What is Russian Elegy? Computational Study of a Nineteenth-Century Poetic Genre

Antonina Martynenko[1,2]

[1] University of Tartu, Ülikooli 18, 50090 Tartu, Estonia
[2] Institute of Russian Literature (Pushkin House), nab. Makarowa 4, 199034 Saint-Petersburg, Russia
`antonina.martynenko@ut.ee`

Abstract. This paper studies the change in formal features of Russian elegies in order to outline the evolution of the genre between 1815 and 1835. Using representative historical corpus study shows metrical and thematic homogenization in Russian elegy: genre's identity shifted from various types of funeral and historical texts to a short love poem written in iambic tetrameter. Despite the previously claimed diffusion of an elegy as a genre, paper demonstrates it is possible to recognize an elegy from the general poetic language using SVM classification, which also highlights interpretable and distinctive lexical features of the genre.

Keywords: Poetry, Topic Modelling, Genre, Classification, Russian Literature.

1   Background

The elegy has a long and tangled history in Ancient and European literatures. In Russian literature the genre first appeared in the middle of the 18th century as a poem about love and sorrow, but was neglected already in the 1780s. In the late 1790s the elegy was reinvented under a strong influence of English and French traditions. While the first writings were translations and adaptations of canonic English, French [2] and Latin elegies [6], since the 1810s Russian poets had produced a large number of original elegies. The period between the late 1810s and the early 1830s is considered to be the most important stage in the development of Russian elegy.

Numerous qualitative studies dedicated to the elegiac texts of this period claim that in the middle of the 1820s the elegy became the main genre which both refined the Russian poetic language [14] and influenced the poetic tradition as a whole, when special elegiac key words, collocations and motives started to be transmitted to other lyric genres [4]. However, even the 19th century literary critics did not agree on the definition of the elegy, because of significant variation in its origins and forms [1]. At the same time, the period of the development of the Russian elegy coincided with the start of the decomposition of the genre system of Russian poetry, when most genres began to lose their forms and functions [4]. As a result, the traditional literary scholarship states that 19th century Russian elegy cannot be defined through its thematic or formal features and the genre's history may only be outlined by studying exemplary "key" texts. Most

literary scholars had analyzed canonical elegiac poems (e. g. Alexander Pushkin's) and paid much less attention to the large number of elegies published either by minor authors or anonymously between 1825 and 1835 [10, 14].

This article aims to study the elegiac genre using quantitative methods at the scale of literary population. Special attention will be given to the development of the elegy after 1825, since this period was assumed to be secondary in the evolution of the genre and had rarely been under consideration. The quantitative approach will allow to test the historical change in elegiac form and content and address the question of genre distinctiveness (under the premise of genre system disintegration) formally.

Computational approaches to literary genres are now widely discussed in Digital Humanities. These approaches are driven by the assumption of common formal features behind vague yet widespread literary categorization. These methods are mostly applied to prose fiction (in particular, novel and its genres [13]) and sometimes to drama [8]. Poetic data is used more rarely because lyrical poems are quite short and may not form a corpus large enough for statistical analysis. The aim of this contribution is to analyze a small poetic corpus using both supervised (SVM) and unsupervised methods (topic modelling) and to discuss the limitations of these approaches to small genre-specific historical data.

2   Data

A corpus of Russian poems that were published with a (sub)title "an elegy" was compiled for this study. Poems were gathered from the printed periodicals (literary journals and almanacs) issued between 1815 and 1835; texts were digitized and converted to modern Russian orthography. The original punctuation, line division and rhymes were retained according to the historical sources; part of the collection introduces previously undigitized and "unread" texts for the first time[1].

Final corpus includes 386 elegiac poems which contain approximately 16,000 lines and 86,000 word tokens. All texts were lemmatized with the Yandex Mystem morphological analyzer commonly used for Russian language [9], the lemmatized corpus contains about 7,200 lemmata (types). The sources were limited only to periodicals because the authors of this time tended to publish new poems in journals and almanacs, so it is possible to assume that the appearance of a poem in a source reflects its actual date of creation. In the case of poetry collections, it is not possible to date a poem precisely. These limitations were applied in order to make the corpus representative and to reflect the historical meaning of the genre title "an elegy" in the period between 1815 and 1835. The distribution of the poems in each year is shown on Fig. 1. Though the corpus is not balanced, this distribution represents explicit appearances of the genre title in periodical printed sources.

---

[1] The corpus and other materials related to the contribution can be found at Github: https://github.com/tonyamart/elegies_dhn.
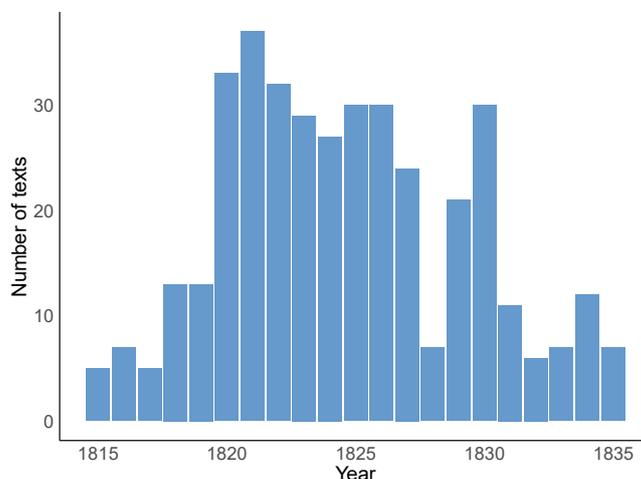
Fig. 1. Distribution of poems in the corpus of elegies.

The corpus of elegies is thus quite small and specific to be studied without a comparison to more general poetic data, so the Poetic subcorpus of the Russian National Corpus was used in the study. The subset from the Poetic subcorpus includes more than 3,000 poems dated between 1815 and 1835 which corresponds to approximately 808,000 word tokens (about 28,000 lemmata).

Another corpus used for comparison with the elegies is the corpus of "Russian songs", a specific Russian poetic genre of the first half of the 19th century which imitated a folk song[2]. The corpus of "Russian songs" was also limited only to well-dated poems published in periodicals in order to be comparable with the corpus of elegies. The subset of the corpus of "Russian songs" used in the study comprises 327 poems containing about 35,000 word tokens and 4,900 lemmata.

The corpus of elegies contains the following additional annotation and metadata: metrical features of a poem (meter, number of feet), number of lines, information about the place of publication, and author's sign. Simple metadata overview suggests a prestige bias in the studied period. At the beginning of the period, in early 1820s, the most remarkable young poets such as Alexander Pushkin and Evgenij Baratynski, were engaged in writing elegies. These young poets' focus on the elegy thus implies that the elegy was initially a promising new genre, prestigious and appreciated, so it was elaborated by popular and soon honored Russian poets.

However, already in the early 1830s the elegy seemed to be popular mostly among novice non-professional poets who signed their texts with acronyms and remained unknown for scholars (e. g. N. Stavelov, an unknown poet, published 12 elegies during 1825–1835). Since poets who wrote elegies in the late 1820s – early 1830s were often considered to be "epigones" their impact on the development of the genre had not been thoroughly studied.

---

[2] The corpus was prepared by Artjoms Šeļa and it is accessible at his Github page: https://github.com/perechen/russian.songs.fin.thesis.

Based on the overview above, my hypothesis is that the appearance of unknown, non-professional authors might have influence not only on the decline in the prestige of the elegy but also on the content and the form of the poems themselves. So, the insufficiently studied "epigone" period of the elegy's history will differ from the elegies published in the beginning of 1820s. At the same time, I assume that the internal genre changes were gradual, so that the elegy as a genre will be still distinguishable from other poetic genres of this period.

3   Methods

To test the hypothesis of the content change in elegies I used LDA topic modelling algorithm since this method is proven to be applicable both for short poems [5] and genre analysis [8]. A model with 20 topics was created using R implementation of LDA in "topicmodels" package[3]. In comparison with other studies mentioned before, small number of topics is explained by the small size of the corpus of elegies: models with more than 20 topics led to topics with only one highly probable word while all other words had equal probabilities close to 0; at the same time, topics with more than one probable word were similar to those from 20 topics LDA model.

The present model shows that some of the topics are distributed unequally during the period under consideration. In Fig. 2 the mean probability of each topic is aggregated for the time window of 3 years.
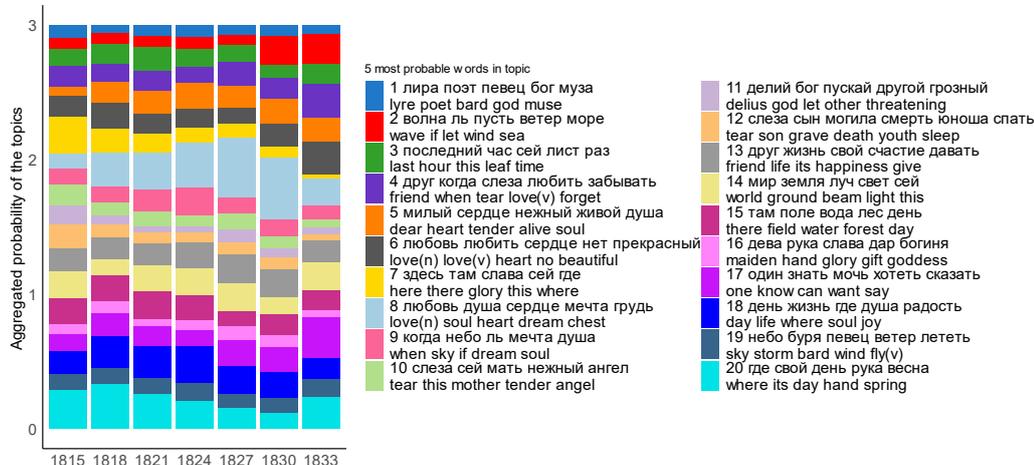

Fig. 2. Distribution of the topics on the timeline.

---

[3]   Before creating the LDA model, a list of 120 stop words was removed from the corpus (these are mostly prepositions, conjunctions, and other frequently used function words). The full list of stop words can be found in the script dedicated to topic modelling at the contribution's Github repository.

It can be seen that in the beginning of the period, the elegy is displayed as a quite heterogeneous genre with no prevalent theme. Among the topics particularly specific to the period of late 1810s – early 1820s the 11th and the 12th should be given particular attention. Topic 11 is related to the translations and imitations of Latin love elegies containing elements of the idyllic nature (for example, this topic is highly probable in one of the most famous Russian loose translation from Tibullus I, 3 written by Kostantin Batyushkov). Topic 12 featured in the late 1810s also relates to the European exemplary elegies, namely the funeral elegy describing a young man's death on a battlefield (among the 20th most probable words in this topic there are also "dream", "memory", "fear", "to perish", "to cry"). In this regard, a short presence of a topic related to the historical and, more precisely, battle scenes[4] (topic 7) seen before 1821 is very noticeable as well. Poems with high probability of these topics reflect the short impact of the poetry of the Napoleonic wars that had disappeared already at the beginning of the 1820s.

The period between 1824 and 1832 could be described as the emergence of the theme of romantic love and the theme of regret for the loss both of love and youth (topics 6 and 8). What is important, both of these topics are not directly related to the theme of mourning one's death and graveyard self-reflection which are considered traditional for the European elegy. The romance topics (6 and 8) appear as highly probable already in the elegies published in the early 1820s (such as the exemplary Baratynsky's elegy "Do not tempt me needlessly..." ("Не искушай меня без нужды...")). However, these topics became prevalent in the population only in 1827–1832 when their inventors had already abandoned the genre. Therefore, topical distribution shows an uncommon development of the elegy in Russian tradition: from the funeral or even "battlefield" meditation to a love poem which was more similar to French "poésie légère" tradition. Irregularities in topics in the latest period may be caused by the small number of elegies published in 1833–1835, but it is still possible to conclude that the content of the elegies did change in a specific direction, which cannot be fully explained by a mere replication of successful examples of the past by so-called "epigone" poets of 1825–1835.

This conclusion – the elegies becoming short love poems at the end of the 1820s – may be supported by other formal features of the texts in the corpus. The study of the metrical repertoire[5] shows that decrease in thematic diversity happens simultaneously with the decrease of metrical variation in the corpus (see Table 1). Elegies published before 1825 were written in a wide range of different meters. More importantly, these were the elegy's specific meters such as free iambic verse, iambic hexameter, and iambic verse with regular alternation of iambic hexameter and iambic pentameter. However, by the end of the 1820s more than a half of the poems in the corpus were composed in iambic tetrameter. Although the rapid expansion of iambic tetrameter happens everywhere in the Russian poetry in the beginning of 1820s, the elegy seems to be a more

---

4   Among the 20 most probable words in this topic there are also keywords such as "glory", "foe", "sword", "a hero", "proud", "field".

5   The identification and calculation of meters were done accordingly to the principles described by Mikhail Gasparov in his canonic work on Russian metrics [3].

conservative genre which retained its metrical features and changed slowly in comparison with the shifts in metrical repertoire of the general poetic corpus [3].

Table 1. Five most frequent iambic forms in the corpus (other forms are not included in the table); values in the brackets are the number of elegiac poems written in this form[6].

	1815—1819 43 poems	1820—1824 158 poems	1825—1829 112 poems	1830—1835 73 poems
Iamb-4	6.9% (3)	20.8% (33)	49% (55)	56% (41)
Iamb-5	—	5.6% (9)	4.4% (5)	2.7% (2)
Iamb-65	16.2% (7)	9.4% (15)	1.7% (2)	1.3% (1)
Iamb-6	27.9% (12)	22.7% (36)	16% (18)	12.3% (9)
Free iambic verse	27.9% (12)	20.8% (33)	14.2% (16)	17.8% (13)

Another important formal feature for understanding the evolution of elegiac poems is the length in lines. The length of a poem is strongly correlated[7] with the year of publication: both mean and median lengths of poems aggregated for each year demonstrate significant decrease in elegies' length roughly from 60 to 30 lines during the period from 1815 to 1835 (Fig. 3). The similar conclusion about poems' decline in length at this period based on the study of all-genre corpus of Russian poetry was made by Artjoms Šeļa and Oleg Sobchuk [11]. The decrease in the size of elegies', therefore, proves that the corpus of elegies adequately represents the processes that were happening in poetry of this period.
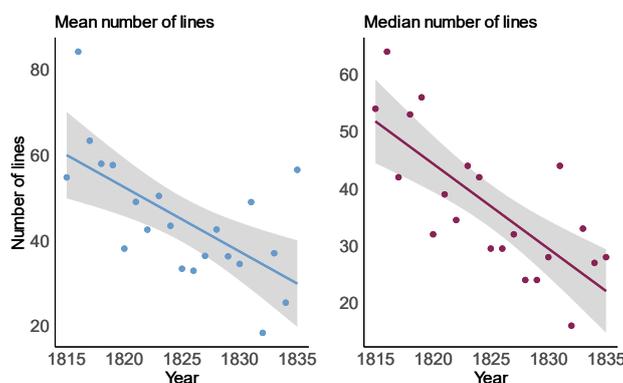

Fig. 3. Mean and median length of elegies aggregated for each year.

---

6   According to Gasparov's data [3] (Gasparov's corpus is comparable with the poetical subcorpus of Russian National Corpus), for the 10-years period between 1821 and 1830 the distribution among these iambic forms is as follows: Iamb-4 – 55%, free iambic verse – 17%, Iamb-5 – 11.5 %, Iamb-6 – 5.5% (no data for Iamb-65).

7   r = -0.64 and -0.75 for correlation between year and mean and median lengths respectively.

To summarize the processes discussed above: between 1815 and 1835 an elegy became thematically and metrically more homogeneous, converging to romantic love theme and iambic tetrameter meter; at the same time poem's size has reduced significantly. These findings based on quantitative evidence explicate the process specific to Russian elegy in the 1820s and 1830s and prove the hypothesis that the last period of massive adoption of the genre was not a simple repetition of the canonic elegies written in early 1820s but has its own evolutionary dynamics.

However, all poems assembled into the corpus still had the explicit genre title, so one would hope to find evidence that the internal changes in the genre have not led to its complete collapse. Did elegy survive as a genre?

This question can be addressed through a classification experiment since genre-specific corpora (elegies and "Russian songs") may be used as labeled data for a genre classification experiment.

In the study on the tradition of "Russian songs" Artjoms Šeļa has shown that the folklore imitations may be quite easily distinguished by its lexical features from the general poetic corpus (such as the Poetic subcorpus of the Russian National Corpus) [12]. However, the poetic language used in elegies is assumed to be closer to the general poetic style and will not be as simply recognizable. Although the elegy should be closer to neutral poetic style, the main hypothesis is that the genre should yet remain distinguishable by its lexical features.

To test this hypothesis two pairs of datasets were used. First, the corpus of elegies (386 poems labeled as an elegy) was compared with the subset from the Russian National corpus (the set comprises 3,157 poems from which a sample of 386 non-elegies was extracted randomly). In the second step, the results obtained from the classification of elegies were compared with the similar experiment where the model tried to categorize texts from the corpus of "Russian songs" (326 poems) and the corresponding subset of poems from the National corpus dated between 1821 and 1848 (according to the chronological span in the corpus of songs; the subset includes 5394 poems from the National corpus from which 326 non-songs were extracted randomly for the experiment).

Among different sets of most frequent words ranked from the 1st to the 400th the best accuracy was obtained with relative frequencies of 250 words ranked from the 50th to 300th. This can be explained by a small size of the two corpora where thematically important words start to appear already in the first hundreds of most frequent words. The frequencies of lemmatized words were used in order to reduce the impact of metrical features of the texts. The classification was performed with Support Vector Machines[8] using linear kernel and 10-fold cross validation[9], the proportion for the training and test corpus was set to 75/25% that result in 580/192 and 448/164 training/test texts in elegies and songs experiments respectively.

---

[8]   On the high performance of SVM applied to the poetic data see the article by Petr Plecháč, Klemens Bobenhausen, and Benjamin Hammerich [7].

[9]   The models were build using R package "tidymodels"; similar results were obtained with the package "e1071" which was used to plot the most distinctive words on the Fig. 4.

In this setting the accuracy of predicting the folklore imitation, a "Russian song" from the general corpus is about 81% which seems to be a good result for such a small collection and non-aggregated single poetic texts. Under the same experimental setting the prediction accuracy for elegies and non-elegies was lower, about 75%. The most distinctive words for the model remain interpretable for a human reader (see Fig. 4): these are the key-notions of the love elegy emerging in the middle of 1820s such as "tear", "grave", "beauty", "wish", "forget", etc.
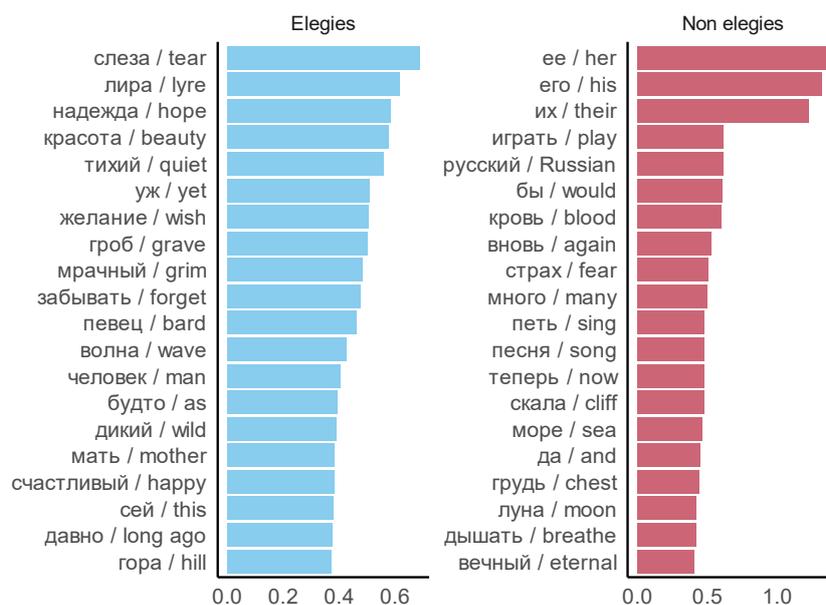

Fig. 4. Most important features for SVM model in distinguishing two groups of texts.

One can imagine that this result was biased by the unequal distribution of metrical forms between two samples (it is easier for the model to recognize elegies because they are very homogenous metrically compared to the diverse sample of Russian poetry). To scrutinize this, I have repeated the experiment with the strict selection of iambic forms for both elegies and non-elegiac poems (351 poems for each group). It did not affect the classification results: the prediction accuracy has not changed significantly and remained around 74%. In additional experiment the set of elegies written in iambic tetrameter was compared with other poems written in this meter. This, however, led to a significant drop in corpus size (despite the proportional prevalence of this form, only 132 elegies in the corpus are in iambic tetrameter). The results of classification in this case are unstable and range from approximately 60 to 70% of accuracy. In this case, the prediction accuracy highly depends on the random sample of poems taken from the general corpus and it cannot be stated whether the elegies in iambic tetrameter are less distinctive or there is just not enough data for the classification task.

4   Conclusions

This study on poetic genre shows that analysis of a small-scale poetic corpus might give meaningful results that are not only consistent with the previous scholarly research of the genre, but also provide new insights about neglected periods in literary history (late 1820s – early 1830s for the Russian elegy). As it was shown above, quantitative methods help to describe the development of poetic features used in elegies: reduction in size, changes in meter and themes. At the same time, the classification methods may be used to test whether a genre with all its internal changes remains distinguishable from general poetic language. Formal classification experiments show that elegies are overall similar to the general poetic language of the first half of the 19th century and cannot be classified against the general corpus as easy as folklore imitations that are very distinct in their lexical features. Nevertheless, the classification accuracy remains better than guessing at random, which suggests that elegy as a genre was not entirely diluted in the poetic language of the time and retained distinctive lexical features. This gives hope that computational methods could be used on a small scale and give meaningful answers to the questions of localized literary history and support the investigation of "micro-genres".

References

1. Fleishman, L.: Iz istorii elegii v pushkinskuju epokhu [From the history of the elegy in Pushkin's time]. In: Pushkinskij sbornik, vol. 106, pp. 24–53. Riga (1968).
2. Frantsuzskaja elegia XVIII-XIX vekov v perevodach poetov pushkinskoj pory [French elegy of the 18th and 19th century translated by the poets of Pushkin's time]. Raduga, Moscow (1989).
3. Gasparov, M.: Sovremennyj russkij stikh. Metrika i ritmika [Contemporary Russian verse. Metrics and rhytmics]. Nauka, Moscow (1974).
4. Ginzburg, L.: O lirike [About lyric poetry]. Sovetskij pisatel', Moscow (1974).
5. Navaro-Colorado, B.: On poetic topic modeling: Extracting themes and motifs from a corpus of Spanish poetry. Frontiers in Digital Humanities 5, 15 (2018).
6. Pilshchikov, I.: Problema mezhjazykovoj intertekstualnosti: na materiale klassicheskoj elegii [The problem of interlingual intertextuality: on the material of classic elegy]. Moscow (2007).
7. Plecháč, P., Bobenhausen, K., Hammerich, B.: Versification and authorship attribution. A pilot study on Czech, German, Spanish, and English poetry. Studia Metrica et Poetica 5(2), 29–54 (2018).
8. Schöch, Ch.: Topic Modeling Genre: An Exploration of French Classical and Enlightenment Drama. Digital Humanities Quartely 11(2), 266–285 (2017).
9. Segalovitch, I.: A Fast Morphological Algorithm with Unknown Word Guessing Induced by a Dictionary for a Web Search Engine. In: Proceedings of the International Conference on Machine Learning; Models, Technologies and Applications. MLMTA'03, June 23–26. Las Vegas, Nevada, USA (2003).
10. Semenko, I.: Poety pushkinskoj pory [Poets of Pushkin's time]. Khudozhestvennaja literatura, Moscow (1970).

11. Shelya, A., Sobchuk, O.: The shortest species: how the length of Russian poetry changed (1750–1921). Studia Metrica et Poetica 4(1), 66–84 (2017).
12. Šeļa, A.: "Russkaja pesnja" v literature 1800–1840 ["Russian song" in the literature of 1800–1840s]. University of Tartu Press, Tartu (2018).
13. Underwood, T.: Distant Horizons. Digital Evidence and Literary Change. The University of Chicago Press, Chicago, London (2019).
14. Vatsuro, V.: Lirika Pushkinskoj Pory [Lyrics of Pushkin's time]. Nauka, Saint-Petersburg (1994).