

# Thematic Tagging of Literary Fiction: The Case of Early 20<sup>th</sup> Century Russian Short Stories

Tatiana G. Skrebtsova<sup>[0000-0002-7825-1120]</sup>

St. Petersburg State University, 7/9 Universitetskaya emb., 199034 St.Petersburg, Russia  
t.skrebtsova@spbu.ru

**Abstract.** The paper deals with the problem of thematic tagging in works of fiction, Russian short stories of the early 20<sup>th</sup> century (1900-1930) serving as research data. The very concept of discourse theme, or topic, is argued to be fuzzy and ambivalent, all the more so in the case of literary prose. In the present study, theme is conceived as a set of keywords basically (but by no means exhaustively) defining the story's plot. A list of 89 themes was empirically formed, embracing a wide range of topics. A sample corpus of 310 stories was manually tagged, with each story being mapped onto a set of themes. This corpus was divided into three parts corresponding to three periods, 1900-1913, 1914-1922, 1923-1930. These periods of Russian history being radically different, the stories' content varies greatly, too, not only in what concerns political and social themes, but also in quite personal and mundane matters. The paper traces the themes' frequency rates across the three periods, accounting for their change dynamics in terms of sociopolitical context. The sample corpus will be further used as training data in devising computational techniques for automated thematic tagging of literary fiction.

**Keywords:** literary corpus, thematic tagging, discourse theme, Russian literature, short stories

## Introduction

Thematic tagging is an essential kind of annotation used in digital libraries and large national corpora. Thus, in Russian National Corpus non-fiction texts are assigned tags roughly corresponding to manifold spheres of human life, e.g. *politics, economics, law, religion, medicine, business, art*, etc. Fiction texts, however, are provided solely with tags indicating their genre and chronotope.

This is by no means accidental. The concept of topic, or theme, in literary prose is an under-studied and problematic issue. The lack of a practicable definition has a negative effect on developing literary text corpora. In large-scale corpora, texts have to be thematically annotated but the very principles guiding such annotation are not clear. There is no agreement on what constitutes a theme in works of fiction, or whether a text should have a single theme or may be assigned a few of them. Nor is there a fixed list of themes to choose from. Progress in digital humanities (and artificial intelligence, in general), though, crucially depends on semantic text processing.

In what follows, an individual attempt at manual thematic tagging of a literary texts sample is presented. It will be further tested against similar annotations performed by other researchers on the same data, to check out similarities and differences and together work out a final annotation of the sample.

The resulting tagged sample will be used twofold. First, it will serve as training data for a learning model to be used in annotating a larger corpus of Russian short stories dated between 1900-1930. Second, it will be set against the automatic topic modeling results, to assess the latter's adequacy and thus the prospects of using standard topic modeling algorithms for annotating the above-mentioned corpus. Once a reliable thematic tagging procedure is found, it can be extended to other literary corpora and text collections.

## 1 Russian Short Stories Corpus (1900-1930)

The paper deals with recent research within the ongoing project “The Russian language on the edge of radical historical changes: the study of language and style in pre-revolutionary, revolutionary and post-revolutionary artistic prose by the methods of mathematical and computer linguistics (a corpus-based research on Russian short stories)” [1;2]. The project aims to give a comprehensive account of early 20<sup>th</sup> century Russian short stories from the viewpoint of genre theory which presupposes their thematic, structural, and linguistic analysis [3: 237]. Although the project's particular emphasis is on language and style, it seems impossible to research into them without taking account of the short stories' themes and plot structure.

The first three decades of the 20<sup>th</sup> century in Russia were a tumultuous time marked by a series of dramatic historical events (Russo-Japanese war, World War I, February and October revolutions, the Civil War) that radically changed the course of Russian history and made an impact on Russian literature and language style. Thus, in a closer perspective, the project's goal is to trace the way all the three aspects (themes, narrative structure and language style) kept changing over time by comparing different chronological periods [4].

To accomplish this, a text corpus was created, containing several thousands of short stories written in Russia and later, the Soviet Union, from 1900 to 1930 and published in literary journals or story books. This timespan is divided into 3 parts, 1900–1913, 1914–1922 and 1923–1930, the first covering the time before the great cataclysms, the second embracing World War I, February and October revolutions and the Civil War, and the third corresponding to the post-war socialist period. Thus there are peace periods at the beginning and at the end linked by an intermediate war-time period.

Each author is represented by a single, randomly selected, story per period. To ensure robustness of the results, the project aims to account for as many professional writers as possible, both famous (e.g. Anton Chekhov, Leo Tolstoy, Ivan Bunin, Maxim Gorky) and lesser-known ones, metropolitan and provincial alike [5].

From the text corpus, a random sample was taken, containing 310 stories by 300 authors (some writers featuring in more than one period, this accounts for a slight

discrepancy in numbers) [6]. This sample serves as an initial testbed for linguists and literary scholars enabling them to put forward and prove (or disprove) preliminary conceptions concerning Russian short stories of the early 20<sup>th</sup> century as a special genre, with its specific themes, plot structure and stylistic features.

## 2 Themes and Their Dynamics of Change

### 2.1 Mapping Short Stories onto Sets of Themes

Identifying themes in works of literature is a challenging job. Theme, or topic, is a fuzzy concept allowing for an array of interpretations. This holds for any discourse type, but it is in fiction that the theme notion is most ambivalent. The main problem is that literary texts are often heavily laden with implicit meanings, or subtext. A short story theme may never manifest itself in a particular word or phrase and remain an unspoken message, thus being quite different from the so-called “word themes” [7: 30-31].

In this respect, literary prose is opposed to both academic and mass media discourse [8]. It is no accident that the merits of various information extraction and data mining techniques are demonstrated on these discourse types rather than on works of fiction. Thematic content of literary texts by and large defies automatic detection although certain noteworthy advances have been made in recent years [9;10].

Another difficulty concerning the thematic content of a literary text is that it normally contains a handful of themes, like *love*, *war*, *death* and *desolation*, or, say, *art*, *poverty*, and *suicide*. In fiction, unlike some other text types, themes are not hierarchically arranged, so that one cannot confidently tag one of them as dominant, or global, and others as subsidiary, or local. Rather, themes may come in a bunch or be consequentially arranged, being linked by causal relations.

In theory, a story's themes can all be put together in a single proposition, as suggested for discourse topics in general [11: 134ff], e.g. *A poverty-stricken artist desperately needs money and, unable to sell his paintings, commits suicide*. Obviously enough, each story then will have an individual topic and there will be little chance for finding regularities.

In the present paper, thematic tagging presupposes the identification of all semantic components that contribute to the plot, determine the protagonist's motives and actions and directly bear on the conflict and its resolution. Themes are considered akin to keywords [12], so that each story is mapped onto a clustering of themes. A parallel can be drawn here with componential analysis that aims to present word meaning in a bundle of semantic features. The difference, though, is that while componential analysis tries its best to bring out the total semantic content of a word, a set of themes is not meant to fully define the short story plot.

As there are no universally accepted algorithms of themes detection in works of fiction, a careful qualitative analysis is needed at the outset. Topic modelling techniques [13; 14] will be considered later, once there is a certain amount of data at hand. Thematic tagging discussed in the present paper was done manually on the above-

mentioned sample comprising 310 stories which will be later used as training data for a learning model.

We proceeded as follows. A rough set of themes was drawn from the 1<sup>st</sup> period stories. It was subsequently tested against the short stories of the two other periods, with inevitable corrections, deletions and additions. The final set for the whole sample currently numbers 89 themes, ranging from political to personal, and from philosophical to mundane.

In the next sections, groups of themes are identified and figures are presented showing their frequency rates over the three periods. One should be aware that the figures in the tables below show the absolute frequencies of themes, i.e. the actual number of stories in which a particular theme occurs. However, as the total number of stories per period is close to 100, these figures can be roughly viewed as percentage.

## 2.2 Political Themes: Wars and Revolutions

The initial three decades of the 20<sup>th</sup> century proved a difficult time in Russian history. Defeat in the Russo-Japanese war (1904–1905), the subsequent political and social unrest, World War I, February and October revolutions of 1917, resulting in a radical transformation of economic, political and social life, and finally the Civil War (1917–1922) with its long aftermath period could not fail to affect Russian literature. It is but natural that these events are used in many stories as settings. We treat such political events as themes if they play a key role in the plot. This is often the case with the war themes. With the revolutions, however, things are different. In a sense, almost all stories of the third period and some of the second one could be marked by such tag since their plot would be deemed unrealistic had not the revolutions taken place. Nevertheless, we think it completely unnecessary to introduce a special tag for the February bourgeois democratic revolution as it was but a prelude to a much more radical socialist revolution in October. As for the latter, only a couple of stories, specifically highlighting the role of this event for the plot, are tagged accordingly.

It is important to note that themes pertaining to the political context are likely to be evoked in fiction long after the events concerned. Thus, the Civil War is a theme in twice as many stories of the third period as those of the second one. The greater the event, the stronger the postponed effect. One should remember this when comparing the figures.

Together with wars and revolutions are listed themes directly bearing on them, including pre-revolutionary unrest (strikes, military mutinies, peasant riots), political activities which ultimately led to the revolution (revolutionary movement), death in the war and execution (Table 1).

The disastrous Russo-Japanese war failed to get a wide coverage in fiction, probably because of its local character, or the shame of the defeat. It was largely forgotten by the time a new war broke out. World War I dragged inconclusively for three and a half years, causing heavy losses in the Russian Army. Soon after the October revolution Russia signed Brest-Litovsk Treaty and emerged from the war. World War I is a frequent theme in the 2<sup>nd</sup> period stories, but seems no longer significant in the 3<sup>rd</sup> period ones.

The October revolution had brought about so fundamental transformations in all spheres of life that World War I began to look like a remote and much less topical event. This accounts for a sharp drop in frequency rate, from 21 to 1. The number of stories about pre-revolutionary activities also declined.

**Table 1.** Political themes

Themes	1900-1913	1914-1922	1923-1930
Russo-Japanese war	3	1	0
Pre-revolutionary unrest	7	3	0
Revolutionary movement	12	6	5
World War I	0	21	1
October revolution	0	2	1
Civil war	0	12	24
Death in the war	6	13	5
Execution	3	5	8

The impact of the Civil War, in contrast, is on the rise, from 12 during the war to 24 after the end. This is by no means surprising. The war's outcome was crucial for the socialist regime, determining the future life of the whole country. In the decades to come, the Red Army victory would be praised and glorified in numerous works of fiction, movies, etc.

The death-in-the-war theme reaches its peak in the 2<sup>nd</sup> period stories, which is but natural, given that wars went on over the whole period, and then decreases. The execution theme, however, gradually goes up, probably due to an increase in the number of stories dealing with the Civil War.

### 2.3 Economic and Social Developments

A thematic block associated with the country's industrial and social policy became particularly prominent after the end of the Civil War, during the third period (1923–1930). Below, in Table 2, major issues are covered.

The construction of Trans-Siberian railway between 1891 and 1916 was a strong impetus toward migration over the Ural mountains into Siberia. At the turn of the 20<sup>th</sup> century, resettlement was promoted and sponsored by the Russian government. However, the process was by no means easy and smooth. Disappointment, frustration and desperation plagued many settlers who found it difficult to survive in the new lands and were on the verge of giving up the whole idea and going back. These attitudes appear in three stories of the first period. During World War I, the resettlement program was dropped. After the end of the Civil war, it was subsumed by larger-scale projects aimed at the total reconstruction of the country's life.

In economics, such large-scale projects were meant to promote technical modernization and inventions. Hence, a noticeable rise in the number of stories tackling these issues during the third period.

**Table 2.** Economic and social themes

Themes	1900-1913	1914-1922	1923-1930
New lands development	3	0	0
Industrial advance	1	0	6
Explorations and inventions	1	1	5
Mass education	0	0	3
Women's emancipation	2	3	7
New social order	1	9	38
The old vs. the new	1	1	17
Bright future	3	0	5
Young people	2	4	9

In the social sphere, the same period is marked by the policy of mass education and women's emancipation, the corresponding figures being higher than previously. The concept of bright future is gaining ground due both to the Soviet propaganda efforts and the universal human need to believe in a better tomorrow which was particularly acute after the end the Civil war. The 3<sup>rd</sup> period stories tend to focus on the young, active and romantic people, architects of the future happy society.

The October revolution totally destroyed the former structure of the Russian society and its old ways of life. The process of instituting a new social order is a key theme of the third period, hitting a record 38 stories. Sometimes the new order is explicitly set against the old one, with the former always evaluated positively and the latter, negatively. Such a neat divide is due to the fact that people disapproving of the October revolution and the subsequent transformations either had to leave the country or keep silent. It was impossible for writers denying the new ideas and values to get their work published.

Changes inflicted by the revolution affected every person and every family (Table 3). In particular, they totally eliminated the familiar pre-revolutionary pattern of the well-to-do families dwelling in the large cities in winter and moving to their countryside estates in summer. The opposition of city and country life is no longer relevant. Instead, the characters of the 3<sup>rd</sup> period stories either reside in the city (workers, clerks) or, most often, are to be found in the rural settlements. There they are trying to survive in the absence of food, cattle, seeds, agricultural implements, horses or any other facilities. The peasant life was undergoing major transformations, perhaps the greatest of all, hence the skyrocketing frequency of this theme in the third period.

The concept of land is central to the third period, too, as one of the first Bolshevik decrees, Decree on Land, abolished private property and proclaimed that peasants should divide up rural land among themselves. This decree led to a split between the poorer farmers who supported the revolution and sided with the Red Army in the Civil War and the wealthier ones, who did their best to retain the traditional lifestyle. The two groups fought over land and the new ways of things in general, sometimes with violence. The latter theme is increasingly present in short stories across the three decades.

**Table 3.** Social structure and lifestyle

Themes	1900-1913	1914-1922	1923-1930
Country vs. city life	9	2	2
Peasant life	3	2	24
Land as property	1	2	3
Violence	5	7	10
Non-peasant work	4	4	11
Working class	1	2	2
Country life	4	3	12
Pets and animals	4	4	13

The number of stories about the non-peasant work and the working class (another target audience of the early Bolshevik decrees) also grows, the dynamics, though, being less spectacular than with the peasantry.

Surprisingly, quite a number of stories in the second and particularly the third period are about country life, pets and animals (embracing the love of nature, the unity of man and nature, and hunting). This bunch of themes seems to have been a last resort for writers hating to praise the Soviet State's new policies and eager to step aside.

## 2.4 Religion

As the Bolsheviks explicitly pursued to eliminate religion with the goal of establishing state atheism, noticeable change can be seen in the relative frequency of such themes as Christian God (incorporating the concepts of faith, saints, sin and even devil) and religion as a social institution across the second and third periods (Table 4).

**Table 4.** God and religion

Themes	1900-1913	1914-1922	1923-1930
Christian God	10	16	4
Religion as a social institution	3	7	7

In the wartime period from 1914 to 1922, the concept of God, quite naturally, was quite prominent. After the ultimate victory of the Red Army, a peace time ensued, marked by an active anti-religious campaign launched by the Soviet government. Spiritual issues are seldom mentioned in the literature of the third period. This is not the case, however, with religion as a social institution. Although from the quantitative viewpoint the third period looks exactly as the second one, the situation is different in two respects. First, in the 3<sup>rd</sup> period stories, the Christian church no longer enjoyed the monopoly and had to make room for the Jewish and Buddhist religions. Second, the references to the church, priests, worshippers, etc. became outright derogatory or ironic, at best. This fact, incidentally, warrants an interpretative, rather than a purely quantitative, approach to the semantic analysis of literary fiction.

## 2.5 Personal matters

One might think that there are timeless, core values in the human life, unlikely to be affected by political whirlpools and social-life transformations. This may well be so as regards individual lives, but in the literature, the impact of dramatic public events prevails. As a result, strictly personal themes like love, marriage, unfaithfulness, jealousy, children, parental love in the second period shifted out of focus and gradually declined, becoming less prominent and frequent (Table 5).

Interestingly, the sexual aspect of love and, more broadly, the body life (including accidental affair, sexual desire and lust) is on the rise. The same holds for rape. Prostitution theme vanished from the 3<sup>rd</sup> period stories following the attempts of the Soviet government to combat and repress this social practice.

**Table 5.** Personal themes

Themes	1900-1913	1914-1922	1923-1930
Romantic love	22	19	17
Unrequited love	1	2	1
Marriage	28	12	13
Unfaithfulness	14	7	6
Jealousy	10	9	8
Children	17	14	11
Parental love	8	6	4
Mutual heterosexual love	6	2	9
Body life	3	11	14
Rape	1	3	3
Prostitution	4	2	0

Other human relationships embrace fraternity and solidarity, friendship, mentorship, and rivalry. Together with them are listed ill effects of interpersonal dealings, including envy, greed, deceit, treachery, and revenge (Table 6).

The dynamics of the fraternity and friendship themes reaching their maximum in the 2<sup>nd</sup> period stories is quite understandable. In a war, mutual help and reliability are greatly appreciated and needed and thus it is but natural that stories written during the war should profile these aspects of human life.

Revenge-theme frequencies display an opposite pattern with an absolute minimum in the second period. Rivalry slightly increases over time, while envy, greed and deceit stay roughly at the same level. Treachery (not to be confused with treason!) steeply goes up in the 2<sup>nd</sup> period stories and remains at the same point in the 3<sup>rd</sup> period ones. Figures for the mentorship theme rise, reaching roughly 5% in the 3<sup>rd</sup> period stories. Their rather high frequency rate may be due to the fact that in Soviet times this type of relationship was actively promoted in professional and ideological domains.

Poverty and hunger plagued people's life more or less steadily. In the wartime period, hardships obviously increased. They did not end after the Civil War as the country was exhausted and near ruin. The economy was devastated, people were starving and dying from epidemics and lack of health-care. The number of stories highlighting



the contrast between the rich and the poor and the crucial role of money went down in the 3<sup>rd</sup> period stories, as there no longer remained wealthy people (Table 7).

**Table 6.** Interpersonal relations

Themes	1900-1913	1914-1922	1923-1930
Fraternity, solidarity	5	14	11
Friendship	2	5	3
Mentorship	1	4	5
Rivalry	1	3	3
Envy	1	1	2
Greed	1	1	2
Deceit	6	6	5
Treachery	0	7	7
Revenge	4	0	5

The alcoholism theme gradually declines over decades which must be due to prohibition introduced in Russia at the beginning of World War I. It carried on through the turmoil of the revolutions and the Civil War until 1925.

The difference between the times of war and peace is most obviously reflected in the figures related to the death-in-the-war theme (see Table 1 above). Surprisingly, there is yet another thematic marker of peace times, and that is boredom. In the epoch of cataclysms, people do not have the luxury of monotonous everyday life.

**Table 7.** Mundane aspects of life

Themes	1900-1913	1914-1922	1923-1930
Poverty, hunger, hardships	9	12	14
The rich vs. the poor; money	12	8	9
Alcoholism	5	3	1
Death from natural causes	10	6	6
Sudden and accidental death	6	3	2
Suicide	6	8	4
Monotonous life; boredom	12	7	11

Some stories touch on the characters' inner life (Table 8). The frequency rates of such themes across the periods can be accounted for by two factors. One is the Soviet ideology that fundamentally "reformatted" the former social structure, overturning the system of norms and values and abolishing charity institutions and philanthropy on the false pretext of an all-sufficient state.

Another is the above-mentioned shift of focus, both in real life and fiction. In the aftermath of the Civil War, people haunted by poverty, starvation, homeliness, and misery could not afford to brood over being lonely, but rather had to survive. The same applies to hopes and illusions.

**Table 8.** Inner life

Themes	1900-1913	1914-1922	1923-1930
Willingness to help, be of use, philanthropy	4	2	0
Nobility of character, magnanimity, self-sacrifice	8	5	1
Readiness to forgive, spiritual rebirth	2	6	2
Loneliness	7	8	3
Mysticism, hallucinations, presentiments	9	12	5
Dreams vs. reality	9	9	4
Frustrated hopes, disillusionment	9	1	1
Insanity	4	2	2
Shame	1	1	0
Remorse	1	1	1
Passion for life	1	1	1

The mysticism theme enjoys a rise during the war and drops sharply in the Soviet period, stories of spiritual experience being unwelcome by the new regime. The insanity theme also diminishes. Shame, remorse and passion for life get their small but steady numbers.

## Conclusion

In this paper, early 20<sup>th</sup> century Russian short stories are analyzed from a thematic viewpoint. Drawing on a large corpus currently containing a few thousand stories, a sample of 310 stories was randomly taken to serve as a testbed for a pilot study. It was further divided into three periods, 1900-1913, 1914-1922, and 1923-1930.

To detect themes typical of the whole timespan (1900-1930), an empirical approach was used. Themes were identified manually in a bottom-up fashion (starting from individual stories) and then tested against the whole sample. In the end, a set of 89 themes was formed, most of them featuring in the tables above. Tagging was also done manually, to ensure the reliability of results. Each story thus is mapped onto a set of themes deemed essential for its content.

For each theme, its occurrences were calculated in the stories of particular periods and then the frequency rates were compared. The figures vary greatly across periods almost for every theme, which is not surprising given the turbulent time the country was going through. Fundamental transformations affected all spheres of life, from public to private.

Short stories published in the third period are quite different from those of the first period. New themes emerged while some old ones dropped in priority, vanished, or radically changed in evaluation. However, the 2<sup>nd</sup> period stories cannot be viewed as a “bridge” between the literature of the two peace periods. They have a distinct character of their own shaped by the large-scale political and military events.

Future research into the Russian short stories thematic tagging will run along several lines. First, a few annotators will perform independent tagging of the sample, to test the degree of subjectivity, pinpoint controversial issues and work out guidelines

for tackling them. Second, each theme will be provisionally linked to a set of semantically associated lexical items (words or collocations). It will be then checked to what extent each theme manifests itself explicitly via such words (it may be presumed that themes will vary greatly in this respect). The results will determine the prospects for their automatic detection in literary texts.

**Acknowledgement.** The paper is supported by the Russian Foundation for Basic Research, project # 17-29-09173 "The Russian language on the edge of radical historical changes: the study of language and style in prerevolutionary, revolutionary and post-revolutionary artistic prose by the methods of mathematical and computer linguistics (a corpus-based research on Russian short stories)".

## References

1. Martynenko, G.Ya., Sherstinova, T.Yu., Melnik, A.G., Popova, T.I. Methodological problems of creating a Computer Anthology of the Russian story as a language resource for the study of the language and style of Russian artistic prose in the era revolutionary changes (first third of the 20th century). In: Computational linguistics and computational ontologies. Issue 2. Proceedings of the XXI International United Conference 'The Internet and Modern Society', IMS-2018, St. Petersburg, May 30 - June 2, 2018 Collection of scientific articles, pp. 99–104 (2018). [In Russian].
2. Martynenko, G.Ya., Sherstinova, T.Yu., Popova, T.I., Melnik, A.G., Zamirajlova, E.V. On the principles of creation of the Russian short stories corpus of the first third of the 20th century. In: Proceedings of the XV International Conference on Computer and Cognitive Linguistics 'TEL 2018', pp. 180–197 (2018). [In Russian].
3. Bakhtin, M.M.: Verbal Creativity Aesthetics. Moscow, Iskusstvo (1979). [In Russian].
4. Martynenko, G., Sherstinova, T. Emotional Waves of a Plot in Literary Texts: New Approaches for Investigation of the Dynamics in Digital Culture. In: Alexandrov, D., Boukhanovsky, A., Chugunov, A., Kabanov, Y., Koltsova, O. (eds.). Digital Transformation and Global Society. DTGS 2018. Communications in Computer and Information Science, vol 859, pp. 299–309 (2018). [https://doi.org/10.1007/978-3-030-02846-6\\_24](https://doi.org/10.1007/978-3-030-02846-6_24).
5. Martynenko, G., Sherstinova, T. Linguistic and Stylistic Parameters for the Study of Literary Language in the Corpus of Russian Short Stories of the First Third of the 20th Century. In: R. Piotrowski's Readings in Language Engineering and Applied Linguistics, Proc. of the III International Conference on Language Engineering and Applied Linguistics (PRLEAL-2019), Saint Petersburg, Russia, November 27, 2019, CEUR Workshop Proceedings. Vol. 2552, pp. 105–120 (2020). <http://ceur-ws.org/Vol-2552/>.
6. Martynenko G., Sherstinova T. Corpus of Russian Short Stories of the First Third of the 20th Century: Theoretical Issues and Linguistic Parameters. In: Structural and Applied Linguistics. Vol. 14. St. Petersburg.
7. Zhirmunskii V.M.: Literary theory. Poetics, Stylistics. Leningrad, Nauka (1977). [In Russian].
8. Blummer, B., Kenton, J.M. Academic Libraries' Outreach Efforts: Identifying Themes in the Literature. Public Services Quarterly, 15 (3), 179–204 (2019).
9. Rhody, L.M. Topic Modeling and Figurative Language. Journal of Digital Humanities, 2(1) (2012). URL: <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-figurative-language-by-lisa-m-rhody/>.

10. Iyyer, M., Guha, A., Chaturvedi, S., Boyd-Graber, J., Daumé III, H. Feuding Families and Former Friends: Unsupervised Learning for Dynamic Fictional Relationships. In: Proceedings of the 2016 Conference of the North American Chapter of the Association of the Computational Linguistics. (2016). URL: <https://www.aclweb.org/anthology/N16-1180>.
11. van Dijk, T.: Text and Context: Explorations in the Semantics and Pragmatics of Discourse. Longman, London and New York (1977).
12. Watson Todd, R. Discourse Topics. John Benjamins, Amsterdam & Philadelphia (2016).
13. Blei, D.M., Ng, A.Y., Jordan, M.I. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(4–5). (2003). URL: <http://jmlr.csail.mit.edu/papers/volume3/blei03a/blei03a.pdf>
14. Mitrofanova, O.A. Probabilistic Topic Modelling of the Russian Text Corpora by Means of GenSim Toolkit. In: Proceedings of the International Conference «Corpus Linguistics – 2015». St.-Petersburg (2015).