AUTHORSHIP ATTRIBUTION WITH A VERY NAÏVE BAYES MODEL AND WHAT IT CAN TELL US ABOUT RUSSIAN POETRY[1]

Piperski A. (apiperski@gmail.com)
Russian State University for the Humanities / National Research University Higher School of Economics, Moscow, Russia

This paper presents the results of application of a Naïve Bayes classifier based on character n-grams and words to a corpus of Russian poetic texts annotated for author coming from the Poetic subcorpus of the Russian National Corpus. It turns out that character 4-grams allow us to make very good prediction with respect to the authorship of individual texts. The model achieves a quality of 72.03% on a corpus of 69 poets, which is very good as compared to the Most-Frequent-Class classifier baseline which has an accuracy of 5.09% only. The confusion matrices resulting from the classification can be interpreted in various ways. First, they show the idiosyncraticity of a poet's oeuvre; second, they make it possible to identify pairs of similar authors, and these pairs can be represented as a graph illustrating the history of Russian poetry.

Keywords: authorship attribution, Naïve Bayes classifier, Russian poetry, Russian National Corpus, text similarity.

ЧТО МОЖЕТ СКАЗАТЬ О РУССКОЙ ПОЭЗИИ ОПРЕДЕЛЕНИЕ АВТОРСТВА С ПОМОЩЬЮ НАИВНОГО БАЙЕСОВСКОГО КЛАССИФИКАТОРА?

Пиперски А. Ч. (apiperski@gmail.com)
РГГУ / НИУ ВШЭ, Москва, Россия

В данной статье представлены результаты применения наивного байесовского классификатора, основанного на символьных n-граммах и словах, к корпусу русских поэтических текстов с размеченным авторством из Поэтического подкорпуса Национального корпуса русского языка. Показано, что символьные 4-граммы дают высокое качество определения авторства отдельных текстов. Модель достигает качества 72,03% на корпусе из 69 поэтов, что заметно превышает качество бейслайн-классификатора, приписывающего все тексты к наиболее частотному класса (он дает лишь 5,09% верных ответов). Матрицы ошибок, полученные в результате классификации, могут интерпретироваться различными способами. Во-первых, они показывают особенности творчества отдельных поэтов; во-вторых, они позволяют идентифицировать пары похожих авторов, и эти пары можно представить в виде графа, иллюстрирующего историю русской поэзии.

Ключевые слова: определение авторства, наивный байесовский классификатор, русская поэзия, Национальный корпус русского языка, сходство текстов.

[1] This work is supported by the Russian Science Foundation under grant 17-78-10196.

1. Introduction

Authorship attribution has been a long-standing and important task in computational linguistics (for recent comprehensive surveys, see Juola 2008; Koppel et al. 2009; Stamatatos 2009; Luyckx 2011; Neal et al. 2017). However, this kind of studies has provided philology only with a limited number of recognized successes. Perhaps the most frequently cited example is the study by Mosteller & Wallace (1964) attributing a number of disputed *Federalist Papers* to James Madison. Authorship attribution studies are widely covered in press; however, they only rarely put an end to a discussion about the authorship of a certain literary work. Some examples of this kind include the attribution of *The Cuckoo's Calling* by Robert Galbraith to J.K. Rowling (Juola 2013), the "unmasking" of the famous Italian novelist Elena Ferrante (Savoy 2017), a paper on the disputed authorship of Mikhail Sholokhov's (or someone else's) *And quiet flows the Don* presented at this very conference (Mikheev & Erlikh 2017), or a recent paper on Old English poetry that aims to disprove the hypothesis that *Beowulf* was composed by more than one author (Neidorf et al. 2019). Many papers of this kind contain interesting claims, but the cause is only solved if an author decides to uncover their identity (as it was the case with Robert Galbraith and J.K. Rowling), or a final decision is yet to be made by philologists, who regard authorship attribution as a forensic enterprise in a Holmesian style, where unsystematic, but irrefutable pieces of evidence are worth more than statements of the kind "This work was composed by A with a probability of *x* % or by B with a probability of *y* %". This way of thinking is justified when we are dealing with individual masterpieces, even though it might seem unreasonable to a statistically-minded scientist. The results of authorship attribution studies are more readily accepted in forensics and in social media studies, where their application has proven to be very useful (Coulthard 2013; Rocha et al. 2017)

Even though authorship attribution studies rarely serve to actually identify authorship of literary works, they are valuable from a methodological point of view since they discuss various approaches to measuring text and corpus similarity dealing with corpora by different authors and contribute to our understanding of individual style. For instance, Holmes (1994) claimed that lexical items are especially important for authorship attribution. A seminal study by Burrows (2002) highlighted the importance of function words for defining the style of an author. Another line of research that has proven to be even more successful starts with Kjell et al. (1994) who proposed to identify authorship based on character n-grams; this approach was used by Kešelj et al. (2003) among others, and its superiority is recognized by Juola (2008), Kestemont (2014) and many others. Kestemont et al. (2012) compare this to the developments in art history that happened in the 19th century, when experts agreed that the attribution of a painting to an author cannot be based on its content, but turns out to be more successful if subtle and unnoticeable features are taken into consideration. Stamatatos (2018) goes even further in hiding content from authorship attribution algorithms by removing low-frequency words that may bear topic-related information. However, the linguistic nature of the features represented by character n-grams remains far from clear; one might say that character n-grams capture 'a bit of everything', but further analysis is needed (Kestemont 2014; cf. also Sundararajan & Woodard 2018; Sari et al. 2018). To sum up, a successful experiment in authorship attribution may not bring us new insights about who authored some literary masterpieces, but it may shed some light on what style is and how it is constructed.

Another type of information that can be gained for an authorship attribution experiment concerns the nature of the analyzed texts. For instance, if an author is hard to recognize automatically, this might mean that their works are either very heterogeneous or very similar to some other authors in the set. On the other hand, if an author is accurately recognized, this can serve as evidence for homogeneity and originality of their corpus (the word "originality" is not a value judgment here; an author with a high degree of originality in terms of authorship attribution was not only uninfluenced by other authors, but also did not influence others). In fact, the similarity relations between authors identified from the mistakes frequently made by the classification algorithm are more valuable than correct attributions given that the algorithm performs well in general, because these mistakes are indicators of affinity between authors that can be further studied from the point of view of the history of literature.

The aim of this paper is to show that (partially unsuccessful) authorship attribution can provide insights into the history of Russian poetry and also into what cons

2. Methods

For the purpose of the experiment, I use the Poetic subcorpus of the Russian National Corpus (www.ruscorpora.ru). The Poetic subcorpus includes 11m tokens with rich metadata covering more than 250 years of Russian poetry starting with the first half of the 18th century. There are 69 authors that are represented by more than 50,000 tokens in the corpus. The full list of these authors is given in Table A1 (see Appendix). The set of texts used for the experiment includes 36,446 poems with 5.7m tokens.

For these 69 authors, the following experiment was undertaken. Each individual text was excluded from the corpus (leave-one-out cross validation) and then assigned to one of the authors using a Naïve Bayes classifier width additive smoothing (a previously unseen feature was assumed to be seen 0.5 rather than 0 times). The experiment was performed using four different levels of segmentation, namely, words, character 2-grams, 3-grams, 4-grams, and 5-grams.[2] All texts were converted to lowercase. For the purposes of tokenization, sequences of alphanumeric characters and hyphens were regarded as words. Spaces, punctuation marks, and line breaks were regarded as regular characters and were included into character n-grams. Only texts longer than 400 characters were used for the cross-validation procedure. The number of such texts is 25,596, and they comprise 70.2% of the corpus.

3. Classifier accuracy and error analysis

When evaluating the results of authorship attribution, it is important to select appropriate baselines. For more practically oriented studies, various n-gram based classifiers may themselves serve as a baseline (Vilariño Ayala et al. 2011; Ge et al. 2016). However, for the purposes of this paper some number of mistakes made by the classifier is necessary, so there is no need to search for an ideal classifier: a classifier with an accuracy of 100% would be of absolutely no use to analyze similarity between poets. One only needs to show that the classification results achieved by the Naïve Bayes classifier are reasonable enough so that the classifier's mistakes can be deemed to be reasonable, too; to satisfy this goal we only need very unsophisticated baselines, such as a classifier randomly assigning a class to a poem or a Most-Frequent-Class classifier (Luyckx 2011). The former classifier achieves an accuracy value of 1 / 69 = 1.45%. The latter classifier ascribes any text to Valery

[2] It would be possible to use other levels of segmentations, e.g. word n-grams, lemmas and lemma n-grams, or character n-grams of size larger than 5. However, as mentioned in the Introduction, character n-grams have been shown to be the best type of features for authorship attribution; also, it will be shown below that the performance drops for character 5-grams as compared to character 4-grams (apparently due to data sparseness), which was the reason not to test the performance of the classifier with longer character n-grams. Some more sophisticated techniques such as using n-grams of variable length (Houvardas & Stamatatos 2006) may be applied in future research.

Bryusov, who authored the biggest number of texts in the corpus that are more than 400 characters long. This classifier's accuracy is 1303 / 25,596 = 5.09%. Experiment results for "real" classifiers as compared to the baselines are presented in Table 1:

Classifier	**Accuracy**
Random	1.45%
Most Frequent Class	5.09%
Words	44.60%
Character 2-grams	54.41%
Character 3-grams	67.99%
Character 4-grams	72.03%
Character 5-grams	57.76%

Table 1. Classifier accuracy.

The results achieved using character n-grams are quite impressive. An accuracy of 72.03% when choosing the author of an unknown text among 69 poets is hardly possible even for a very well-versed human expert. For further analysis, I will use the results produced by the classifier based on character 4-grams which has the best performance among the classifiers tested and by the classifier based on words; the quality level of the latter is rather poor, but it leaves some room for interpretation.

The performance of the classifier is captured in more detail by a confusion matrix, which has the size of 69×69 in this case. A part of the matrix for character 4-gram classifier is presented in Table 2:

Actual / Predicted	A.A. Akhmatova	A.A. Blok	A.A. Fet	A.K. Tolstoy	A.N. Apukhtin	…
A.A. Akhmatova	193	5	1	1	0	…
A.A. Blok	5	450	7	1	3	…
A.A. Fet	0	7	290	18	6	…
A.K. Tolstoy	0	0	1	47	1	…
A.N. Apukhtin	1	30	27	24	218	…
…	…	…	…	…	…	…

Table 2. A fragment of the confusion matrix for character 4-gram classifier

Looking at Table 2, one can say that Akhmatova's and Apukhtin's texts are very easy to classify, which is not true for A.K. Tolstoy, whose texts are often confused with Apukhtin's and Fet's poems. Blok's and Fet's recognizability lies somewhere in between. This can be summarized by the proportion of correctly identified texts for each of the poets. These proportions are given in Table A2 (see Appendix) for the character 4-gram classifier.

The recognizability figures presented in Table A2 can serve as a measure of idiosyncraticity of a poet in the Poetic corpus. Thus, S.V. Petrov, B.P. Kornilov, and V.V. Khlebnikov turn out to be very idiosyncratic, whereas S.I. Kirsanov, M.M. Kheraskov, and A.K. Tolstoy are not. However, idiosyncraticity is not to be confused with originality, since its lack can be explained by an influence exerted on later poets; this is probably the explanation for the fact that A.S. Pushkin and V.A. Zhukovsky are so hard to classify. A likely correlate of idiosyncraticity is the homogeneity of a poet's oeuvre. It is worth noting that recognizability is not correlated to the mean length of text by an author (Spearman's $\rho = -0.06$).

The picture is to some extent different if we look at the confusion matrix based on words (Table A3, see Appendix). The top part of the table is not much different from Table A2. The most recognizable poets are B.P. Kornilov, B.Yu. Poplavsky, N.M. Yazykov, V.V. Khlebnikov, M.V. Lomonosov, and S.V. Petrov, and this list of six includes three most recognizable poets based on character 4-gram classifier. At the bottom of the table, one finds N.A. Nekrasov who was not identified correctly even a single time; his poems are most often assigned to his contemporaries A.N. Apukhtin, P.F. Yakubovich, and I.S. Nikitin. Also hard to recognize using words are M.I. Tsvetaeva, A.S. Pushkin, I.L. Selvinsky, A.N. Maykov, and V.V. Mayakovsky. For Pushkin, this might be explained by the great influence he exerted on all subsequent Russian poets. Tsvetaeva, Selvinsky, and Mayakovsky are known for their experiments with form, but the words they use are evidently not so specific. Once again, it is worth noting that recognizability is not correlated to the mean length of text by an author (Spearman's $\rho = 0.07$).

Actually, recognizability based on character n-grams can be interpreted as phonological (i.e., euphonic) recognizability, whereas the data from the word-based classifier show lexical recognizability. The ranks of the poets in these two tables are strongly correlated (Spearman's $\rho = 0.66$), but some discrepancies are very telling. There are nine poets with ranks that are at least twice smaller in Table A3 (lexical recognizability) than in Table A2 (euphonic recognizability) and with an absolute difference of ranks exceeding 5, namely:

Andrei Bely, I.A. Brodsky, I.I. Dmitriev, A.A. Akhmatova, V.A. Lugovskoy, Z.N. Gippius, V.I. Maykov, M.A. Voloshin, I.S. Nikitin.

On the other hand, for nine other poets the opposite is true, i.e. their rank based on euphonic recognizability is smaller than their rank based on lexical recognizability:

P.G. Antokolsky, A.T. Tvardovsky, G.R. Derzhavin, K.K. Sluchevsky, K.D. Balmont, N.A. Klyuev, B.A. Slutsky, B.L. Pasternak, V.V. Mayakovsky.

In other words, the comparison of Tables A2 and A3 shows that the former group of poets, including Akhmatova and Brodsky, is characterized by idiosyncratic vocabulary, whereas the latter group of poets, including Pasternak and Mayakovsky, is characterized by idiosyncratic sound patterns.

Another kind of information that comes from confusion matrices pertains to similarity between authors. For the character 4-gram classifier, 10 greatest confusion values are as follows:

Poet	**Classified as**	**% of texts**
M.M. Kheraskov	V.I. Maykov	24.14
M.M. Kheraskov	A.P. Sumarokov	20.69
V.I. Maykov	M.V. Lomonosov	20.48
A.P. Sumarokov	V.I. Maykov	19.07
V.A. Zhukovsky	I.I. Dmitriev	16.04
N.A. Nekrasov	A.N. Apukhtin	14.24
A.S. Pushkin	N.M. Yazykov	11.47
A.K. Tolstoy	A.N. Apukhtin	11.43
I.P. Myatlev	A.N. Apukhtin	11.11
D.S. Merezhkovsky	S.Ya. Nadson	10.44

Table 3. Pairs of authors with greatest relative confusion values
(character 4-gram classifier)

Table 3 shows that classification results are especially unreliable for poets from the same period. In other words, poets from the same period are the ones that are deemed to be similar to each other, which is not counterintuitive. The results coming from the word-based classifier are not much different, which confirms that this method of finding pairs of similar authors provides reliable results:

Poet	**Classified as**	**% of texts**
V.A. Zhukovsky	I.I. Dmitriev	30.77
N.A. Nekrasov	A.N. Apukhtin	30.30
V.V. Mayakovsky	B.P.Kornilov	25.57
M.M. Kheraskov	V.I. Maykov	24.14
A.S. Pushkin	N.M. Yazykov	22.02
A.P. Sumarokov	V.I. Maykov	21.65
V.Ya. Brysuov	P.F. Yakubovich	21.03
V.I. Maykov	M.V. Lomonosov	20.48
A.K. Tolstoy	A.N. Apukhtin	19.04
M.M. Kheraskov	A.P. Sumarokov	18.97

Table 4. Pairs of authors with greatest relative confusion values
(word-based classifier)

4. Attribution errors and the history of Russian poetry

The errors made by the character 4-gram classifier can be visualized as a graph. In this graph, 69 vertices correspond to the poets included in the sample. A pair of vertices (*X*, *Y*) is connected with an edge if at least 3% of poems written by X are assigned to Y or vice versa. The resulting graph is shown in Figure 1. 12 poets (including Khlebnikov, Pasternak, and Tsvetaeva) are identified as not being similar to anyone else in particular, and the remaining 57 authors form a connected component.

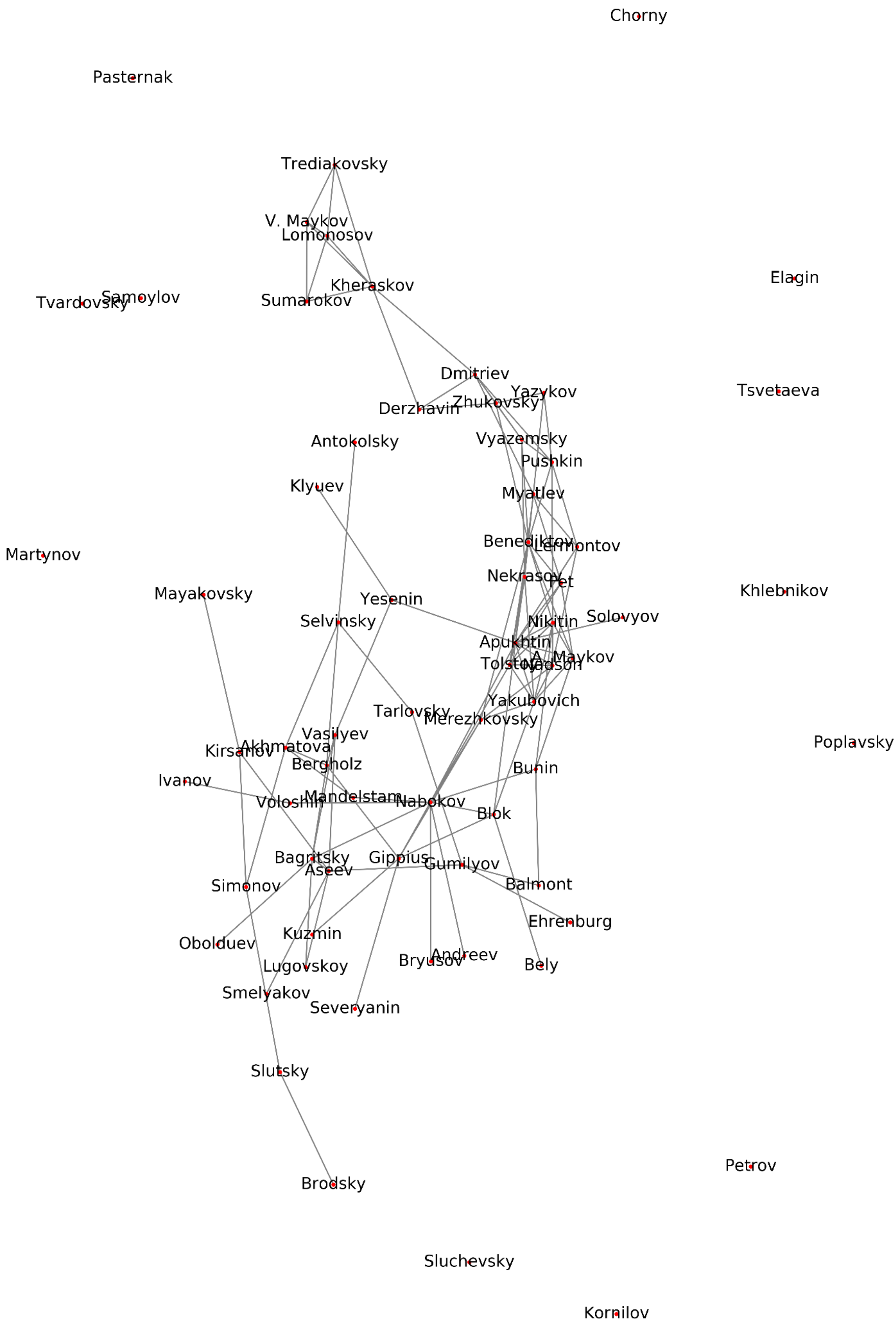

Figure 1. Similarity of Russian poets
based on the confusion matrix of the character 4-gram classifier.

This graph demonstrates that the confusion matrix resulting from the classification based on character 4-grams can be converted into a fairly convincing graph depicting the history of Russian poetry. It shows a picture that is justified from a chronological point of view: the biggest age difference between two poets connected by an edge is 55 years (namely, between Apukhtin, who was born in 1840, and Yesenin, born in 1895), whereas in most cases poets are linked to their close contemporaries. Some adjacencies in the graph correspond to facts known from other sources, e.g.:

- Benediktov occupies a very central position in the graph, which highlights his important role in the literature of the 19th century in spite of the fact that he was almost forgotten later;
- Mayakovsky had a strong influence on Kirsanov and Aseev;
- Yesenin and Klyuev were the most prominent representative of the peasant poetry at the beginning of the 20th century;
- Tarlovsky was very fond of Gumilyov, to whom he privily dedicated a collection of poems called *Bumerang*, which was a very brave move in 1930’s (see notes in Tarlovsky 2009);
- Balmont and Gumilyov are known for the prevalence of exotic topics in their poetry;
- Blok and Bely are the key figures of Russian Symbolism.

This list can be continued further, and this indicates that it might be fruitful to analyze the links that are shown on the graph but whose origin is not so evident; a qualitative analysis of these links might provide interesting insights into the history of Russian poetry.

5. Text distortion for identifying the principal components of style

As mentioned in the Introduction, authorship attribution performs best with character n-grams, but it is far from clear what kind of information is captured by well-performing classifiers. In order to find this out, text distortion is most commonly used (Kestemont 2014; Sundararajan & Woodard 2018).

In the case of Russian poetry, there may be various kinds of information hidden within character 4-grams that turn out to be important for authorship attribution. This information may be linked to phonology (or, in other words, euphony), to morphology, or to lexical meaning expressed by stems. In order to test which is more important for authorship attribution, texts used for training were subjected to the following distortion procedure: in 20% of words longer than one character, two randomly selected adjacent characters were swapped. There were three distortion conditions: 1) a pair of adjacent characters may be swapped anywhere in the word; 2) a pair of adjacent characters may be swapped in the left-hand half of the word only; 3) a pair of adjacent characters may be swapped in the right-hand half of the word only. In two-letter words, the two letters were swapped in any case; in longer words with an odd number of character bigrams (i.e., with an even number of characters), the medial bigram was randomly assigned either to the first or to the second part of the word. Some examples of distortion under these conditions are given in Table 5:

Distortion condition	**Text**
None	*Kak gosudarstvo bogateet, I čem živět, i počemu Ne nužno zolota emu*
Anywhere	*Kak g**so**udarstvo bogateet, I čem živět, i počemu Ne nužno zolo**at** emu*
Left	*Kak gosudarstvo **ob**gateet, I čem živět, i p**čo**emu Ne nužno zolota emu*

Right	*Kak gosudarst**ov** bogateet, I čem živ**të**, i počemu Ne nužno zolota emu*

Table 5. Examples of distortion under various condition

One may assume that the distortion on the left-hand side of the word is especially harmful to lexical meaning; distortion on the right-hand side is especially harmful to morphology; distortion anywhere in the word is less harmful to both of these and has a substantial effect on euphony. The accuracy of the character 4-gram classifiers after the application of various types of distortion is presented in Table 6:

Distortion condition	**Accuracy**
None	72.03%
Anywhere	66.26%
Left	67.54%
Right	67.37%

Table 6. Accuracy for character 4-gram classifier after distortion.

These results are difficult to interpret. Probably, euphony plays the most important role for the character 4-gram classifier and is the most characteristic feature of individual poetic style, given that distortion in any position in the word leads to the worst classification performance. However, this conclusion must be tested further, since the relationship between the three distortion conditions remains far from clear.

6. Conclusion

The results of the experiment presented in this paper show that classification methods used for authorship attribution can serve not only this very purpose, but also a more general purpose of measuring stylistic idiosyncraticity and similarity between authors. An interesting interpretable result obtained from the comparison of the character 4-gram classifier and the word-based classifier shows that some poets (including Akhmatova and Brodsky) are characterized by idiosyncratic vocabulary, whereas other poets (including Pasternak and Mayakovsky) are characterized by idiosyncratic euphony. Confusion matrices also make it possible to study similarity between authors and represent it as a graph. It is shown that a Naïve Bayes classifier often treats poets from the same epoch as similar, which agrees with intuitive feelings about possible sources of similarity. An attempt at finding out what plays the most important role in identifying authorship was also made, with a tentative conclusion that a Naïve Bayes classifier based on character 4-grams pays special attention to euphony.

References

1. Burrows, J. (2002). 'Delta': a measure of stylistic difference and a guide to likely authorship. *Literary and linguistic computing*, *17*(3), 267-287.
2. Coulthard, M. (2013). On Admissible Linguistic Evidence. *Journal of Law and Policy*, *21*(2).
3. Ge, Z., Sun, Y., & Smith, M.J. T. (2016). Authorship Attribution Using a Neural Network Language Model. *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI'16).* ArXiv:1602.05292 [Cs.CL]
4. Holmes, D.I. (1994). Authorship attribution. *Computers and the Humanities*, *28*(2), 87–106. https://doi.org/10.1007/BF01830689

5. Houvardas, J., & Stamatatos, E. (2006). N-Gram Feature Selection for Authorship Identification. In J. Euzenat & J. Domingue (Eds.), *Artificial Intelligence: Methodology, Systems, and Applications* (pp. 77–86). Berlin, Heidelberg: Springer.
6. Juola, P. (2008). Authorship attribution. *Foundations and Trends in Information Retrieval*, *1*(3), 233-334.
7. Juola, P. (2013). How a computer program helped reveal J.K. Rowling as author of A Cuckoo's Calling. *Scientific American*. Aug. 20.
8. Kešelj, V., Peng, F., Cercone, N., & Thomas, C. (2003). N-gram-based Author Profiles for Authorship Attribution. *Proceedings of the Pacific Association for Computational Linguistics Conference*.
9. Kestemont, M., Daelemans, W., & Sandra, D. (2012). Robust Rhymes? The Stability of Authorial Style in Medieval Narratives. *Journal of Quantitative Linguistics*, *19*(1), 54–76. https://doi.org/10.1080/09296174.2012.638796
10. Kestemont, M. (2014). Function Words in Authorship Attribution. From Black Magic to Theory? *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*, 59–66. https://doi.org/10.3115/v1/W14-0908
11. Kjell, B., Woods, W.A., & Frieder, O. (1994). Discrimination of authorship using visualization. *Information Processing & Management*, *30*(1), 141–150. https://doi.org/10.1016/0306-4573(94)90029-9
12. Koppel, M., Schler, J., & Argamon, S. (2009). Computational methods in authorship attribution. *Journal of the American Society for information Science and Technology*, *60*(1), 9-26.
13. Mikheev M., Erlikh L. (2017). Connectors frequencies as a distinctive sign of the individual style (in view of the couple Fomenko hypothesis). *Computational Linguistics and Intellectual Technologies. Papers from the Annual International Conference "Dialogue" (2017)*. Issue 16.
14. Luyckx, K. (2011). *Scalability Issues in Authorship Attribution*. Asp / Vubpress / Upa.
15. Mosteller, F., & Wallace, D. (1964). *Inference and Disputed Authorship: The Federalist*. Reading, MA: Addison-Wesley.
16. Neal, T., Sundararajan, K., Fatima, A., Yan, Y., Xiang, Y., & Woodard, D. (2017). Surveying Stylometry Techniques and Applications. *ACM Comput. Surv.*, *50*(6), 86:1–86:36. https://doi.org/10.1145/3132039
17. Neidorf, L., Krieger, M.S., Yakubek, M., Chaudhuri, P., & Dexter, J.P. (2019). Large-scale quantitative profiling of the Old English verse tradition. *Nature Human Behaviour*, 1. https://doi.org/10.1038/s41562-019-0570-1
18. Rocha, A., J. Scheirer, W., W. Forstall, C., Cavalcante, T., Theophilo, A., Shen, B., Carvalho, A.R.B., & Stamatatos, E. (2016). Authorship Attribution for Social Media Forensics. *IEEE Transactions on Information Forensics and Security*, *12*, 5. https://doi.org/10.1109/TIFS.2016.2603960
19. Sari, Y., Stevenson, M., & Vlachos, A. (2018). Topic or Style? Exploring the Most Useful Features for Authorship Attribution. *Proceedings of the 27th International Conference on Computational Linguistics*, 343–353. Retrieved from https://www.aclweb.org/anthology/C18-1029
20. Savoy, J. (2017). Elena Ferrante Unmasked. *Drawing Elena Ferrante's Profile: Workshop Proceedings*. Padova, 7 September 2017. Padova: Padova UP.P. 123–143.
21. Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for information Science and Technology*, *60*(3), 538-556.

22. Stamatatos, E. (2018). Masking topic-related information to enhance authorship attribution. *Journal of the Association for Information Science and Technology*, *69*(3), 461–473. https://doi.org/10.1002/asi.23968
23. Sundararajan, K., & Woodard, D. (2018). What represents “style” in authorship attribution? *Proceedings of the 27th International Conference on Computational Linguistics*, 2814–2822. https://www.aclweb.org/anthology/C18-1238
24. Tarlovsky, M. (2009). *Molčalivyj polët* [A silent flight]. Introduction and notes by E. Vitkovskyij and V. Rezvyj. Moscow: Vodolej.
25. Vilariño Ayala, D., Castillo, E., Pinto, D., Leon, S., & Mireya, T. (2011). Baseline Approaches for the Authorship Identification Task. *CLEF 2011 Labs and Workshop, Notebook Papers*.

Appendix

Author	Texts	Tokens		Author	Texts	Tokens
V.A. Zhukovsky	672	248750		L.N. Martynov	675	65383
A.S. Pushkin	905	194733		Ya.V. Smelyakov	372	64662
V.Ya. Bryusov	1684	185856		P.N. Vasilyev	245	63377
N.A. Nekrasov	447	172322		I.S. Nikitin	232	63069
M.I. Tsvetaeva	1469	170661		S.I. Kirsanov	441	62533
V.V. Mayakovsky	635	135289		V.V. Nabokov	582	62504
M.Yu. Lermontov	475	132972		N.N. Aseev	376	62067
V.I. Ivanov	1181	124346		N.M. Yazykov	359	60372
B.A. Slutsky	1288	124032		I.P. Myatlev	115	60227
A.K. Tolstoy	302	119909		O.E. Mandelstam	680	59844
Sasha Chorny	706	116620		S.A. Yesenin	428	58517
A.T. Tvardovsky	347	114348		I.A. Brodsky	278	58374
A.N. Maykov	563	113711		G.N. Oboldduev	388	57891
A.A. Blok	1351	113111		I.V. Elagin	434	57752
K.D. Balmont	1005	108098		S.Ya. Nadson	491	57466
K.K. Sluchevsky	736	95245		I. Severyanin	638	57362
G.R. Derzhavin	416	93503		M.M. Kheraskov	65	56707
N.A. Klyuev	573	93493		A.A. Akhmatova	946	56504
D.L. Andreev	453	92628		O.F. Bergholz	393	55679
M.A. Kuzmin	838	91851		Z.N. Gippius	475	55622
I.L. Selvinsky	459	84631		I.I. Dmitriev	392	55605
D.S. Merezhkovsky	339	83807		A.N. Apukhtin	355	55046
K.M. Simonov	283	82275		E.G. Bagritsky	203	54542
I.A. Bunin	768	80839		Andrei Bely	506	54389
A.A. Fet	921	80042		V.V. Khlebnikov	291	54270
B.L. Pasternak	532	77868		M.V. Lomonosov	146	53310
A.P. Sumarokov	284	76534		V.A. Lugovskoy	151	53309
S.M. Solovyov	433	75347		B.Yu. Poplavsky	548	53073
D. Samoylov	901	73358		V.I. Maykov	113	52326
I.G. Ehrenburg	711	70441		M.A. Tarlovsky	316	51343
V.K. Trediakovsky	163	69050		M.A. Voloshin	300	51341
N.S. Gumilyov	524	67511		B.P. Kornilov	189	51288

P.A. Vyazemsky	340	67452		P.F. Yakubovich	367	50590
S.V. Petrov	609	66402		V.G. Benediktov	272	50556
P.G. Antokolsky	371	65976				

Table A1. Authors represented by more than 50,000 tokens in the Poetic subcorpus of the Russian National Corpus.

	Author	% Correctly Attributed			Author	% Correctly Attributed
1	S.V. Petrov	96.10		36	M.Yu. Lermontov	73.38
2	B.P. Kornilov	95.40		37	N.S. Gumilyov	72.92
3	V.V. Khlebnikov	93.49		38	I.G. Ehrenburg	72.50
4	P.G. Antokolsky	89.11		39	V.V. Nabokov	71.97
5	B.Yu. Poplavsky	88.95		40	P.A. Vyazemsky	71.61
6	G.R. Derzhavin	88.54		41	V.I. Maykov	71.43
7	V.V. Mayakovsky	87.56		42	M.A. Tarlovsky	70.08
8	N.M. Yazykov	87.38		43	G.N. Obolduev	69.46
9	A.T. Tvardovsky	87.20		44	S.A. Yesenin	69.23
10	M.V. Lomonosov	87.07		45	V.Ya. Bryusov	68.61
11	S.Ya. Nadson	86.97		46	M.A. Voloshin	67.62
12	Ya.V. Smelyakov	86.03		47	O.F. Bergholz	66.56
13	N.A. Klyuev	85.69		48	A.P. Sumarokov	66.49
14	P.F. Yakubovich	84.74		49	O.E. Mandelstam	66.44
15	K.K. Sluchevsky	84.72		50	I. Severyanin	65.86
16	K.D. Balmont	84.23		51	I.A. Bunin	65.51
17	Andrei Bely	82.25		52	M.I. Tsvetaeva	65.22
18	V.K. Trediakovsky	82.14		53	A.A. Blok	62.59
19	V.G. Benediktov	81.07		54	P.N. Vasilyev	60.40
20	B.A. Slutsky	80.91		55	A.A. Fet	59.79
21	S.M. Solovyov	80.20		56	M.A. Kuzmin	59.50
22	B.L. Pasternak	80.05		57	I.P. Myatlev	57.58
23	I.A. Brodsky	79.91		58	D. Samoylov	56.49
24	I.I. Dmitriev	79.89		59	V.A. Zhukovsky	56.48
25	D.L. Andreev	79.49		60	I.S. Nikitin	54.50
26	V.I. Ivanov	79.23		61	D.S. Merezhkovsky	53.41
27	E.G. Bagritsky	78.35		62	N.N. Aseev	48.76
28	Sasha Chorny	78.31		63	A.N. Maykov	47.59
29	K.M. Simonov	78.13		64	A.S. Pushkin	46.79
30	I.V. Elagin	77.14		65	N.A. Nekrasov	44.24
31	V.A. Lugovskoy	76.39		66	I.L. Selvinsky	43.41
32	A.A. Akhmatova	75.69		67	S.I. Kirsanov	35.51
33	L.N. Martynov	74.94		68	M.M. Kheraskov	27.59
34	A.N. Apukhtin	74.66		69	A.K. Tolstoy	22.38
35	Z.N. Gippius	74.24				

Table A2. Recognizability of poets (character 4-gram classifier).

	Author	% Correctly Attributed			Author	% Correctly Attributed
1	B.P. Kornilov	93.10		36	V.V. Nabokov	52.53
2	B.Yu. Poplavsky	89.78		37	M.Yu. Lermontov	51.54
3	N.M. Yazykov	86.71		38	K.D. Balmont	51.41
4	V.V. Khlebnikov	86.39		39	L.N. Martynov	50.35
5	M.V. Lomonosov	84.48		40	I. Severyanin	50.00
6	S.V. Petrov	83.55		41	P.A. Vyazemsky	49.15
7	S.Ya. Nadson	83.10		42	N.S. Gumilyov	45.13
8	Andrei Bely	82.51		43	N.A. Klyuev	45.04
9	P.F. Yakubovich	82.47		44	G.N. Obolduev	44.97
10	I.A. Brodsky	81.66		45	I.G. Ehrenburg	43.13
11	I.I. Dmitriev	81.48		46	D.L. Andreev	42.17
12	V.G. Benediktov	79.42		47	V.I. Ivanov	40.17
13	A.A. Akhmatova	79.22		48	N.N. Aseev	39.13
14	V.A. Lugovskoy	79.17		49	M.M. Kheraskov	36.21
15	Z.N. Gippius	78.98		50	A.A. Fet	32.58
16	V.I. Maykov	77.55		51	D.S. Merezhkovsky	32.53
17	E.G. Bagritsky	76.80		52	M.A. Tarlovsky	32.28
18	V.K. Trediakovsky	76.79		53	M.A. Kuzmin	25.86
19	A.N. Apukhtin	76.03		54	O.È. Mandelstam	25.50
20	M.A. Voloshin	74.59		55	Sasha Chorny	24.50
21	Ya.V. Smelyakov	72.91		56	S.I. Kirsanov	24.15
22	P.G. Antokolsky	72.78		57	A.A. Blok	23.09
23	S.A. Yesenin	71.68		58	B.A. Slutsky	19.50
24	K.M. Simonov	69.20		59	B.L. Pasternak	19.27
25	I.V. Elagin	68.57		60	D. Samoylov	17.54
26	A.T. Tvardovsky	67.82		61	V.Ya. Bryusov	17.50
27	G.R. Derzhavin	66.87		62	V.A. Zhukovsky	13.41
28	O.F. Bergholz	63.25		63	A.K. Tolstoy	12.86
29	I.S. Nikitin	63.00		64	V.V. Mayakovsky	9.84
30	K.K. Sluchevsky	58.97		65	A.N. Maykov	8.50
31	S.M. Solovyov	58.12		66	I.L. Selvinsky	7.72
32	A.P. Sumarokov	57.73		67	A.S. Pushkin	5.73
33	I.A. Bunin	56.02		68	M.I. Tsvetaeva	3.06
34	I.P. Myatlev	55.56		69	N.A. Nekrasov	0.00
35	P.N. Vasilyev	55.45				

Table A3. Recognizability of poets (word-based classifier).