prozhito project. Prozhito is a large database of intimate papers written by people in the Russian Empire and modern Russia but mostly in the Soviet Union. Started in 2015, Prozhito has become a source for reflection on history both from national and personal perspective.

The Prozhito database offers an opportunity to explore diaries on a large scale, with computable methods. However, opportunities go hand-in-hand with challenges, and from this perspective, a researcher who studies the Prozhito database faces certain

intricacies due to the data's heterogeneity. A closer overview of these peculiarities is given in the second chapter of this work.

As we stated earlier, the Prozhito database contains diaries of different time periods: from the XVIII century to the present day. These records are the *subject* of the current study. We examined them at the first stage of the research in order to make a corpus of texts that include descriptions of love-related emotions.

The research is based on the *hypothesis*: there are certain patterns laying behind the description of the experience of love. These patterns can be expressed as idioms, literary references, and recurring collocations. At the second stage of the study, they are retrieved from the corpus and analyzed.

In our study, we will try to address several *questions*:

- What recurring collocations are used by the authors of diaries when they describe love-related emotions?
- Do these patterns change over time?
- Do the authors of diaries use literary references when they describe love-related emotions?

The *purpose* of the study is to identify and compare patterns that occur in records describing love-related emotions in different time periods.

In order to achieve this purpose, we defined several *tasks*:

- to make a corpus of records describing love-related emotions and divide them into several groups according to the time period
- to retrieve and analyze idioms and recurring collocations found in records
- to draw conclusions about trends observed in the records describing love-related emotions

There have been no studies similar to Zorin's work that would be conducted on a large amount of data. It is much more difficult to analyze the cultural basis of emotions by examining thousands of diaries. However, it seems reasonable to acquire a larger amount of data due to the purpose of the current study—to compare patterns of different periods.

The work consists of an introduction, three chapters, a conclusion, a list of references, and an appendix. The first chapter gives an overview of theoretical literature and research methodology. The second chapter describes data and procedure of text processing. The third chapter presents an analysis of diary records and compares patterns of different time periods. The conclusion part provides the prospects of the study. The appendix contains the tables used in the analysis part.

2. Literature review

2.1. Novel autobiographical methodologies

As it was stated in the introduction, autobiographical documents have been examined through the historical approach for a long time. Diaries were considered as a source of everyday practices of the past, and researchers were therefore mostly focused on *what* is written in a diary than on *how* it's written. However, women's diaries were approached from a different perspective; it was examined how women were constructing their selves under the male gaze (Savkina 2007).

Nevertheless, diaries of the past not only described daily life and habits. Philippe Lejeune stated that the first personal diaries were written in the late XVIII and early XIX centuries (Lejeune 2009: 100). Remarkably, these diaries had a formula *dear diary*, which "indicated that diarists begin to address their diary as though it was an intimate friend in whom they could confide" (Popkin 2009: 1). Lejeune emphasizes that "every diarist found original words—always the same ones—to speak to his dear diary" (Lejeune 2009: 97).

The main idea of Lejeune's autobiographical theory reveals the difference between memoir and diary narratives. According to Lejeune, diary, unlike autobiography, is "virtually unfinishable from the beginning" (191). That means that diary is turned toward the future, while autobiography is turned to the past. Despite the fact that a diary record can observe the events already passed, there always will be a "time lived beyond the writing, making it necessary to write anew" (191).

It is reasonable that even unpublished private diaries address a potential reader, which sometimes is expressed as a future self. Being *strictly assessed* by these future readers, diarists often scrutinize their own emotions and deeds. A diary is therefore deemed suitable to study emotional culture.

In the above mentioned work *The Emergence of a Hero*, Andrey Zorin introduced a similar approach. Examining Andrey Turgenev's emotions, Zorin implemented the methodology of Dutch psychologists Nico Frijda and Batja Mesquita (Frijda and Mesquita 1994), who divided the emotional process into several parts.

Coding is the first part; it takes place when a person distinguishes the event that triggers an emotion. The attitude to this event is expressed through *evaluation*—it can

be fear, joy, or anger. According to Frijda and Mesquita, coding and evaluation are conditioned by cultural norms, rules, and taboos. This concept constitutes a crucial premise of the current research.

From the methodological perspective, another remarkable research is *The History of Emotions* by Jan Plamper. In his book, Plamper combined psychological and cultural approaches to emotions. Plamper also addressed the question to what extent emotions are universal and culturally conditioned.

Thus, in several works, emotions were considered as a manifestation of a collective self, not an individual one. Nevertheless, it should be also taken into account that emotional analysis was mostly conducted on the basis of texts written by noble people from the upper classes. In addition, behavioral patterns were neither examined on a large scale nor compared with each other depending on the time period. In the current study, we combine these two features. Although, it involves several problems, which will be discussed in the second chapter of the work.

2.2. Computable approaches

Quantitative methods were also fruitful for analyzing autobiographical documents. In the paper *Topic Modeling Martha Ballard's Diary* (Blevins 2010), Cameron Blevins used quantitative methods to identify topics and patterns that occured in the diaries of Martha Ballard who lived in Massachusetts in the XVIII century.

Due to quantitative methods, it was possible to identify patterns invisible at first glance. Remarkably, it was observed that the words *informed* or *hear* were associated with the *death* topic. The possible reason is the context—the news about death was circulating "through face-to-face interactions."

It should be also mentioned that "short, content-driven entries" (Blevins 2010) in Ballard's diary were suitable for this kind of analysis. Therefore, they afford the opportunity to produce "remarkably cohesive and accurate topics."

Topic modeling seems to have a lot of potential in examining autobiographical documents. Another possible method considers retrieving frequent or meaningful collocations.

Similar method was applied in the research *A "wind of change"—shaping public opinion of the Arab Spring using metaphors* (Núñez et al. 2018). In order to define how

the Arab Spring was covered in German media, researchers annotated various collocations like *wave of revolutions*, *self-consciousness of the generation* and clusterized them.

This method can be productive for retrieving patterns, although it was applied to specific data that could be more saturated with rhetorical expressions. The third chapter of this work shows whether this method is similarly productive for examining emotions.

In order to settle on a method for extracting collocations, we will refer to the article *Evaluation of collocation extraction methods for the Russian language* (Pivovarova et al. 2018). While mostly all the methods showed relatively similar results, such methods as t-score, log-likelihood, and Dice gave better performance. According to the researchers, in some cases, even the frequency method could compete with the others.

To sum it up, several methods are deemed efficient for detecting patterns—topic modeling and extracting collocations. In the third chapter, we compare the outcomes of these methods applied to the emotional corpus. The next chapter describes how this corpus was built.

3. Building the corpus

This chapter describes how data from the Prozhito database was preprocessed in order to build a corpus of love-related records.

3.1. Data preprocessing

Data from the database was contained in dump files with interconnected csv-files. The total number of diary entries was 384587. In order to process the files, we used the Python library prozhito-tools (Usov 2018) developed by the Prozhito team members.

This library has a special method that combines csv-files into the format of *prozhitotools.dump.Wrapper*. With the help of special attributes, you can get the necessary information from the combined tables. For example, the *ID* attribute outputs the ID of the diary entry, and the *date* attribute displays the creation date. The text of the record can be obtained using the *text* attribute.

When the corpus was built, the data was not divided into time periods. In our research, we focused on the records of the XIX, XX and XXI centuries. Despite the fact that the Prozhito database also has records of the XVIII century, there are quite a few of them. There are 623 records compared to 50467 in the XIX century.

Suitable records were retrieved with keywords and manual help. We used phrases with keywords *to love* (lyubit'), *to be jealous of* (revnovat'), *to stop loving* (razlyubit'). The obtained records were examined for other keywords and expressions that could have potential in reaching love-related records. Thus, using the bootstrapping method, a list of keywords and expressions for building a corpus was completed.

As we mentioned above, records were not divided into groups before the selection process. For this reason, certain words and expressions gave better results in the XX century compared to the XIX century. For instance, the word *date* (svidanie) was less productive in the XIX century because it was more often used to denote a meeting in general.

Records created in the XIX century were therefore additionally examined with AntConc and VoyantTools. These instruments made it possible to detect words and phrases associated with the love topic that would be specific to this period. For example, the word *to dangle* (volochit'sya) was productive for the beginning of the XIX century since it occurred mostly in the love-related records.

Nevertheless, it should be also mentioned that the results are not free from error and incompleteness. For this reason, keywords were put into phrases, otherwise the resulting corpus would contain many false positive entries.

On the whole, the chosen method was not absolutely accurate. However, presenting formal criteria of love-related emotion seems to be quite challenging. In this work, we made an attempt to grope for them. The detailed working process can be seen in the repository on GitHub (Vorobiova 2022).

3.2. Corpus description

The corpus was built as a dictionary with a key as a record ID and a value as a record itself. The metadata of each entry was also collected: the diarist's name, date of entry, etc. The resulting corpus consisted of 6283 entries.

One of the objectives of the study was to compare patterns of different time periods. Therefore, the resulting corpus was divided into five parts.

The first period includes the entire XIX century. As it was mentioned above, due to language differences, the XIX century is set apart from the XX century. The second period is associated with the fall of the Russian Empire and the Formation of the Soviet Union.

The third period comprises the Second World War and a post-war period. In order to define the boundaries of this period, we compared the frequency of the word *war* with different years. Since this word turns out to be frequent in 1946–47 (the 61st rank), these years were included in the third period. In 1948–49, this word was ranked 157th in frequency.

The fourth period comprised postwar reconstruction and the era of stagnation in the USSR, and the last period saw the dissolution of the Soviet Union and the establishment of modern Russia. The boundaries of the periods can be seen in Table 1.

Table 1. Allocated periods' boundaries

Period	When it begins	When it ends
Period I	January 1, 1799	December 31, 1898
Period II	January 1, 1899	December 31, 1938
Period III	January 1, 1939	December 31, 1947
Period IV	January 1, 1948	December 31, 1979
Period V	January 1, 1980	December 31, 2017

An important peculiarity of the Prozhito database and diaries in general is heterogeneity. Unsurprisingly, diaries written by prominent people or created during dramatic times have more chances to be saved. The Prozhito database reflects this specifics as well.

Therefore, there is no surprise that it is hard to divide the database evenly. The table below shows how the records of the resulting corpus were distributed over time periods. It should be also emphasized that there are more records of the XX century compared with the XIX century.

Table 2. Records' distribution

Period	Total number of records in the database	Total number of records in the corpus
Period I	50467	818
Period II	130594	2207
Period III	73111	1301
Period IV	73045	936
Period V	53151	1017

The graph below shows how corpus records are distributed over time periods.

Figure 1. Corpus records' distribution over time periods
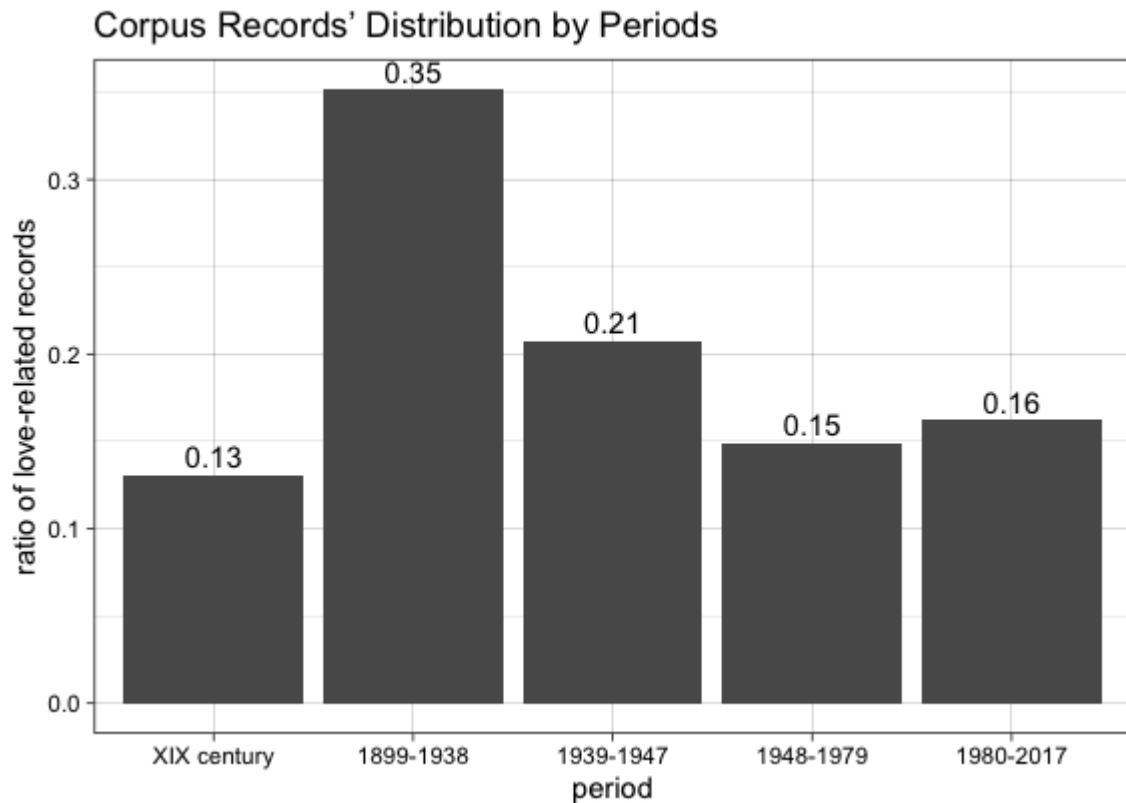
Another consequence of the diaries' heterogeneity is the presence of extremely large diaries in the database. For example, in the fourth period (1948–1979), we discovered the diary of paleontologist Oleg Amitrov (9399 entries) devoted to his expeditions. Another significant diary belongs to Vladimir Shvets (9127 entries), a composer and music teacher. In his diary, Shvets mostly described the teaching process. The presence of these two large diaries may be a possible reason why there are not so many love-related records in the fourth period.

Another way to analyze the number of love-related records over time periods is to compare how many relevant entries per diary are there in a period. We took the diaries that have more than 10 entries and calculated the ratio of love-related records per diary.

Figure 2. Ratio of love-related records per diary

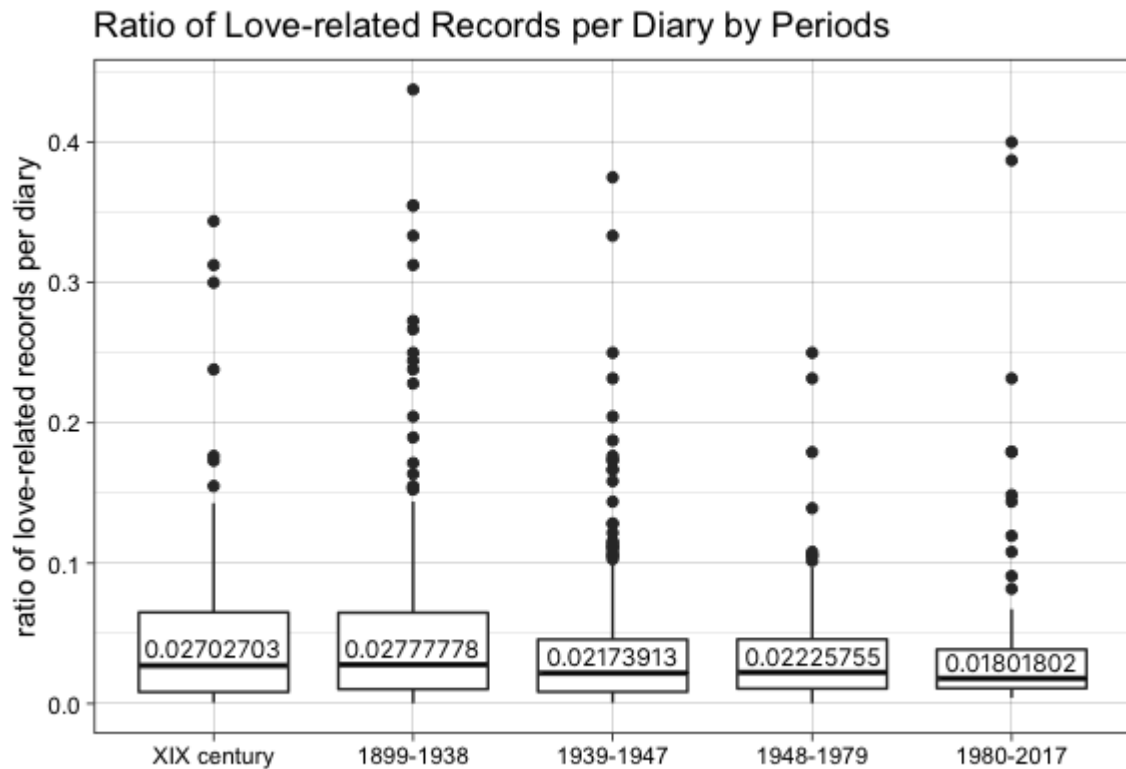## Ratio of Love-related Records per Diary by Periods

Figure 2 shows that the ratio of love-related records is slightly bigger in the XIX century and at the beginning of the XX century than in other periods. It is also noteworthy that the second (1899–1938) and the third periods (1939–1947) have more outliers.

The diary with the biggest ratio (0.438) belongs to Ernestine Edelman. We took a closer look at it and found out that the diarist indicated the intention for keeping a diary: "The diary comes to an end. And it will describe two of my beautiful romances, those days when I was happy" (Valieva 2016: 698). Therefore, the number of relevant records in this diary is reasonable.

Gender composition of the corpus is heterogeneous as well. Table 3 shows how many female and male diarists are there in the database and in the corpus.

Table 3. Gender composition of the corpus

	Female diarists	Male diarists
Database	810	2882
Corpus	243	444

The graph below shows how female and male records included in the corpus are distributed throughout time periods.

Figure 3. Distribution of female and male records in the corpus by periods
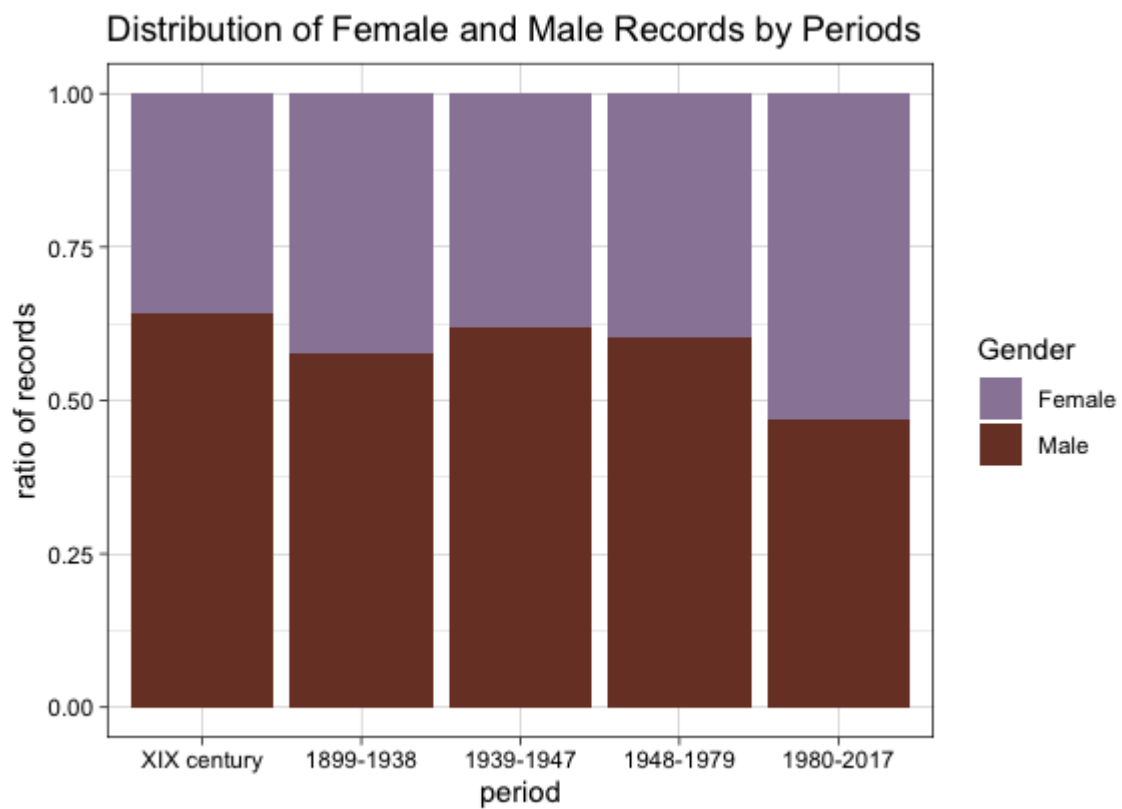
Regarding age, it turned out to be technically difficult to ascertain age groups of diarists in the corpus. The cause lies in the peculiarity of the diary itself and its dynamic temporality. The age of a diarist is literally changing from record to record.

In the next chapter, we describe methods applied to the corpus and their outcomes.

4. Analyzing the corpus

In this section, we describe the outcomes of four methods applied to the corpus. We practiced topic modeling, collocations retrieval, morphological markup, and close reading.

4.1. Corpus preprocessing

Before proceeding to the analysis step, corpus data was preprocessed. The corpus was tokenized and lemmatized with nltk and pymorphy2 libraries. It was processed with a Python function that returned lemmatized tokens with all punctuation removed.

The preprocessing stage revealed the Prozhito database peculiar properties. There were spelling mistakes and author's abbreviations found, and it turned out to be difficult to identify tokens with errors. Therefore, they were not excluded from the corpus.

4.2. Topic modeling

We used the LDA method from the gensim library for topic modeling. Words that were found in more than 10% of documents and less than 20 documents were removed. Topics were retrieved from the entire corpus and each period separately.

We got 20 topics of 10 words each in the outcome. Topics were formed mostly by words of historical and socio-political context like *worker, soviet, government*, *Russia*, and *war*. This could happen due to the mathematical modeling algorithms that cut off the most frequent words associated with emotions in this corpus.

To analyze the resulted topics, the words included in them were divided into 5 clusters:

- The words *emperor*, *government*, *war* and others were classified as the historical cluster.
- The words *quarrel*, *smile*, *sadness*, *friendship* were attributed to the emotional cluster.
- The cultural cluster included the words *art*, *actor*, *portrait*, *theater*, etc.
- The words *monk*, *blessing*, *temple* were attributed to the religious cluster.

The graph below shows how these clusters are distributed over time periods.

Figure 4. Ratio of words belonging to a particular cluster by periods
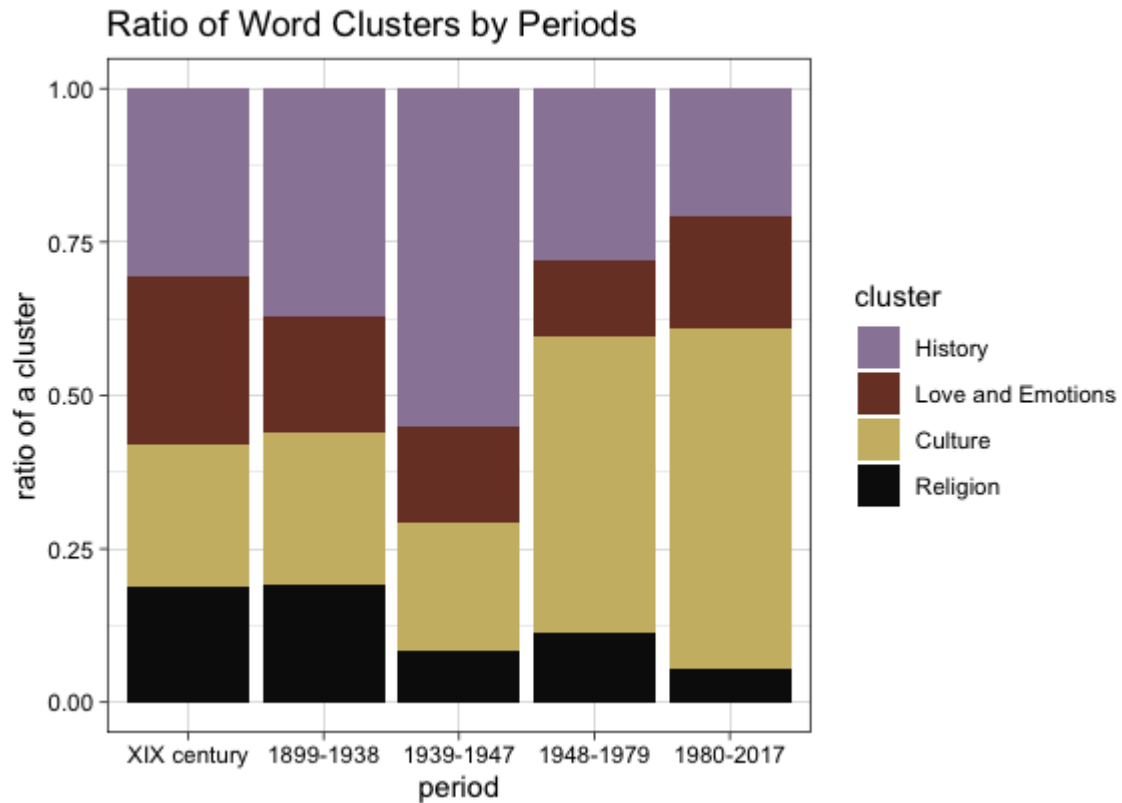
The graph shows that the words of historical context are increasing in the third period. The fact that war-related words like *soldier*, *battle,* and *enemy* were much more common that time is explainable.

It is also noteworthy that in the fourth and fifth periods the words of the cultural cluster are rising. The possible reason is that the database has numerous diaries of theater, cinema and literature workers created in these periods. For example, the corpus included diaries belonging to the literary figures Lydia and Korney Chukovsky, the actress Tatiana Doronina.

To sum up, topic modeling did not turn out to be productive for retrieving patterns; with its help, it was only possible to make the features of the data being on the surface even more obvious.

4.3. Collocations

As we showed in the Literature Review section, collocations can be productive for identifying patterns. Relying on the research by Pivovarova, Kormacheva, and Koptev (Pivovarova et al. 2018), we examined empirical collocations, which are understood as

co-occurrences "that 1) are frequent enough to be extracted automatically and 2) may be semantically and/or syntactically bounded to various extents" (Pivovarova et al. 2018: 137).

For our purposes, we retrieved collocations based on their frequency. Collocations were extracted using the NLTK Python library. The sample of the resulting table with a normalized number of resulting bigrams is given below.

Table 4. Sample of the table with collocations' ratio

collocation	XIX century	1899-1938	1939-1947	1948-1979	1980-2017
bit'sya serdce		0.00003272 0871334989 9734			
bolshoy lyubov'		0.00003490 2262757322 38	0.00006508 1735002488 42		
verit' lyubov'			0.00003828 3373530875 54		
govorit' lyubov'	0.00009555 3253336820 19	0.00010688 8179694299 8	0.00010719 3445886451 52	0.00007756 4475470234 63	0.00006673 5112936344 97
dusha lyubit'	0.00004526 2067370072 724				0.00004106 7761806981 516

The correlation between the periods was calculated using the Pearson coefficient. The resulting graph can be seen below.

Figure 5. Bigrams correlation between the periods (Pearson coefficient)
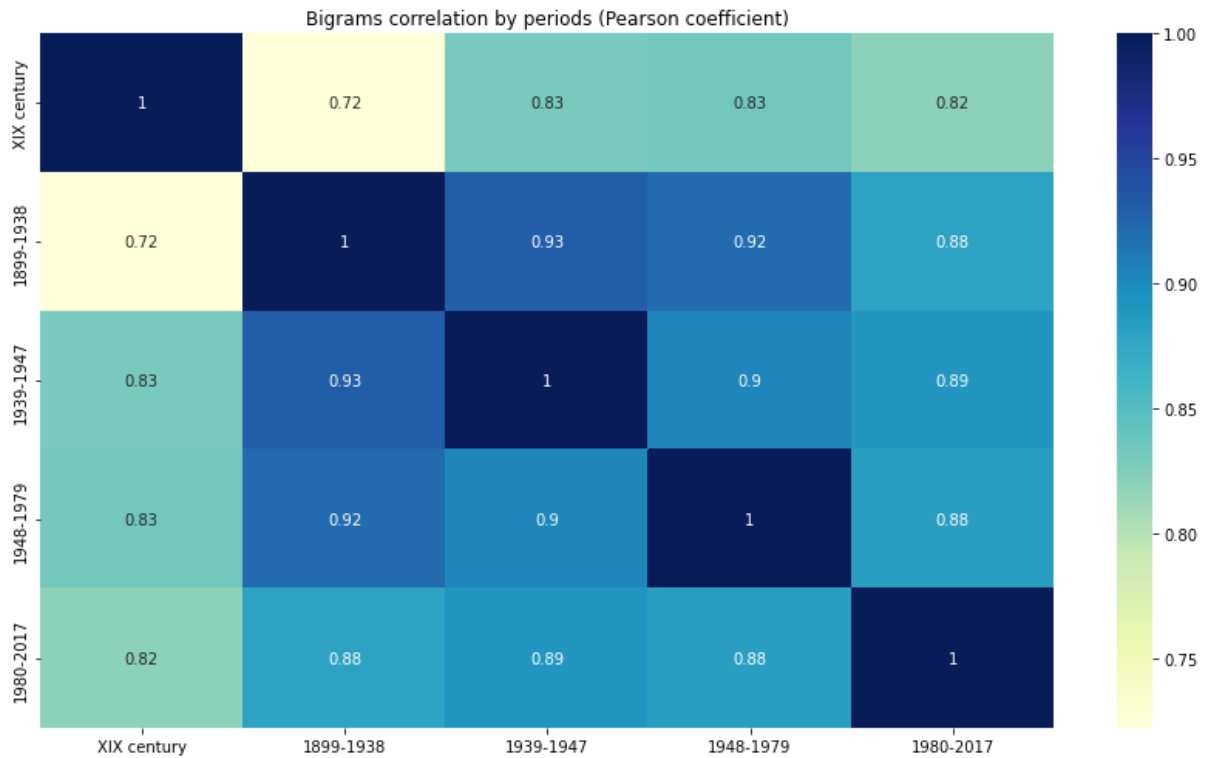
Remarkably, the XIX century is set apart from the other periods and especially from the beginning of the XX century. Correlation between the other periods is stronger. The closest connections can be observed between the beginning of the XX century and the period of the Second World War and between the beginning of the XX century and the postwar period.

In order to conduct a deeper analysis of the collocations, we grouped data by gender and compared the outcomes. We selected a few bigrams and took a closer look at the records with them. The analysis of the selected collocations is given below.

4.3.1. Bit'sya serdce

Technically, we combined collocations *heart to be beating* (serdce bit'sya) and *beating heart* (bit'sya serdce) into one to evaluate the frequency correctly. Below we consider them as a single collocation.

Regarding the ratio of the collocation, it increased during the XIX century (0.0000704077) and at the beginning of the XX century (0.0000741673).

Figure 6. Frequency of the collocation *heart to be beating* (bit'sya serdce) in the corpus
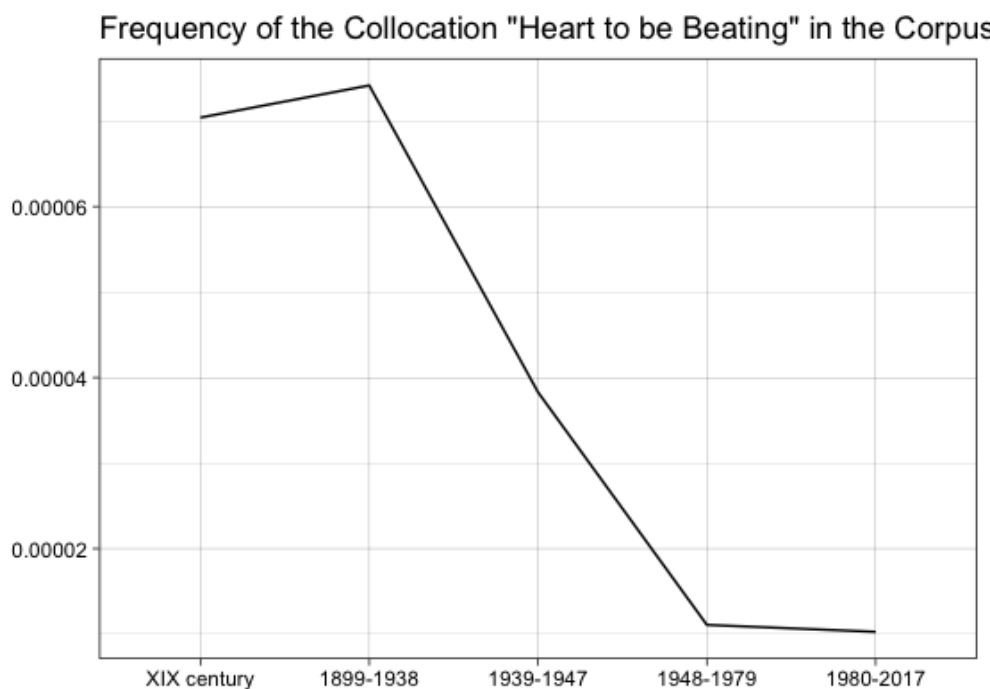
After examining the examples closely, we ascertained that the collocation was frequently used in the love-related context.

(1)   *I don't know if anyone has experienced it; but I was afraid that **my heart was beating** so hard that others would hear it. (Ne znayu, ispytyval li eto kto-nibud'; no ya boyalas', chto moe **serdce b'etsya** tak sil'no, chto eto uslyshat drugie.)* [Maria Bashkirtseva, May 18 (6), 1873]

(2)   *I left, and my **heart was beating** as if it wanted to shake my chest. (Ushel, a **serdce bilos'** tak, slovno ono hotelo rastryasti grudnuyu kletku.)* [Valentin Smyshlyaev, January 16 (3), 1918]

Example 1 is noteworthy due to the expression "my heart was beating **so hard that others would hear it**." We kept looking for the same formula, but the only example found belonged to the same diarist. The expression can be literary-based, as Example 3 shows.

(3)   ***My heart was pounding** so hard that I was afraid it would be heard, as **they say in novels**. (**Serdce moe stuchalo** tak sil'no, chto ya boyalas', kak by etogo ne bylo slyshno, **kak govoryat v romanah**.)* [Maria Bashkirtseva, May 31 (19), 1876]

Also, we examined the difference in the frequency depending on gender. We found out that the collocation was much more frequent in female diaries. Remarkably, the overall trend of frequency change is comparatively similar both for female and male diaries.

Figure 7. Frequency of the collocation *heart to be beating* (bit'sya serdce) by gender
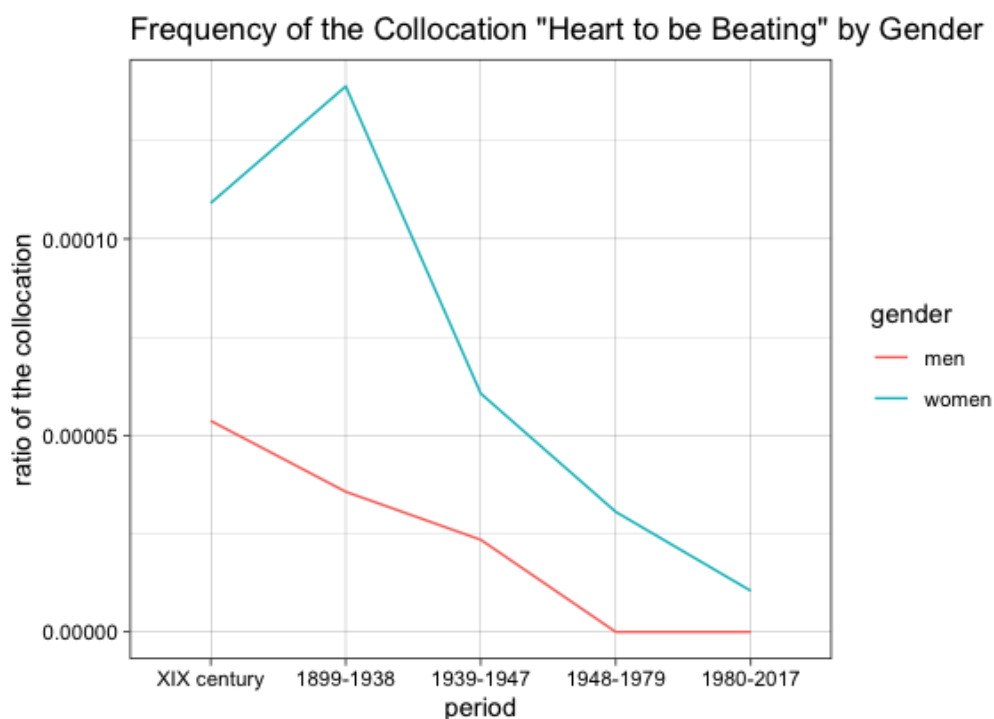
Also, we examined the context of the collocation during the Second World War and in the postwar period. Remarkably, at this time, the collocation was used in other contexts more often, as Examples 4 and 5 prove.

(4)   *Father died, although his **pulse and heart were still beating**, and his right temple was gushing "torrents" of blood. (Papa byl mertv, hotya **pul's i serdce esche bilis'**, a krov' lilas' «klyuchom» iz ego pravogo viska.)* [Anna Aratskaya, November 14, 1942]

(5)   *The sight of a dying man is terrible. **My heart was beating** with fear. (Vid umirayuschego cheloveka uzhasen. Ot straha **bilos' serdce**.)* [Galina Larskaya, 1969]

Whether the observed difference is a consequence of the gradually disappearing pattern or the heterogeneity of the autobiographical data is debatable.

We also compared the outcomes with the frequency of the collocation in the database. Similarly, the collocation saw a dramatic rise in frequency at the beginning of the XX century, as Figure 8 proves.

Figure 8. Frequency of the collocation *heart to be beating* (bit'sya serdce) in the database

Frequency of the Collocation "Heart to be Beating" in Data
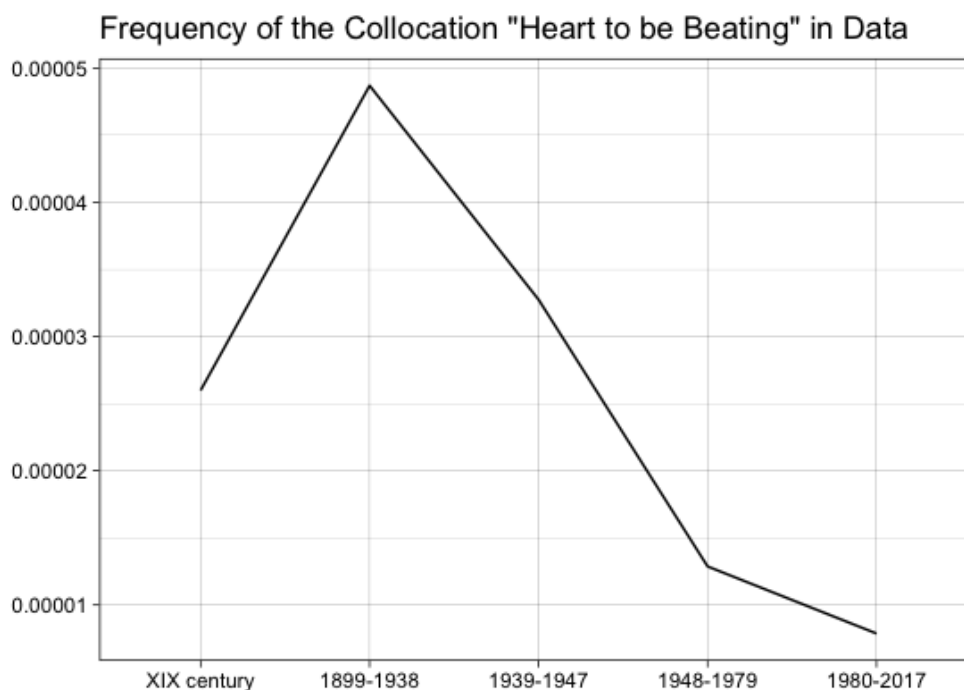
We examined the relevant records found at the beginning of the XX century. It turned out that the ratio of records with the collocation used in the love-related context was 0.2. Accordingly, the ratio of other contexts was 0.8.

4.3.2. Vzyat' ruka

The collocation *to take hand* (vzyat' ruka) was mostly used in the love-related context in the corpus.

Most of the examples were found in the records of three periods: in the XIX century (with a ratio of 0.00011064), at the beginning of the XX century (with a ratio of 0.00010253), and during the Second World War (with a ratio of 0.00015696). The frequency is shown on Figure 9.

Figure 9. Frequency of the collocation *to take hand* (vzyat' ruka) in the corpus
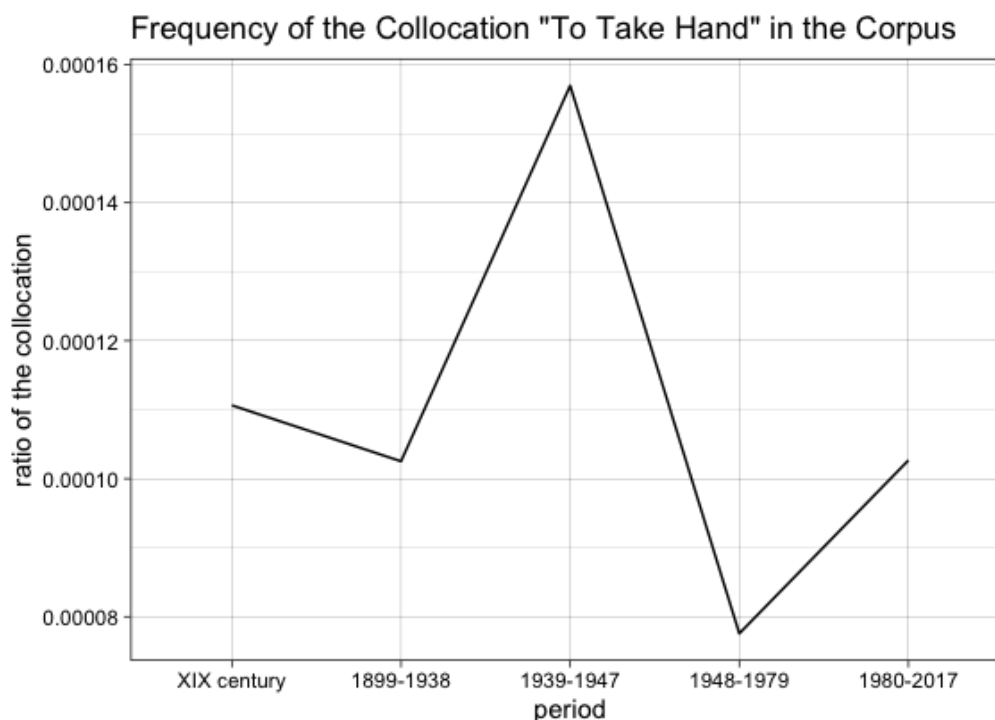
The most notable examples with the collocation are given below.

(6)   *I was amazed. He wanted **to take my hands**, but I didn't let him do it. (Ya byla porazhena. On hotel **vzyat' menya za ruki**, no ya ne dala.)* [Apollinaria Suslova, September 29 (17), 1865]

(7)   *[...] while chatting and laughing, I could have been near her, could imperceptibly **take her hand** and shake it. ([...] boltaya i smeyas', ya mog byt' vozle nee, mog nezametno **vzyat' ee ruku** i pozhat'.)* [Leonid Andreev , August 24 (12), 1898]

(8)   *I just **took his hand** and looked at him intently, intently. It was painful and scary. (Ya tol'ko **vzyala ego ruku** i posmotrela na nego pristal'no-pristal'no. Bylo bol'no i strashno.)* [Irina Knorring, January 21, 1927]

It's also noteworthy how the collocation is distributed in female and male records. In the XIX century and at the beginning of the XX century the collocation was seen more frequently in female diary records (with ratios of 0.00027284 and 0.00016901 respectively). Male diarists started using the bigram more frequently during the Second World War (0.00014659). Figure 10 shows changes in frequency that happened over periods.

Figure 10. Frequency of the collocation *to take hand* (vzyat' ruka) by gender
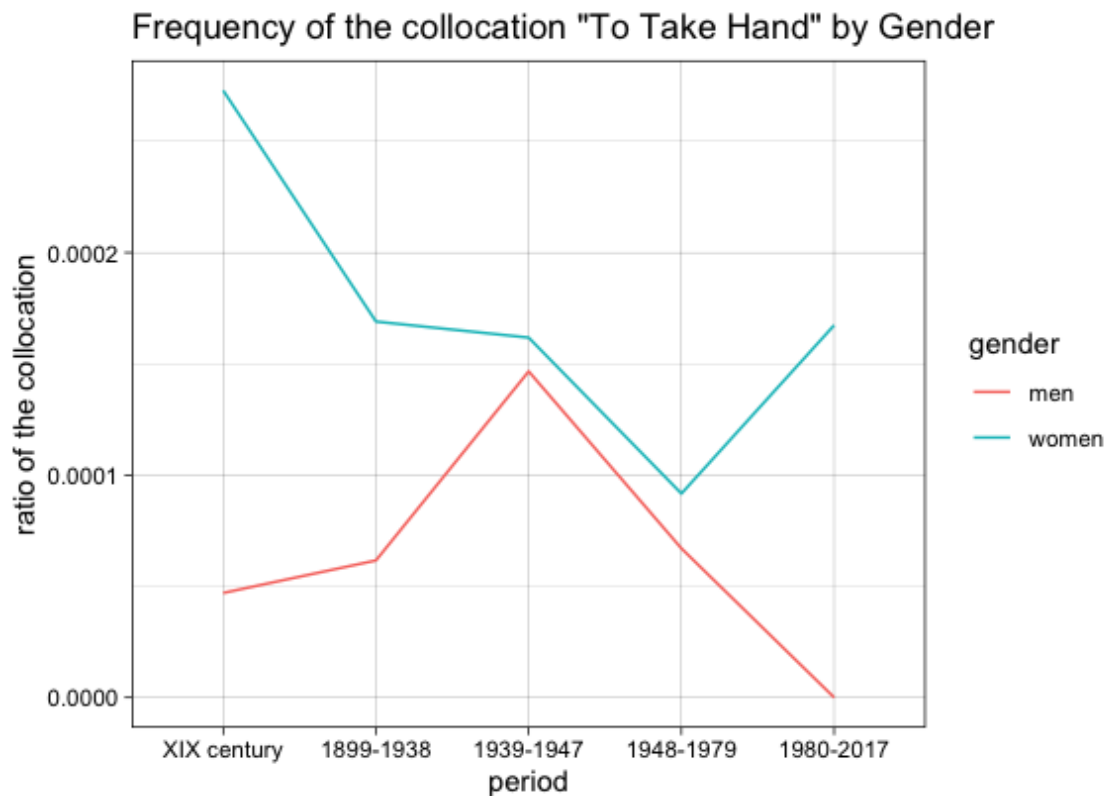
Example 9 shows that diarists could also use the collocation to describe military events during the Second World War. This can be the possible cause of the increase in the frequency at this time. The bigram was also used in the love-related context, which is proven by Example 10.

(9)   *I took the patient's hand — there was no pulse, his eyes were closed. (Ya vzyal ruku bol'nogo — pul'sa ne bylo, glaza byli zakryty.)* [Yakov Polonsky, January 20, 1942]

(10)   *I wanted to take her hand and bring her closer to me, but she shrank back and hid both hands under her cape. (Ya hotel vzyat' ee ruku, priblizit' k sebe, no ona otpryanula nazad i spryatala obe ruki pod nakidku.)* [Georgiy Knyazev, February 17, 1942]

In the two subsequent periods, diarists rarely used the bigram in love-related records. We found only one relevant record given in Example 11.

(11)   *When he **took my hand**, I gently took it away and began to speak. (Kogda on*
***vzyal ruku**, ya ostorozhno otnyala ee i nachala govorit'.)* [Nina Byalosinskaya-Evkina,
April 16, 1957]

Also, we compared the results with the database. We encountered a dramatic rise
of the collocation during the Second World War.

Figure 11. Frequency of the bigram *to take hand* (vzyat' ruka) in the database
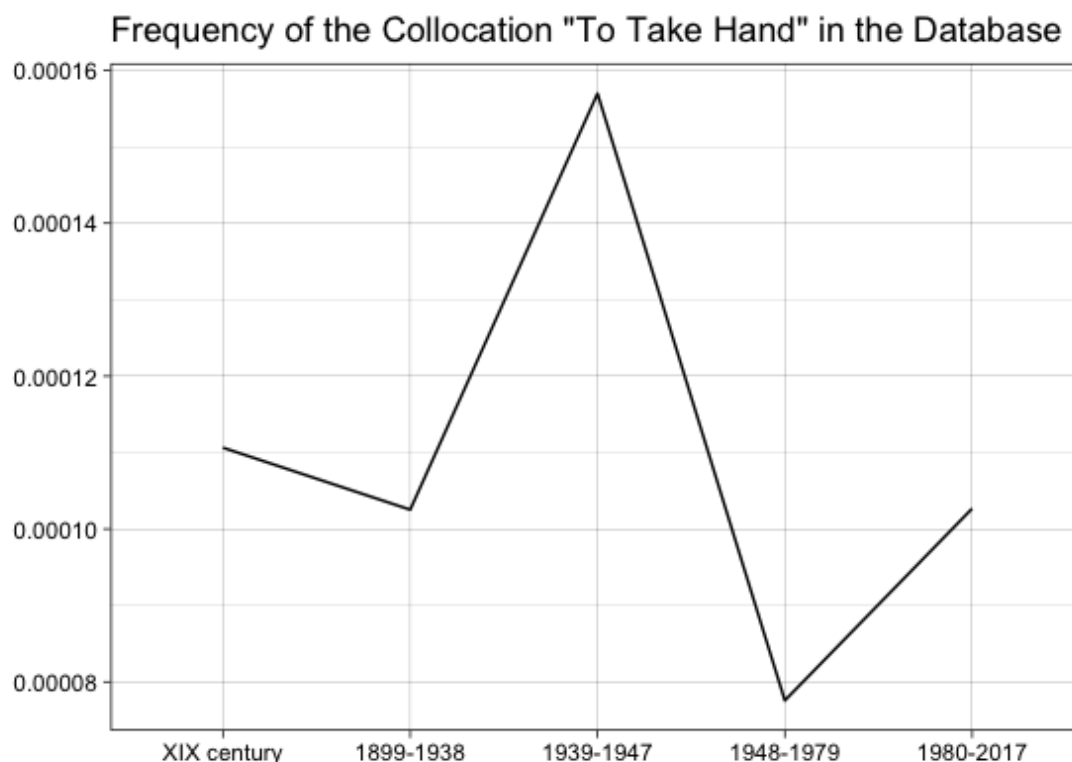
We examined the context in the records with the collocation found. The love-
related usage was comparatively frequent (with ratio 0.36 among all relevant records),
and along with that, the collocation occurred in the war context (in 0.27 of all entries),
as Example 12 shows.

(12)   *An intelligent woman K. 40 years old received a notice of the death at the front*
of her only son 18 years old. [...] A soldier enters with two rhombuses. [...] With a firm,
*military step, he approached her bed, knelt down, **took her hand**... said nothing...*
(Intelligentnaya zhenschina K. 40 let poluchila izveschenie o smerti na fronte
edinstvennogo syna 18 let. [...] Vhodit voennyj s dvumya rombami. [...] Tverdym,

*voennym shagom podoshel k krovati, opustilsya na koleni, **vzyal za ruku**... molchit...*)
[Alexander Dreitzer, October 4, 1942]

4.3.3. Govorit' lyubov'

The collocation *to talk love* (govorit' lyubov') occured in records of all periods, especially at the beginning of the XX century and during the Second World War (with ratios 0.00010689 and 0.00010719 respectively).

Figure 12. Frequency of the collocation *to talk love* (govorit' lyubov') in the corpus
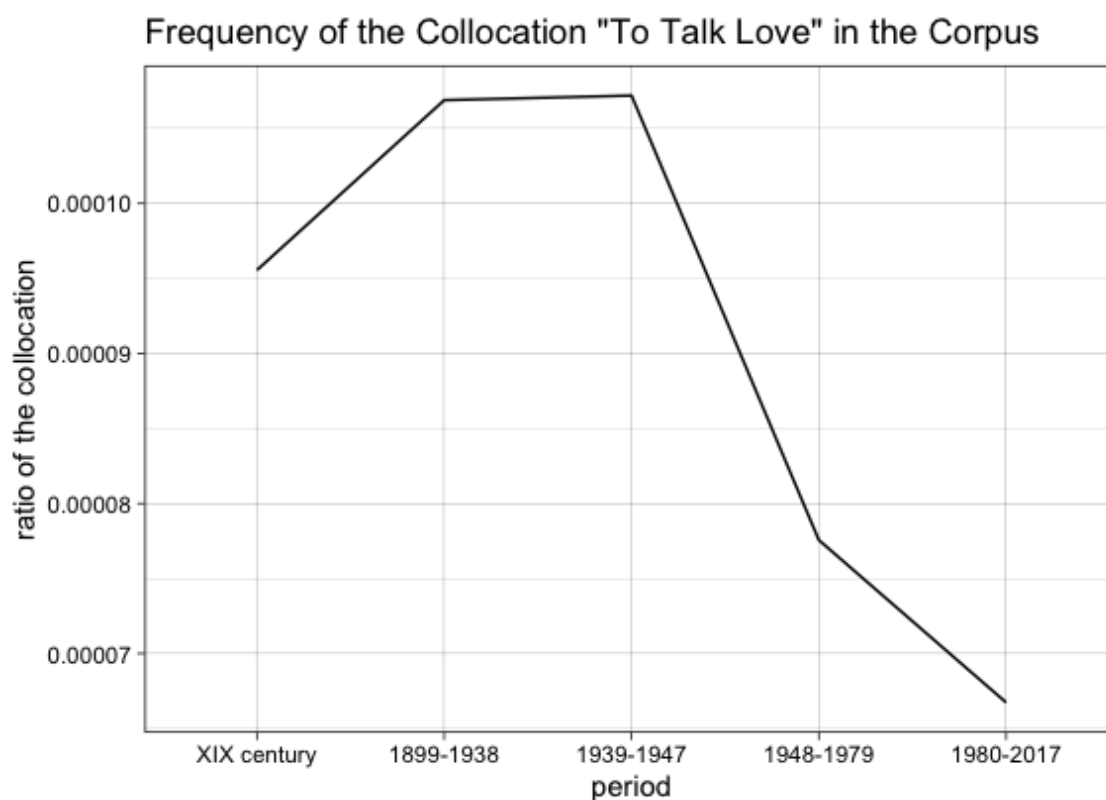
The most notable examples with the collocation are given below.

(13)   *During the mazurka we **talked about love** and marriage. (Za mazurkoj my **govorili o lyubvi i o zamuzhestve**.)* [Tatiana Sukhotina-Tolstaya, January 8 (December 27), 1884]

(14)   *I am very glad to have such a friend, but if he **talks about love**, he will become unbearable to me. (Imet' takogo druga ya ochen' rada, no esli on budet **govorit' o lyubvi**, on stanet dlya menya nevynosim.)* [Marina Allendorf, September 10, 1939]

(15)   *Now Juraev is constantly **talking to me about love**. But I don't like him. (Teper'
mne nepreryvno **govorit o lyubvi** Dzhuraev. No ne nravitsya on mne.)* [Rosalia Serdnak,
February 3, 1944]

As is proved by Examples 14 and 15, the collocation was sometimes used in the
meaning of the bigram *to declare love*.

We compared the use of the bigram among men and women: it was more frequent
in female diaries in the XIX century and in the first half of the XX century, which
Figure 13 shows.

Figure 13. Frequency of the collocation *to talk love* (govorit' lyubov') by gender
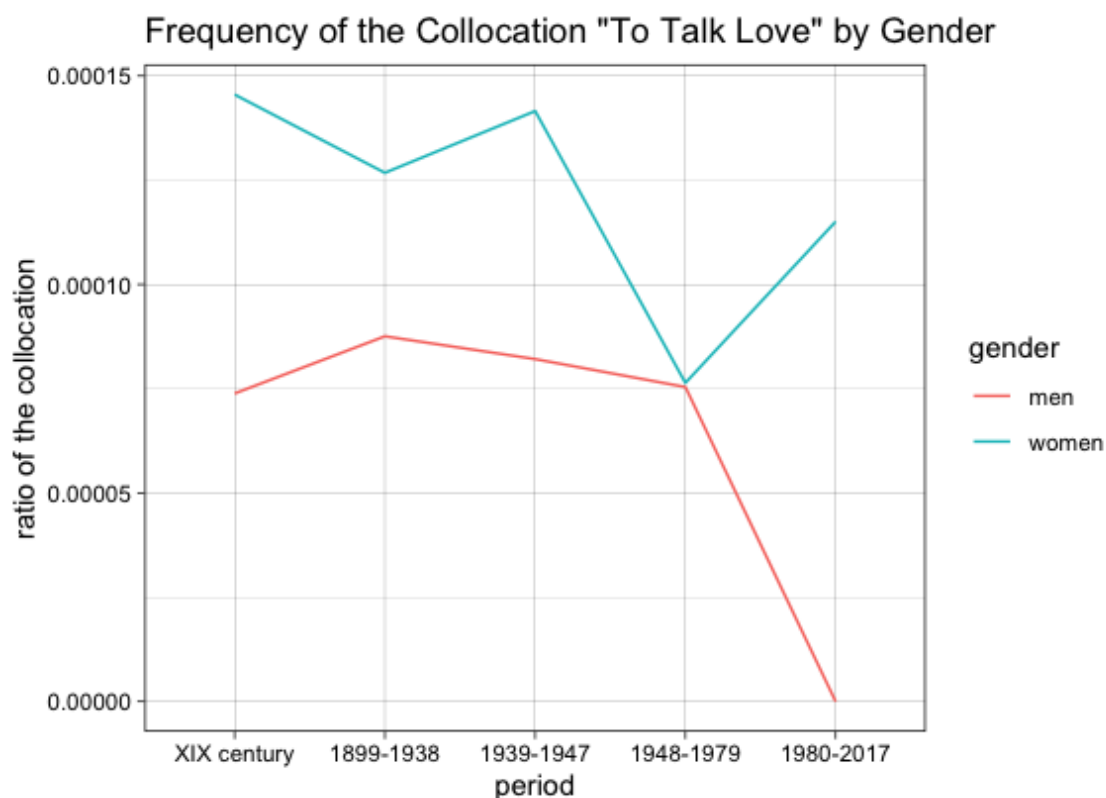
Nevertheless, Example 16 proves that male diarists also used the bigram but more
rarely. The ratio is 0.00007385 in the XIX century and 0.00008756 at the beginning of
the XX century.

(16)   *I bowed down to her with an expression of affection. I kiss her hands, **I talk
about my love** for her. (Ya sklonilsya pered neyu s iz"yavleniem lask. Celuyu ruki,
govoryu o lyubvi svoej k nej.)* [Ivan Yuvachev, August 12, 1892]

We also compared the outcomes with the database. Like in the example of the collocation *heart to be beating*, the collocation *to talk love* saw a significant rise at the beginning of the XX century and gradually decreased afterwards.

Figure 14. Frequency of the collocation *to talk love* (govorit' lyubov') in the database
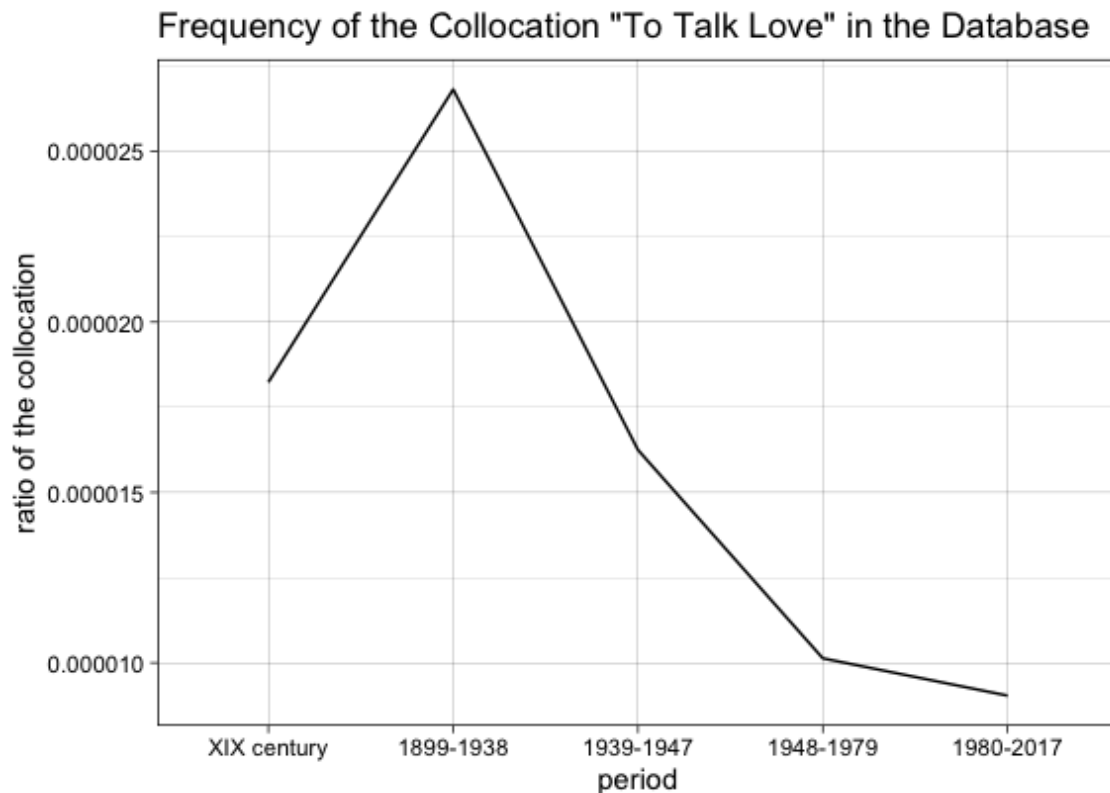
The possible reason of these similar patterns can be seen in language changes; collocations deemed *high-sounding* could gradually get out of use.

4.3.4. Lyubimyj devushka

The ratio of the collocation *beloved girl* (lyubimyj devushka) considerably increased during the Second World War (0.0000382834), as Figure 15 proves.

Figure 15. Frequency of the collocation *beloved girl* (lyubimyj devushka) in the corpus
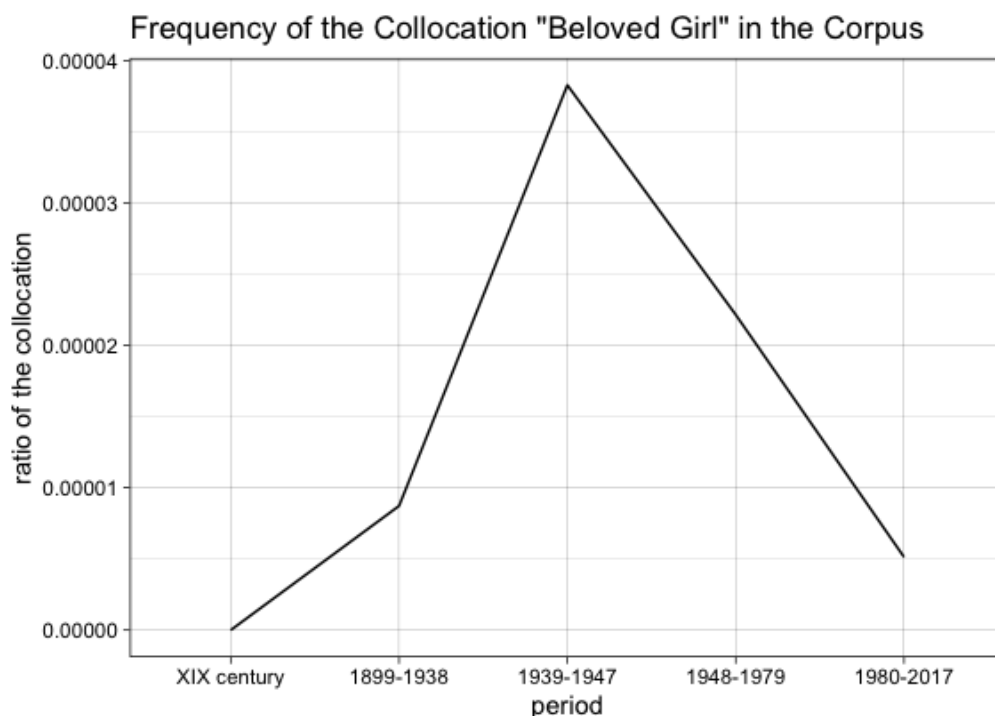
We assume that diarists used the bigram when they kept thinking about their loved ones or described the joy of seeing them. Two remarkable examples are given below.

(17)   *The main purpose of my visit was to have fun. Forget about the surroundings and find quiet peace in the arms of your **beloved girl**. (Osnovnoj cel'yu moego prihoda bylo ─ veselo provesti vremya. Zabyt'sya ot okruzhayuschego i najti tihij pokoj v ob"yatiyah **lyubimoj devushki**.)* [Dmitry Bespalov, October 2, 1945]

(18)   *It's good that the future is empty; after all, I'm not jealous of Taska, and if my **beloved girl** wrote so rarely, it would be a torment. (Horosho, chto buduschee pusto; ved' Tas'ku ya ne revnuyu, a esli by tak redko pisala **lyubimaya devushka**, eto bylo by muchen'e.)* [Yuri Ryabov, December 11, 1942]

4.3.5. Nikto nuzhnyj

The collocation *to need nobody* (*nikto nuzhnyj*) became noticeable during the second half of the XX century. The collocation was used in two possible contexts. Firstly, diarists could express someone's unwillingness to feel love (Example 19). As Example 20 shows, diarists could convey their disappointment and resentment after

encountering the lack of reciprocity. Secondly, as in Example 21, diarists could emphasize that they didn't need anybody but their lover.

(19) *And he says he **doesn't need anyone**. Why did he tame me? (A on govorit, chto emu **nikto ne nuzhen**. Zachem zhe on menya priruchal?)* [Galina Larskaya, May 8, 1966]

(20) *I'm tired. **I don't need anyone**. (Ya ustala. Mne **nikto ne nuzhen**.)* [Olga Kovaleva, December 8, 1983]

(21) *Probably, he will keep me for a long time and **no one will be needed** except him. (Navernoe, on eshe dolgo budet derzhat' menya i **nikto ne nuzhen** budet, krome nego.)* [Olga Kovaleva, December 27, 1983]

4.3.6. Conclusions

Analyzing collocations turned out to be a productive method of retrieving patterns. The frequency of bigrams compared over time periods can shed some light on hidden patterns. The examples given show that the increase in the frequency of a certain bigram can be explained by various reasons.

Firstly, the frequency of the collocation can reflect a custom or an act of decency related to both private and public displays of affection common in a certain period. The collocations *to take hand* and *to kiss hand* are the examples of this case. In several diaries, it is described that taking a lover's hand could be an act of expressing love. According to Zorin, it can be considered as *a public mode of feeling* within a certain emotional matrix. Possibly, this public act was common among particular social groups in particular historical periods—in the XIX century and, probably, earlier.

Secondly, the usage of the collocation can be connected with historical events, as the collocation *beloved girl* proves. Being away from home, military people were thinking about their loved ones more often.

What's more, collocations can also reflect the language and its changing characteristics. The case of the collocation *to talk love* makes the shift happening in meaning of the words explicit. In the XIX century and in the first half of the XX century, the collocation *to talk love* was more often used in the meaning of the expression *to declare love*. Subsequently, the collocation occurred more rarely.

Gender differences in bigrams' use should be also taken into account. Noteworthy, the collocations *to talk love* and *to take hand* occurred more frequently in the records written by female diarists. Nevertheless, examples found in records show that female diarists mostly described their male partners' actions with these words.

Another insight into the origin of patterns was provided by the collocation *heart to be beating*. As the example of Maria Bashkirtseva's diary shows, patterns can be literary-based, and this reference may be both comprehended and clearly indicated by the diarists themselves.

To sum up, examining collocations is fruitful in combination with close reading. Even if collocations are identical, the context can vary greatly. Close reading seems to be the simplest option to notice the difference, which is extremely significant for pattern retrievement.

4.4. Morphological markup

With collocations counted, we could only scratch the surface of the patterns in love-related records. It was possible to define patterns related to history and linguistic process, but culturally-based tendencies mostly remained hidden.

In order to reach deeper patterns, we applied instruments for morphological and syntactic tagging. We parsed data with UDPipe library instruments in R and selected co-occurrences with nouns and verbs found within a sentence. As a result, we got several tables with co-occurrences and their relative frequency depending on a period. Among all the verb co-occurrences, we selected ones with the verb *to love*. The table with the 100 most frequent co-occurrences with this verb is given in the Appendix. The sample of this table can be found below.

Table 5. Sample of the table with co-occurrences

	XIX century	1899-1938	1939-1947	1948-1979	1980-2017
lyubit' tak	0.00096189	0.00097995	0.00089596	0.0011087	0.00087885
lyubit' ochen'	0.00075147	0.0012491	0.0007712	0.0014205	0.00074865
lyubit' chelovek	0.00060118	0.00044857	0.00047633	0.00069294	0.00048825
lyubit' drug	0.00042082	0.00026914	0.00039694	0.0002945	0.00030922
lyubit' potomu	0.00030059	0.00021393	0.00014744	0.00019056	0.00009765

4.4.1. Frequent co-occurrences over time periods

Some of the resulting bigrams occurred with comparatively similar frequency in different periods. The co-occurrences *to love truly*, *to love passionately*, *to love sincerely* are among them.

Figure 16. Frequency of the co-occurrences *to love truly* (lyubit' istinno), *to love passionately* (lyubit' strastno), *to love sincerely* (lyubit' iskrenne)
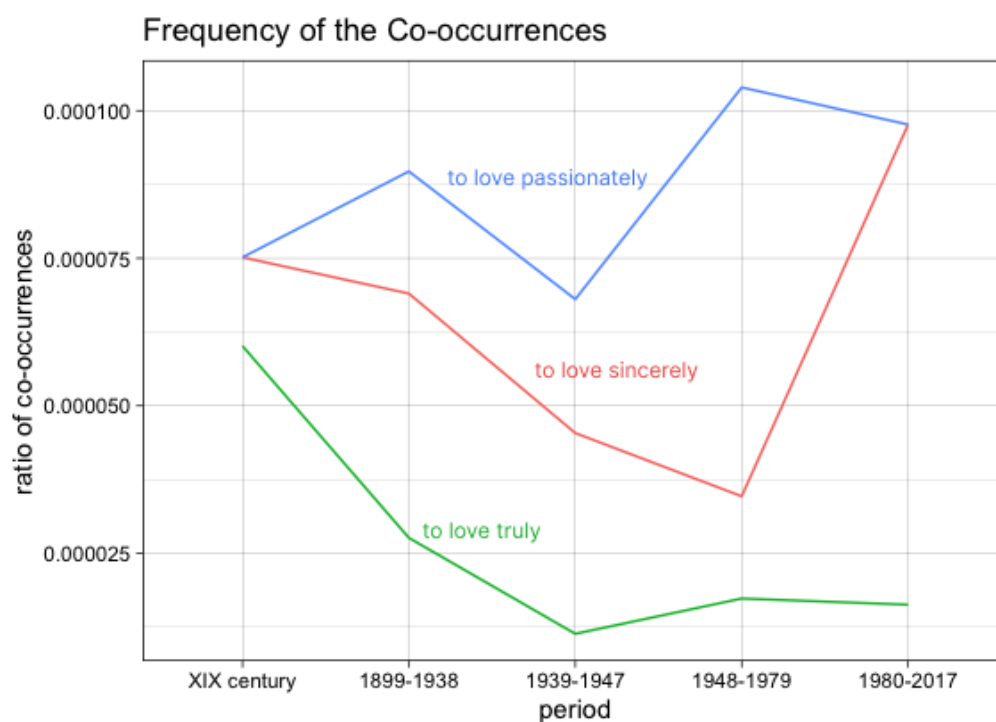
However, some of the co-occurrences saw significant changes, which can be seen on Figures 17 and 18.

Figure 17. Frequency of the co-occurrences *to love really* (lyubit' po-nastoyaschemu), *to love tenderly* (lyubit' nezhno)
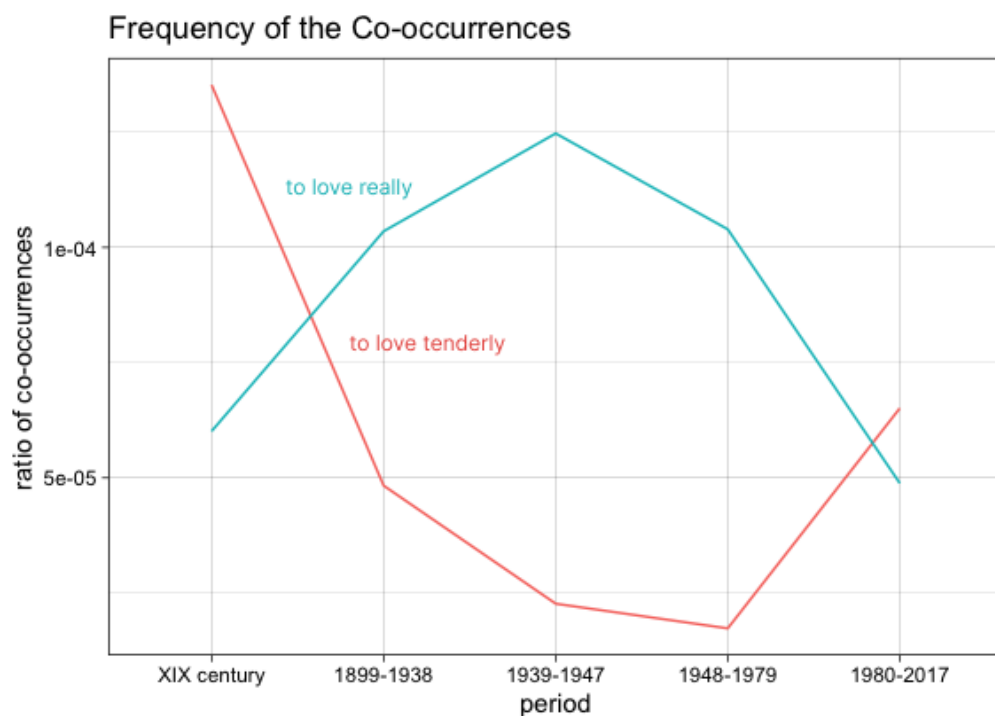
Figure 18. Frequency of the co-occurrences *to love dreadfully* (lyubit' uzhasno), *to love madly* (lyubit' bezumno)
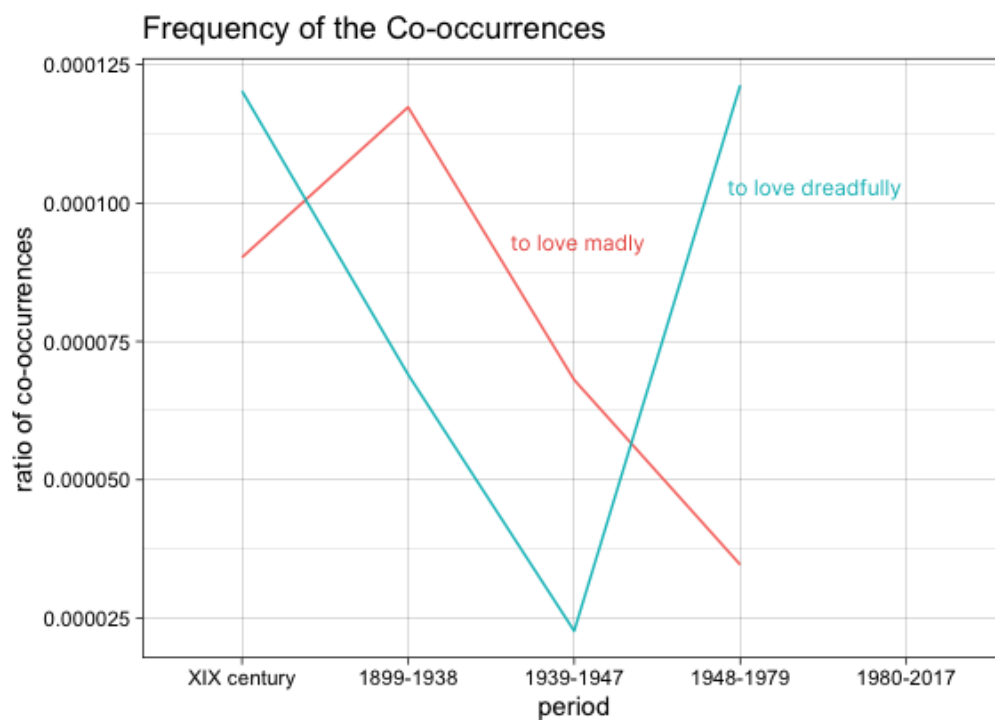
We looked closely at the examples of these co-occurrences in the diaries. Regarding the bigram *to love tenderly*, it occurred mostly in the XIX century. Thereafter, its frequency is gradually decreasing and slightly goes up in the fifth period. Concerning the context, in the XIX century, diarists described their own feelings with this co-occurrence more frequently (Example 22). Thereafter, it was more often used in reference to other people's emotions, as Example 23 shows.

(22)   *Today I **love him tenderly** and devotedly all day. (Segodnya celyj den' **lyublyu ego nezhno** i predanno.)* [Tatiana Sukhotina-Tolstaya, April 5 (March 24), 1897]

(23)   *A man with the skin of a rhinoceros knew how and could **love tenderly**, passionately and selflessly. (Chelovek s kozhej nosoroga umel i mog **lyubit' nezhno**, strastno i samozabvenno.)* [Tatiana Doronina, December 7, 1984]

As regards the bigram *to love madly*, it saw a dramatic growth at the beginning of the XX century (Example 24). Thereafter, it was far less popular. Furthermore, there were no examples of the co-occurrence in 1980–2017.

(24)   *I love you, I **love you madly**, it's a strange love that you don't understand. (Ya lyublyu tebya, **lyublyu bezumno**, eto strannaya lyubov', kotoroj ty ne ponimaesh'.)* [Ernestine Edelman, May 10, 1929]

The way how the frequency of the co-occurrences *to love dreadfully* and *to love really* changed during the Second World War is noteworthy. The frequency of the bigram *to love dreadfully* plummeted while the bigram *to love really* increased. The only example of the bigram *to love dreadfully* is ironic, which Example 25 shows.

(25)   *Apparently, Andronikov is **dreadfully in love** with himself — he can't stop, he keeps talking and talking about himself. (Po-vidimomu, v sebya Andronikov **vlyublen uzhasno** — ne mozhet ostanovit'sya, vse govorit o sebe i govorit.)* [Vsevolod Ivanov, November 5, 1942]

Along with the co-occurrence *to love really*, the bigram *real love* rocketed during the Second World War, as Figure 19 shows.

Figure 19. Frequency of the co-occurrence *real love* (nastoyaschaya lyubov')
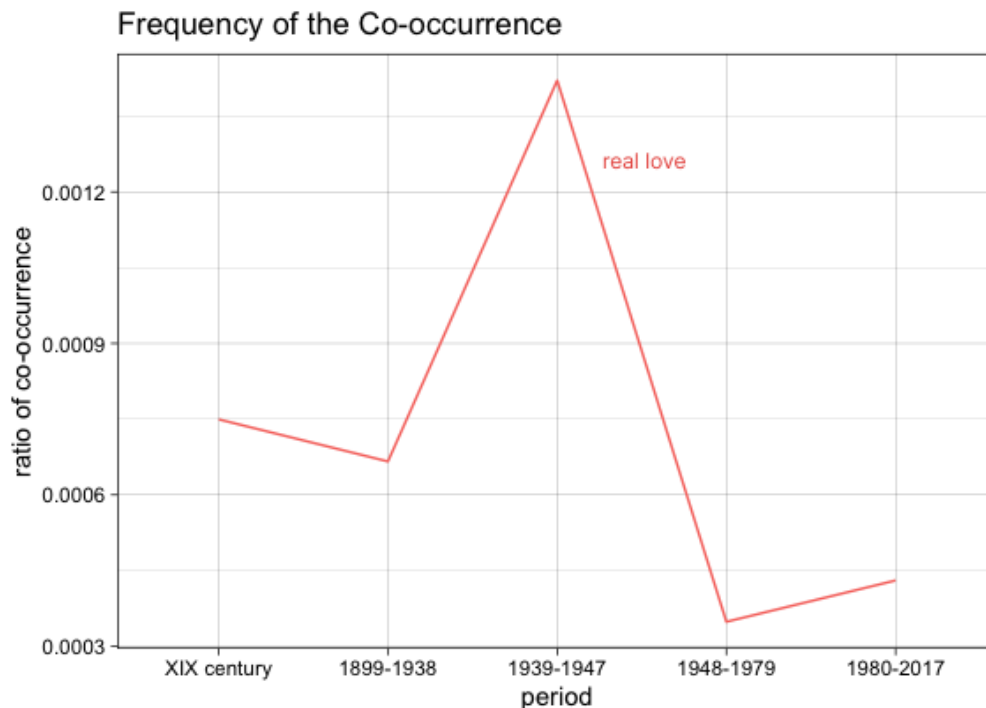
(26)   *That day I was drunk, as on the day before, from vodka, but I did not love Klava, as one should love with **real love**, but only enjoyed her tenderness. (V etot den' ya byl p'yan, kak i nakanune, ot vodki, odnako ya ne lyubil Klavu, kak dolzhno lyubit' **nastoyaschej lyubov'yu**, a tol'ko naslazhdalsya ee nezhnost'yu.)* [Vladimir Gelfand, October 19, 1944]

In conclusion, it still seems vague why certain co-occurrences become popular in a particular period. Nevertheless, the plummeting of several bigrams during the Second World War is conspicuous. The possible reason could be that certain emotional words like *dreadfully* and *madly* become inappropriate for use in the love-related context. However, this fact does not negate the importance of *real love* for the war period, which is indicated by the increase of certain co-occurrences.

In addition to comparing individual co-occurrences, we took a closer look at expressions related to one theme or metaphor. We will examine a few of them below.

4.4.2. Lunnyj svet / lunnaya noch'

In the table of co-occurrences, several frequent bigrams related to the moon (*moonlight night*, *moon light*) caught our attention. We examined the records with these bigrams closely.

First of all, it should be mentioned that literature, poetry, and art are usually rife with moon imagery and descriptions of moonlight walks. There is no wonder that diaries reflect this cultural tradition as well.

In Example 27, poet and diarist Semyon Nadson clearly pointed out this tradition highlighting that these patterns are "never found in real life." It should be also stressed that Nadson was 13 at that time.

(27)   *From this stupid reading, my imagination began to draw mysterious adventures everywhere, **moonlight walks** and similar horrors, which, fortunately, are never or almost never found in real life. (Ot etogo glupogo chteniya voobrazhenie moe vezde nachalo risovat' tainstvennye priklyucheniya, **progulki pri lune** i tomu podobnye uzhasy, kotorye, k schast'yu, nikogda ili pochti nikogda ne vstrechayutsya v dejstvitel'noj zhizni.)* [Semyon Nadson, March 3 (February 20), 1876]

Nevertheless, during the first half of the XX century, there are numerous examples of the proverbial moonlight walks in love-related records.

(28)   *We went home together with Misha. We went out into the **moonlit** snow garden. We were all talking about the same thing. He accompanied me to the Sennaya. (Poshli domoj vmeste s Mishem. Vyshli v **lunnyj** snezhnyj sad. Govorili vse o tom zhe. On provodil menya do Sennoj.)* [Ada Onoshkovich-Yatsyna, February 20, 1921]

(29)   *In the evening, when **the huge moon was rising**, we walked, and, going over the experiences, we perpetuated our love. (Vecherom **pri voskhodyaschej ogromnoj lune** gulyali, i, perebiraya perezhitoe, zakreplyali nashu lyubov'.)* [Mikhail Prishvin, May 21, 1940]

Remarkably, several diarists pointed out that the moonlight night was associated with romantic moonlight walks to them.

(30)   *Yes, it was a **beautiful white night**. On such a night, it would be nice, warmly dressed, to walk through such a forest with your beloved. (Da, byla **belaya prekrasnaya***

noch'. V takuyu noch' horosho by, teplo odevshis', progulyat'sya po takomu lesu s lyubimoj.) [Anatoly Dzyakovich, December 5, 1943]

(31)   *And the weather today is so wonderful: a little frost, snow and a **full moon is shining**. [...] I want, I passionately want a loving and beloved person to walk next to me and lead me by the arm. (I pogoda segodnya takaya chudnaya: nebol'shoj moroz, sneg i **svetit polnaya luna**. [...] Hochetsya, strastno hochetsya chtoby s toboj ryadom shel i vel tebya pod ruku lyubyaschij tebya i lyubimyj chelovek.)* [Rosalia Serdnak, January 8, 1944]

4.4.3. Luch sveta

Metaphor of light was proven to be another widespread love-related imagery. In diaries, the beginning of a new romance is compared to a ray of light. The feeling of love resembles the light coming from within. The diarist's everyday life is illuminated by a sudden ray of love.

Noteworthy, the phrases *ray of light* and *ray of love* occur in records of almost all periods in data. Below we give examples from different periods.

(32)   *After three years of failures in all my hopes [...] it was the first joy that penetrated into my soul, **the first ray of light** that illuminated the darkness that clothed it. (Posle trekhletnih neudach vo vsekh moih nadezhdah [...] eto byla pervaya radost', pronikshaya v dushu moyu, pervyj **luch sveta**, ozarivshij mrak, kotoryj obleg ee.)* [Alexey Wolf, August 30 (18), 1831]

(33)   *From the first meeting, immediately some of your **rays of light** from you entered me. (S pervoj vstrechi, srazu kakie-to tvoi **luchi sveta** ot tebya voshli v menya.)* [Bronislava Nijinskaya, January 11 (December 29), 1920]

(34)   *Hugo is gone for me. In my life, he is like a wonderful **ray of light**. (Ugo propal dlya menya. V moej zhizni on kak chudesnyj **luch sveta**.)* [Alexandra Mikhaleva, September 7, 1945]

(35)   *After all, first of all, **love should shine** from a person, and first of all on the chosen ones, and then on the rest. (Ved' prezhde vsego iz cheloveka **dolzhna svetit' lyubov'**, i prezhde vsego na izbrannyh, a potom na ostal'nyh.)* [Valery Milovatsky, September 22, 1996]

In the examples above, life without love is compared to darkness or mundaneness. Therefore, the beginning of a new romance is extraordinary. This metaphor can be described with a formula "a ray of light in the dark realm" (luch sveta v tyomnom tsarstve) from the eponymous article by Nikolay Dobrolyubov if you take it out of context. It cannot be stated with certainty that this pattern is somehow connected to Dobrolyubov's magnum opus. Nevertheless, the name of Dobrolyubov's work highlights the juxtaposition of light and darkness. This contrast is implied in the records above.

4.4.4. Conclusions

Due to morphological markup with UDPipe instruments, we managed to have a glimpse of patterns that were harder to retrieve. Two of them associated with phrases *moonlight* and *ray of light* have possible cultural origin. It seems undoable to track the exact original source of these patterns because they are ubiquitous and deeply rooted in literature and art. Nevertheless, it is remarkable that diarists got inspired by classic literature and took imagery and examples from it.

4.5. Close reading

It was challenging to retrieve patterns from the cultural realm with methods based on frequency. The reason is that patterns can be represented not only with repeating phrases and words. They can also be distributed throughout the text, which can be hard to track with computable methods.

For this reason, we also examined the corpus with the close reading method applying those insights we got on previous steps.

4.5.1. Feeling of real love

With close reading, we found various noteworthy examples of the repeating pattern when diarists described their feelings for their lover like something they had never experienced before. The new feeling was called *real love* and compared to *something you read of in books*. Previous relationships were inferior to these new emotions.

(36)   *Until that time, **I did not love anyone** and I do not know, strictly speaking, what love is, I do not know the feelings that lovers experience and which I could only judge by hearsay, **according to the book by Turgenev, L. Tolstoy and other authors**. (Do*

etogo vremeni ya nikogo ne lyubil i ne znayu, sobstvenno govorya, chto takoe lyubov',
ne znayu tekh chuvstv, kotorye ispytyvayut vlyublennye i o kotoryh ya mog sudit' tol'ko
ponaslyshke, po knige Turgeneva, L. Tolstogo i drugih avtorov.) [Vasily Trushkin,
February 14, 1938]

(37)   ***I didn't really love anyone***, *but "now, it seems, my hour has come."* (***Ya ne***
lyubila nikogo po-nastoyaschemu, *no «nyne, kazhetsya, moj chas nastal.»*) [Rosalia
Serdnak, January 8, 1944]

We investigated whether this comparison was given more frequently in a
particular period. We calculated the proportion of relevant records to all records in the
period. It turned out that during the Second World War there were found more relevant
records than in the other periods. There were 0.0046 relevant records, while at the
beginning of the XX century and in the post-war period the proportion was 0.0014 and
0.0032 respectively.

These findings correlate with the analysis of co-occurrences: during the Second
World War, there was the rise of the bigrams *love really* and *real love*. It follows that in
the situation of war people could re-evaluate what was important. Therefore, the feeling
of *real love* could become even more valuable.

4.5.2. Concept of real love

In order to delve deeper into the concept of real love, we closely examined the
records with the bigram *real love*. It turned out that the idea of real love that happens
only once was especially common—it was found in the XIX century and in the first half
of the XX century. Some of the examples are given below.

(38)   *Real love may not be eternal, but **it can only be once in a lifetime**.*
(*Nastoyaschaya lyubov' mozhet i ne byt' vechnoj, no ona mozhet byt' **tol'ko odin raz v***
***zhizni**.*) [Maria Bashkirtseva, January 13 (1), 1881]

(39)   *After all, a human loves with real love **only ONCE**… (*Ved' nastoyaschej*
*lyubov'yu chelovek lyubit **tol'ko ODIN raz**…*) [Galina Zaitseva, March 3, 1930]

While in the XIX century real love was more frequently described as a first fresh
experience (Example 40), in the postwar period we found a record in which real love,
on the contrary, was considered as a mature feeling. This record is given in Example 41.

(40)   *Each of them is capable of falling in love, but he cannot fall in love as for the first time in his life, because a person has already lost **the freshness and untouchability of feeling**. (Vsyakij iz nih sposoben vlyubit'sya, no polyubit' tak, kak v pervyj raz v zhizni — on ne mozhet, ibo chelovek poteryal uzhe **svezhest' i netronutost' chuvstva**.)* [Elizaveta Diakonova, February 16 (4), 1894]

(41)   *Probably, real love, real closeness of people **comes with the years lived**, with disappointments and charms lived together, with grief and joy lived together. (Navernoe, nastoyaschaya lyubov', nastoyaschaya blizost' lyudej **prihodit s prozhitymi godami**, s prozhitymi vmeste razocharovaniyami i ocharovaniyami, s prozhitym vmeste gorem i radost'yu.)* [Alexander Chugunov, 1969]

During the Second World War and in other following periods the idea of the only real love lost its popularity. Real love was defined with other words—it *overcomes suffering*, *is incompatible with jealousy*, and *allows you to love your enemies*.

4.5.3. Ideas of love

We particularly examined the corpus in order to find diarists' expectations and ideas of love that face reality. We tried to find signs of possible cultural and emotional matrices and their origin.

The pattern that we managed to discover was based on the expectations of *higher love* associated with literature on knights and princes. It is hard to determine the exact literary source of these notions. Nevertheless, this reference goes through different periods and remains a common place for describing expectations of love.

Records with this pattern are seen mostly in female diaries. There are several examples of comparing lovers to *faithful knights* in female records. Some of them are given below.

(42)   *[...] when he heard my question, **like a faithful knight**, he ran to look at the sundial. ([...] uslyshav vopros moj, **kak vernoj rycar'**, pobezhal smotret' na solnechnye chasy.)* [Anna Andro, October 12 (September 30), 1828]

(43)   *But today, yesterday, the third day, all thoughts are busy with the fact that **my faithful knight** is cheating. (A vot segodnya, vchera, tret'ego dnya vse mysli pogloscheny tem, chto **moj vernyj rycar'** izmenyaet.)* [Lyudmila Dayanova, January 5 (December 23), 1908]

Knights and princes are also associated with the idea of *real love*. Expectations inevitably meet reality, and comparing reality to the idea, diarists tried to figure out what *real love* is.

(44)   *I was looking for a heart in a dry and selfish person, I took a vulgar impulse for true love. Thoughts of the **ideal love of knights** flash through my head, and it seems to me that this is worship, adoration. (Ya iskala serdce v cheloveke suhom i egoistichnom, prinimala poshlyj poryv za istinnuyu lyubov'. V moej golove pronosyatsya mysli ob **ideal'noj lyubvi rycarej**, i mne kazhetsya, chto eto bogotvorenie, obozhanie.)* [Lyudmila Dayanova, May 5 (April 22), 1908]

(45)   *As a child, I dreamed of **"my prince."** [...] **The fairy-tale dreams** of childhood and the fairy tale of my love disappeared like a ghost. (V detstve ya mechtala «o moem prince». [...] Skazochnye sny detstva i skazka moej lyubvi ischezli kak prizrak.)* [Aleksandra Mikhaleva, June 2, 1946]

(46)   *And so, **like in a fairy tale, like a fairy prince**, He appeared. [...] But then for some reason we began to move away. [...] Maybe it wasn't love? (I vot, **kak v skazke, kak skazochnyj princ**, poyavilsya On. [...] No potom pochemu-to my stali otdalyat'sya. [...] Mozhet byt' eto ne byla lyubov'?)* [Nina Kravchenko, December 22, 1949]

Remarkably, diarists could even point out that their partners violated the rules of behavior prescribed by the cultural matrix of a diarist (Example 47).

(47)   *[...] my **"faithful knight"**, sprawled vis-a-vis on the couch, yawned every minute and uttered some incomprehensible sounds like mooing to my attempts to talk. ([...] moj «**vernyj rycar'**», razvalyas' vis-a-vis na divane, pominutno zeval i na moi popytki razgovorit'sya proiznosil kakie-to neponyatnye zvuki napodobie mychaniya.)* [Lyudmila Dayanova, February 18 (5), 1908]

Examples given above show that literature can be a fruitful source for role models and expectations of love. Nevertheless, in male records, similar comparisons of reality and literature are given as well, as Example 48 proves.

(48)   *Is this love? There was nothing sublime, romantic. Something thievish, unclean, even shameful. Only **in novels they write about love**, although what do I know about it… (I eto lyubov'? Nichego vozvyshennogo, romanticheskogo ne okazalos'. Chto-to*

*vorovskoe, nechistoe, dazhe postydnoe. Tol'ko **v romanah pishut o lyubvi**, hotya chto ya ob etom znayu…*) [Leonid Fialkovsky, January 7, 1943]

In addition to general literary references, diarists gave specific examples of literary characters or narratives that resembled real life to them. Below we examine several specific examples of referencing literature in love-related records.

4.5.4. Literary symbols of love

It was very common among the diarists to give comparisons of one's own feelings with imagery from classic literature, as Examples 49 and 50 show.

(49)   *[…] I accepted her **as the divine Beatrice, as Laura**. (*[…] ya prinyal ee **kak bozhestvennuyu Beatriche, kak Lauru**.*) [Erkin Zholdasov, 2006]

(50)   *Sometimes in the presence of Volodya I feel **Blok's heartbeat**. (Inogda v prisutstvii Volodi chuvstvuyu **blokovskoe serdcebienie**.*) [Galina Larskaya, 1972]

Among all comparisons our attention was drawn to the repeating reference to Dulcinea del Toboso, a character from Miguel de Cervantes' novel *Don Quixote*. This reference occurred in several records written by different diarists in different periods.

(51)   *Meanwhile, I'm in love with **"Dulcinea"**, it seems! (Mezhdu tem, ya vlyublen v **«Dul'cineyu»**, kazhetsya!*) [Andrey Petrov, August 23 (11), 1899]

(52)   *The tragedy of Don Quixote is terrible, but beautiful. And if this tragedy is turned inside out? If we make it so that there was no Don Quixote Knight, but he was invented by the amorous **Dulcinea**. (Tragediya Don Kihota strashna, no prekrasna. A esli etu tragediyu vyvernut' naiznanku? Esli sdelat' tak, chto Don Kihota — Rycarya ne bylo, a ego vydumala vlyublennaya **Dul'cineya**.*) [Sofia Ostrovskaya, 1941]

Dulcinea symbolizes a beloved person who is endowed with imaginary qualities being *invented* by his or her lover.

Example 53 reveals the symbolic potential of the pattern with Dulcinea.

(53)   *When I woke up, I thought about how, with my complete love for Lala, I could dream of an Inaccessible one. And so I realized that that love and this is the same process: at that time I had **to make a Dulcinea out of Aldonza** in order to perform feats. (Prosnuvshis', dumal o tom, kak zhe eto pri moej polnoj lyubvi k Lyale mozhet mne snit'sya Nedostupnaya. I tak ponyal, chto ta lyubov' i eto est' odin i tot zhe process:*

*mne togda nuzhno bylo **iz Al'donsy sdelat' Dul'cineyu**, chtoby sovershat' podvigi.*)
[Mikhail Prishvin, August 28, 1945]

The juxtaposition of Dulcinea and Aldonza seen in Prishvin's diary reflects the contrast of Aphrodite Urania and Aphrodite Pandemos in Plato's dialogue *The Symposium*. These Aphrodites embody the distinction between a nobler and a baser kind of love. As Example 54 shows, analogous distinction is personified in Dulcinea and Aldonza.

(54)   *When she handed me a letter to my parents and asked me to read it [...], I suddenly became cold, confused, for a moment I saw her **as a very ordinary being**, the bandage suddenly fell off… It lasted for a moment, but she apparently understood me and soon took the letter back. And as soon as she did that and **became unavailable** again, I again began a crazy **affair with Aldonza, who must be Dulcinea**. (Kogda ona peredala mne pis'mo k roditelyam i poprosila prochest' ego [...], ya vdrug ohladel, smutilsya, na mgnovenie uvidel ya ee kak ochen' **obyknovennoe suschestvo**, povyazka vdrug upala… Eto prodolzhalos' kakoe-to mgnoven'e, no ona, po-vidimomu, ponyala menya i vskore vzyala pis'mo obratno. I kak tol'ko ona eto sdelala i snova **stala nedostupnoj**, ya opyat' nachal bezumnyj **roman s Al'donsoj, obyazannoj byt' Dul'cineej**.*) [Mikhail Prishvin, April 28, 1947]

This is not the only example when a literary character symbolizes a particular aspect of love. Tatiana Larina from Pushkin's *Eugene Onegin* is referenced as a symbol of real and eternal love in diaries. There are a few examples below.

(55)   *It [creative tension] remains forever old and forever new, like Tatiana's love for Onegin. (Ono [tvorcheskoe napryazhenie] ostaetsya vechno starym i vechno novym, kak lyubov' Tat'yany k Oneginu.*) [Andrey Burov, August 2, 1943]

(56)   *But I probably don't really love you. Tatiana, for example, did not think about what would happen if Onegin did not love her. (No ya, navernoe, ne po-nastoyaschemu lyublyu tebya. Vot Tat'yana, naprimer, ne dumala o tom, chto budet, esli Onegin eyo ne lyubit.*) [Lydia Barlas, December 23, 1958]

4.5.5. Conclusions

With close reading, we examined diarists' expectations and ideas of love and their origin. It turned out that literature was a fruitful source of ideas and patterns that diarists compared to real life.

Literary characters and narratives serve as an ideal model of love that diarists can compare their feelings with. On the one hand, examples with knights and princes show that such expectations can hardly pass the reality check. On the other hand, literary imagery and narratives can help people to comprehend what is happening and build their own theories of love, as the Dulcinea case illustrates.

Also, implicit origins of the *real love* ideas should be examined and revealed in the future. Literary and biblical sources may be on the surface, but the exact provenance of these patterns is to be explored.

5. Conclusion

In our research, we examined the corpus of love-related records with several methods in order to retrieve patterns of various origin. At first, we applied topic modeling that proved to be helpful in examining diaries before. However, with topic modeling, we only managed to observe the peculiarities of corpus data and some historical features. The ratio of the words connected with love and emotions was relatively small in topics. Thus, topic modeling turned out to be unsuitable for detecting love-related patterns.

At the next stage, we extracted the most frequent collocations and compared them depending on the period. We closely examined several examples and observed the rise and fall of certain collocations in different periods. Reasons for the change in collocation frequency include transformation in traditional practices (collocation *to take hand*), language change (collocation *to talk love*), and historical background (collocation *beloved girl*).

What's more, we observed gender patterns in using particular collocations by female and male diarists. Most of the given examples belonged to female diarists. Nevertheless, these examples cannot be reduced to any single pattern. Regarding the collocation *to talk love*, it was mostly used to describe male partners' actions. In case of the collocation *heart to be beating*, it was used to refer to the diarist herself.

Considering the above mentioned examples, the method of collocations retrieval allowed us to take a glimpse at hidden patterns of various origin. However, culturally-based tendencies were still unrevealed. Thus, at the next stage, we practiced morphological markup with UDPipe instruments.

The main outcome of the morphological analysis was the observed rise in frequency of co-occurrences *to love really* and *real love* during the Second World War. The value of real love is emphasized in Example 57.

(57)   *Only talkers can claim that there can be no **real love** at the front. The poetry of this love will still be sung by poets subtly and deeply. (Tol'ko boltuny mogut utverzhdat', chto na fronte ne mozhet byt' **nastoyaschej lyubvi**. Poeziyu etoj lyubvi esche vospoyut poety tonko i gluboko.*) [Tatiana Atabek, March 18, 1945]

Also, with morphological markup, we managed to detect patterns distributed throughout the text. We analyzed imagery entrenched in literature and art that appeared in diaries as well.

Nevertheless, deeper research would be impossible without close reading. Considering the outcomes we got on previous steps, we examined ideas and aspirations of love and their confrontation with reality. It cannot be confirmed with certainty that these ideas originated from literature and arts, but, as we have seen in several examples, literature can be a fruitful source of ideas and patterns for diarists.

To summarize, it can be stated that computable methods are better in handling historical specificities than emotional patterns. Nonetheless, we managed to gain some insights on patterns' specifics and origin. These outlines can be subsequently deepened and thoroughly studied. For instance, profound comparison of patterns to literary models can be conducted in future.

What's more, the concept of cultural and emotional matrices can be more thoroughly elaborated. It can be explored how emotional matrices undergo changes over time and what are the reasons for breaking them. In our study, we just scratched the surface of it. Nevertheless, we encountered several remarkable examples of breaking the matrices due to dramatic historical changes, as Example 58 illustrates.

(58)   *There are no knights. And people don't know how to love royally either. All the beautiful poses and beautiful words **turned out to be adapted only to peacetime**, even if you look at these poses and words **only as a theater**. [...] How funny and wild it is to remember now — today! — such solemn and beautiful words!* (*Rycarej ne suschestvuet. I po-korolevski lyudi ne umeyut lyubit' tozhe. Vse prekrasnye pozy i prekrasnye slova* **okazalis' prisposoblennymi lish' k mirnomu vremeni***, esli dazhe smotret' na eti pozy i slova* **tol'ko kak na teatr***. [...] Kak smeshno i diko vspominat' teper' — vot segodnya! — takie torzhestvennye i krasivye slova!* [Sofia Ostrovskaya, 1941]

Another possible benefit of the current study is to provide features for automatic classification of texts and diary entries in particular. As we observed on the available data, records of different periods included various *high-sounding* words that diarists described their feelings with. Below are the examples with the word *awe*.

(59)   *My face was burning and **I was in awe** fortunately I was good at hiding my excitement.* (*Lico gorelo **ya trepetal** schast'e ya horosho umel skryt' svoe volnenie.*) [Kirill Berezkin, September 10 (August 29), 1849]

(60)   *What was it in my heart?! Some **inexplicable awe** and, at the same time, horror.* (*Chto eto bylo v moem serdce?! Kakoj-to **neob''yasnimyj trepet** i, s tem zhe, uzhas.*) [Irina Peskova, April 21, 1941]

Hence, these and other insights could be taken into account while building models for classification.

6. References

Blevins 2010 — C. Blevins. *Topic modeling Martha Ballard's diary.* www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/. 2010.

Vorobiova 2022 — V. Vorobiova. https://github.com/vicintina/prozhito_diploma. 2022.

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

Ginzburg 1999 — L. Ya. Ginzburg. *O psihologicheskoj proze.* M.: INTRADA, 1999. 415 p.

Usov 2018 — B. Usov. https://github.com/kilomeow/prozhito-tools. 2018.

Langford and West 1999 — R. Langford, R. West. *Introduction: Diaries and Margins. Marginal Voices, Marginal Forms: Diaries in European Literature and History.* Amsterdam: Rodopi. P. 6–21.

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

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

Núñez A. et al 2019 — A. Núñez, M. Gerloff, E. Do Dinh, A. Rapp, P. Gehring. I. Gurevych. A 'wind of change'— shaping public opinion of the Arab Spring using metaphors // *Digital Scholarship in the Humanities* 34 (Supplement 1), 2019. P. 142–149.

Paperno 2004 — I. Paperno. What Can Be Done with Diaries? // *The Russian Review* 63(4), 2004. P. 561–573.

Paperno 2014 — I. Paperno. *"Who, What Am I?" Tolstoy Struggles to Narrate the Self.* Ithaka: Cornell University Press, 2014. 240 p.

Pivovarova et al. 2018 — L. Pivovarova, D. Kormacheva, M.Kopotev. Evaluation of collocation extraction methods for the Russian language // *Quantitative Approaches to the Russian Language*, 2018. P. 37–157.

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

Popkin 2009 — J. Popkin. Philippe Lejeune, Explorer of the Diary. // Ph. Lejeune. *On diary*. Honolulu: University of Hawaii Press, 2009. P. 1–16.

Savkina 2007 — I. L. Savkina. *Razgovor s zerkalom i Zazerkal'em: Avtodokumental'nye zhenskie teksty v russkoj literature pervoj poloviny XIX veka.* M.: Novoe literaturnoe obozrenie, 2007. 416 p.

Valieva 2016 — Y. M. Valieva. Pis'ma M. A. Malyh k P. N. Ariyan. Dnevnik Eriki Edel'man 1929–1931 gg. // *Ezhegodnik Rukopisnogo otdela Pushkinskogo Doma na 2015 god.* Saint Petersburg, 2016. P. 689–748.

Zaretsky 2011 — Y. P. Zaretsky. *Strategii ponimaniya proshlogo: Teoriya, istoriya, istoriografiya.* M.: Novoe literaturnoe obozrenie, 2011. 384 p.

Zorin 2016 — A. L. Zorin. *Poyavlenie geroya. Iz istorii russkoj emocional'noj kul'tury konca XVIII – nachala XIX vv*. M.: Novoe literaturnoe obozrenie, 2016. 568 p.

7. Appendix

Table 1. The 100 most frequent co-occurrences with the verb *to love*

		XIX century	1899-1938	1939-1947	1948-1979	1980-2017
0	lyubit' tak	0.0009618853	0.0009799456	0.0008959557	0.0011087051	0.0008788490
1	lyubit' ochen'	0.0007514729	0.0012490856	0.0007712024	0.0014205284	0.0007486492
2	lyubit' chelovek	0.0006011783	0.0004485667	0.0004763309	0.0006929407	0.0004882495
3	lyubit' drug	0.0004208248	0.0002691400	0.0003969424	0.0002944998	0.0003092247
4	lyubit' potomu	0.0003005892	0.0002139318	0.0001474358	0.0001905587	0.0000976499
5	lyubit' sil'no	0.0003005892	0.0002139318	0.0003515776	0.0001732352	0.0001464748
6	lyubit' bol'she	0.0002855597	0.0003381502	0.0002608479	0.0002252057	0.0001464748
7	lyubit' nikogda	0.0002104124	0.0001311195	0.0001701182	0.0002252057	0.0000650999
8	lyubit' vsegda	0.0002104124	0.0001449215	0.0000907297	0.0003118233	0.0002115748
9	lyubit' dusha	0.0001953830	0.0000897133	0.0000793885	0.0001905587	0.0000650999
10	lyubit' kak	0.0001953830	0.0001242185	0.0001360945	0.0002078822	0.0001301999
11	lyubit' teper'	0.0001502946	0.0001380205	0.0001247533	0.0000866176	0.0000488249
12	lyubit' bolee	0.0001502946	0.0000069010	0.0000226824	0.0000173235	0.0000162750

13	lyubit' rebenok	0.0001502946	0.0001242185	0.0000680473	0.0000692941	0.0000813749
14	lyubit' sam	0.0001502946	0.0000345051	0.0000907297	0.0000346470	0.0000976499
15	lyubit' dejstvitel'no	0.0001502946	0.0001449215	0.0001360945	0.0001385881	0.0000488249
16	lyubit' malo	0.0001502946	0.0000621092	0.0000226824	0.0000346470	0.0000162750
17	lyubit' lyubov'	0.0001352651	0.0002415359	0.0002495067	0.0002425292	0.0002441247
18	lyubit' nezhno	0.0001352651	0.0000483072	0.0000226824	0.0000173235	0.0000650999
19	lyubit' drugoj	0.0001352651	0.0001311195	0.0000680473	0.0001039411	0.0000162750
20	lyubit' togda	0.0001202357	0.0000207031	0.0000453648	0.0001559117	0.0000488249
21	lyubit' raz	0.0001202357	0.0001035154	0.0000340236	nan	0.0000162750
22	lyubit' zhena	0.0001202357	0.0000414062	0.0000793885	0.0000519706	0.0000650999
23	lyubit' serdce	0.0001202357	0.0000276041	nan	0.0001385881	0.0000162750
24	lyubit' mnogo	0.0001202357	0.0000207031	0.0000680473	0.0000519706	0.0000650999
25	lyubit' pravda	0.0001202357	0.0000276041	0.0000113412	0.0000346470	0.0000488249
26	lyubit' uzhasno	0.0001202357	0.0000690103	0.0000226824	0.0001212646	nan
27	lyubit' prezhde	0.0001052062	0.0000138021	0.0000113412	0.0000346470	nan

#						
28	lyubit' mat'	0.0001052062	0.0000345051	0.0001928006	0.0000866176	0.0000325500
29	lyubit' zhenschina	0.0001052062	0.0001173174	0.0001360945	0.0001385881	0.0001790248
30	lyubit' segodnya	0.0000901767	0.0000069010	0.0000340236	0.0000866176	0.0000325500
31	lyubit' otec	0.0000901767	0.0000621092	0.0000793885	0.0001039411	0.0000162750
32	lyubit' bezumno	0.0000901767	0.0001173174	0.0000680473	0.0000346470	nan
33	lyubit' vremya	0.0000901767	0.0000621092	0.0000680473	0.0000866176	0.0000162750
34	lyubit' papa	0.0000901767	0.0000414062	0.0000453648	0.0000692941	nan
35	lyubit' sila	0.0000751473	0.0000207031	0.0000113412	0.0000173235	nan
36	lyubit' muzh	0.0000751473	0.0000621092	0.0000793885	0.0000173235	nan
37	lyubit' iskrenne	0.0000751473	0.0000690103	0.0000453648	0.0000346470	0.0000976499
38	lyubit' zhizn'	0.0000751473	0.0001932287	0.0003515776	0.0002425292	0.0002441247
39	lyubit' tam	0.0000751473	0.0000345051	0.0000567061	0.0000519706	0.0000325500
40	lyubit' strastno	0.0000751473	0.0000897133	0.0000680473	0.0001039411	0.0000976499
41	lyubit' takzhe	0.0000751473	0.0000345051	0.0000113412	0.0000173235	0.0000162750
42	lyubit' horoshij	0.0000751473	0.0000345051	0.0000453648	0.0000173235	0.0000162750

43	lyubit' horosho	0.0000751473	0.0000276041	nan	0.0000173235	nan
44	lyubit' pochemu	0.0000751473	0.0000828123	0.0000680473	0.0001732352	0.0001627498
45	lyubit' lyubimyj	0.0000751473	0.0000552082	nan	0.0000346470	0.0000325500
46	lyubit' um	0.0000601178	0.0000138021	0.0000226824	0.0000173235	nan
47	lyubit' lyubit'	0.0000601178	0.0000207031	0.0000113412	0.0000173235	nan
48	lyubit' esche	0.0000601178	0.0001725256	0.0001247533	0.0000692941	0.0000650999
49	lyubit' sestra	0.0000601178	0.0000138021	0.0000453648	nan	0.0000162750
50	lyubit' tyazhelyj	0.0000601178	0.0000207031	0.0000113412	nan	nan
51	lyubit' goryacho	0.0000601178	0.0000966144	0.0001020709	0.0000519706	nan
52	lyubit' po-nastoyaschemu	0.0000601178	0.0001035154	0.0001247533	0.0001039411	0.0000488249
53	lyubit' telo	0.0000601178	0.0000414062	nan	nan	0.0000162750
54	lyubit' muzyka	0.0000601178	0.0000759113	0.0000567061	0.0001039411	0.0000162750
55	lyubit' otchego	0.0000601178	0.0000069010	0.0000340236	nan	0.0000162750
56	lyubit' gde	0.0000601178	0.0000138021	0.0000113412	0.0000346470	0.0000325500
57	lyubit' narod	0.0000601178	0.0000138021	0.0000340236	0.0001212646	0.0000488249

5 8	lyubit' kogda	0.00006 01178	0.00002 07031	nan	0.0000 346470	nan
5 9	lyubit' istinno	0.00006 01178	0.00002 76041	0.0000 113412	0.0000 173235	0.0000 162750
6 0	lyubit' gotovyj	0.00006 01178	0.00006 21092	nan	0.0000 173235	0.0000 162750
6 1	lyubit' svet	0.00006 01178	0.00006 21092	0.0000 113412	0.0000 346470	0.0000 162750
6 2	lyubit' dolgo	0.00006 01178	0.00004 14062	0.0000 226824	0.0000 173235	nan
6 3	lyubit' slovo	0.00006 01178	0.00004 14062	0.0000 340236	0.0000 519706	0.0000 650999
6 4	lyubit' nel'zya	0.00006 01178	0.00002 07031	nan	nan	0.0000 162750
6 5	lyubit' god	0.00006 01178	0.00006 21092	0.0001 474358	0.0000 692941	0.0000 325500
6 6	vlyubit'sya delo	0.00004 50884	0.00000 69010	0.0000 113412	nan	0.0000 162750
6 7	lyubit' nemnozhko	0.00004 50884	0.00002 76041	0.0000 340236	nan	0.0000 162750
6 8	lyubit' zachem	0.00004 50884	0.00002 76041	0.0000 226824	0.0000 173235	nan
6 9	lyubit' syn	0.00004 50884	0.00002 07031	0.0000 113412	0.0000 346470	0.0000 325500
7 0	lyubit' dom	0.00004 50884	0.00002 07031	0.0000 113412	0.0000 519706	0.0000 162750
7 1	lyubit' dolzhen	0.00004 50884	0.00003 45051	0.0000 680473	0.0000 173235	0.0000 325500
7 2	lyubit' chuvstvo	0.00004 50884	0.00001 38021	0.0000 226824	nan	0.0000 162750

7 3	lyubit' brat	0.00004 50884	0.00002 76041	0.0000 340236	0.0000 173235	nan
7 4	lyubit' rodnoj	0.00004 50884	0.00002 07031	0.0000 113412	0.0000 173235	0.0000 162750
7 5	lyubit' den'	0.00004 50884	0.00006 90103	0.0000 453648	0.0000 519706	0.0000 162750
7 6	lyubit' nezhnyj	0.00004 50884	0.00000 69010	nan	nan	0.0000 162750
7 7	lyubit' veroyatno	0.00004 50884	0.00002 76041	0.0000 113412	nan	0.0000 162750
7 8	lyubit' iskusstvo	0.00004 50884	0.00000 69010	0.0000 226824	0.0000 519706	nan
7 9	lyubit' mama	0.00004 50884	0.00006 21092	0.0000 453648	0.0001 212646	0.0000 488249
8 0	lyubit' znachit	0.00004 50884	0.00008 28123	0.0000 680473	0.0000 346470	nan
8 1	lyubit' potom	0.00004 50884	0.00002 76041	0.0000 226824	0.0000 173235	nan
8 2	lyubit' ser'ezno	0.00004 50884	0.00002 76041	0.0000 113412	0.0000 173235	nan
8 3	lyubit' slishkom	0.00004 50884	0.00006 21092	0.0000 340236	0.0000 519706	0.0000 162750
8 4	lyubit' delo	0.00004 50884	0.00003 45051	0.0001 020709	0.0000 692941	0.0000 325500
8 5	lyubit' krepko	0.00004 50884	0.00012 42185	0.0000 793885	0.0000 346470	nan
8 6	lyubit' gluboko	0.00004 50884	0.00008 97133	0.0000 226824	0.0000 346470	0.0000 325500
8 7	lyubit' vechno	0.00004 50884	0.00002 07031	0.0000 226824	0.0000 173235	0.0000 325500

88	lyubit' lico	0.0000450884	0.0000414062	0.0000113412	0.0000346470	0.0000650999
89	lyubit' osobenno	0.0000300589	0.0000759113	0.0000340236	0.0000866176	0.0000162750
90	polyubit' nikogda	0.0000300589	0.0000069010	0.0000113412	0.0000173235	nan
91	lyubit' sovsem	0.0000300589	0.0000276041	nan	0.0000346470	0.0000325500
92	lyubit' neschastnyj	0.0000300589	0.0000207031	0.0000113412	nan	nan
93	lyubit' kogda-to	0.0000300589	0.0000621092	0.0000113412	0.0000519706	0.0000162750
94	lyubit' tochno	0.0000300589	0.0000138021	nan	nan	0.0000162750
95	lyubit' kak	0.0000300589	0.0000069010	nan	0.0000346470	nan
96	lyubit' voobsche	0.0000300589	0.0000138021	0.0000907297	0.0000866176	0.0000650999
97	lyubit' snova	0.0000300589	0.0000207031	0.0000113412	0.0000173235	nan
98	lyubit' nauka	0.0000300589	0.0000138021	0.0000113412	nan	nan
99	lyubit' priroda	0.0000300589	0.0000483072	0.0000453648	0.0000692941	0.0000162750
100	lyubit' razgovor	0.0000300589	0.0000069010	nan	0.0000173235	nan