

Б.В. Ковалев*Санкт-Петербургский государственный университет**Санкт-Петербург, Россия*

УДК 51-78

DOI 10.35785/2782-4195-2024-2-44-49

Новые тексты Б.Л. Васильева: атрибуция методом Delta Берроуза

Ключевые слова: Delta Берроуза; атрибуция; цифровая текстология; Б.Л. Васильев; стилеметрия.

В статье производится стилеметрический анализ рукописей, найденных на жестком диске русского писателя Б.Л. Васильева. Основным инструментом выступила Delta Берроуза – наиболее распространенный и надежный метод, позволяющий оценить стилистическую близость текстов. В качестве сопоставительного материала для эксперимента привлекались романы П.М. Алешковского, три части романа И.Я. Болгарина и В.В. Смирнова «Девять жизней Нестора Махно», тексты Л.А. Юзефовича, а также произведения самого Б.Л. Васильева, опубликованные при жизни автора. Результаты экспериментов позволяют сформулировать гипотезу, подтвердить или опровергнуть которую предстоит текстологам и историкам литературы: найденные рукописи созданы Б.Л. Васильевым.

В 2024 году в распоряжении В.А. Карнюшина – крупнейшего исследователя творчества и наследия Б.Л. Васильева – оказался жесткий диск, принадлежавший писателю. На носителе было обнаружено некоторое количество незаконченных и не публиковавшихся прежде текстов. В этой связи актуализировалась задача атрибуции этих текстов при помощи количественных методов.

Для решения задачи атрибуции новообнаруженных текстов Б.Л. Васильева при помощи лингвостатистических средств была выбрана Delta Берроуза [Burrows, 2002] – один из наиболее надежных, авторитетных и апробированных стилеметрических инструментов на сегодняшний день. Метод работает в предположении, что распределение наиболее частотных слов в тексте отражает индивидуальный авторский стиль писателя – так называемый авторский сигнал. Вычисления осуществляются следующим образом. Для каждого из анализируемых текстов нужно взять некоторое количество самых частотных слов (в первую очередь – служебных), а затем сравнить их частотности (представленные как мера z-score, учитывающая число всех слов в тексте и показывающая разброс значений относительно средних). Соответственно, Delta – это сумма взятых по модулю разниц между мерами z-scores у двух сравниваемых текстов, поделенная на количество слов. Так вычисляется «расстояние» между текстами; чем меньше выявленное «расстояние» между текстами, тем более стилистически близкими они являются.

Среди прочего, метод Delta используется как инструмент в текстологических исследованиях. Метод апробировался на материале разных языков:

например, арабского [AbdulRazzaq, Mustafa, 2014], польского [Eder, 2011], испанского [Calvo Tello, 2019; Ковалев, 2024], английского [Gladwin, Lavin et al., 2017].

В последние годы Delta также активно применяется на русскоязычном материале [Метод Дельты Берроуза, 2018; Об авторстве, 2019; Орехов, Великанова, 2019; Iosifyan, Vlasov, 2020]. Проводились сравнительно-сопоставительные замеры на материале русской межвоенной прозы [Двинятин, Ковалев, 2024].

Следует учитывать, что Delta Берроуза, как и всякий лингвостатистический инструмент, не обладает самостоятельной эвристической силой [В поисках потерянного автора, 2001, б]. Он служит для формирования атрибуционной гипотезы, а также для выявления некоторой стилистической близости между текстами и должен использоваться вкуче с историко-литературными и архивными разысканиями. Только в этом случае можно говорить о полноценной процедуре атрибуции, что мы и попытались предпринять в рамках исследования.

Для корректного осуществления экспериментов необходимо сформировать корпус текстов. В него мы поместили три укрупненных класса произведений: а) романы, опубликованные Б.В. Васильевым в последний период жизни; б) атрибутируемые тексты; в) сопоставительные тексты, априори созданные иными авторами, на похожую тему и в тот же хронологический промежуток.

Таким образом, в анализируемый корпус вошли опубликованные тексты Б.Л. Васильева: «Владимир Красное Солнышко», «Глухомань», «Дом, который построил дед», «И был вечер, и было утро», «Красные жемчуга», «Отрицание отрицания», «Ровесница века», «Скобелев», «Утоли моя печали».

Атрибутируемые рукописи: «Мутные бремена», «И усидеть на кончике ножа», «Откровение», «Рукопись града Китежа», «Рыцарь окровавленной богини».

А также в корпус были помещены следующие тексты как сопоставительные: «Арлекин» и «Крепость» П.М. Алешковского, три части романа И.Я. Болгарина и В.В. Смирнова «Девять жизней Нестора Махно» и тексты Л.А. Юзефовича «Журавли и карлики» и «Зимняя дорога».

Все эксперименты проводились в программной среде R в рамках пакета Stylo [Eder, Rybicki, Kestemont, 2016]. Атрибутируемые тексты были обозначены в корпусе как Dubia.

Существуют разные мнения относительно количества частотных слов, на которых следует проводить опыты. В данном случае мы решили провести эксперименты как на малом словаре, 200–300 наиболее частотных слов (most frequent words, далее – mfw), так и на большом – 1000 mfw.

При объеме частотного словаря 200–300 mfw – учитываются преимущественно служебные слова, тематически немаркированная лексика – четыре обнаруженные рукописи кластеризуются в одну большую ветвь текстов Б.Л. Васильева. Наиболее стилистически близкими оказываются «Красные жемчуга», опубликованные при жизни автора, и «Рукопись града Китежа». Одна-

ко «Откровение» приближается к текстам Юзефовича. То есть из всех текстов корпуса «Откровению» близки не другие тексты Васильева, а два романа Юзефовича (рис. 1). Юзефович, разумеется, не писал «Откровение», о чем свидетельствуют архивные разыскания, но важно отметить, что алгоритм выявляет стилистическое отличие этого текста от других васильевских.

Кроме того, было эмпирически установлено, что Delta надежнее работает с текстами, в которых более 5000 слов. «Откровение» же является одним из самых коротких текстов в корпусе (4455 слов). Этим также может объясняться некоторая обособленность текста.

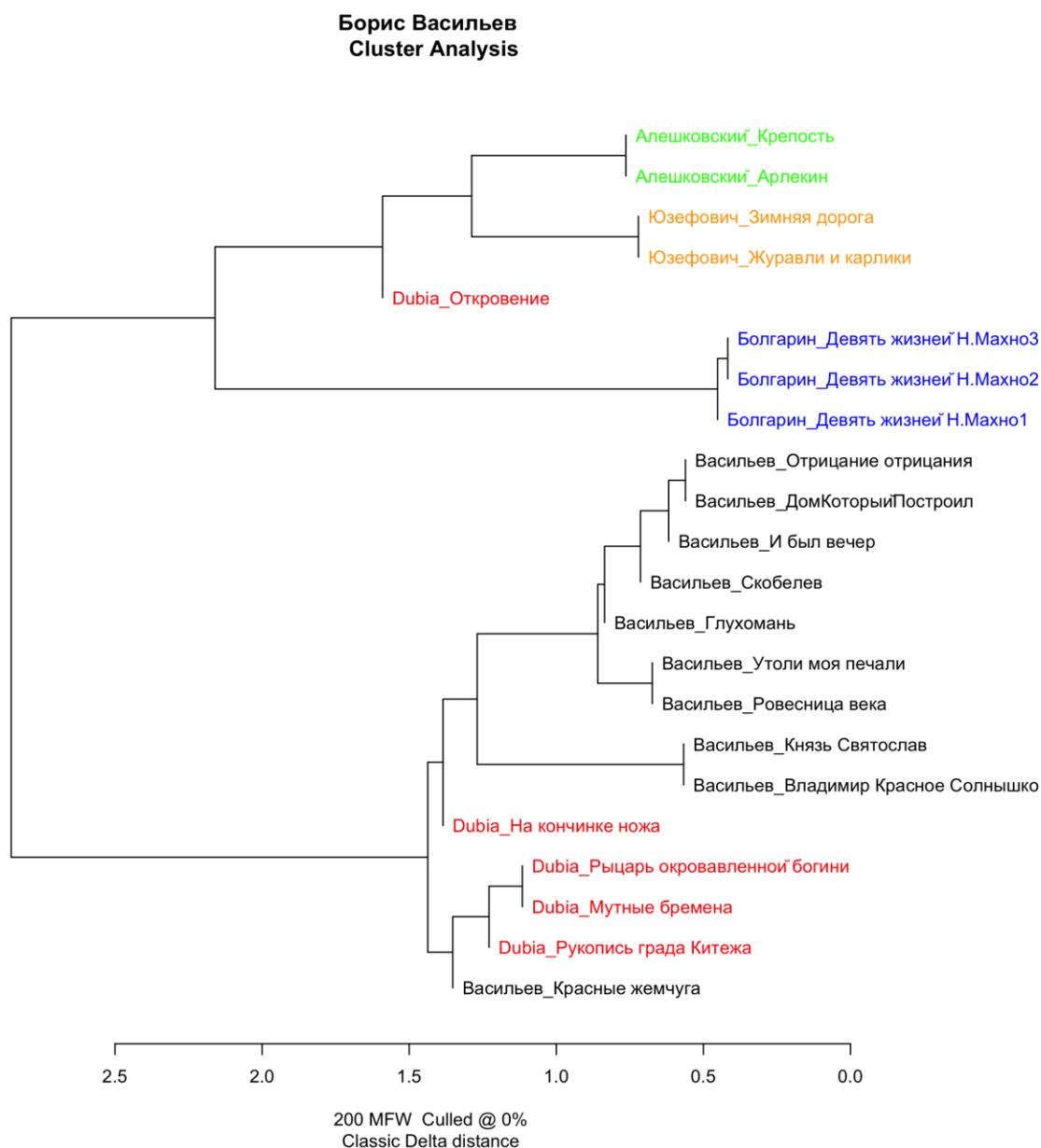


Рис. 1. Иерархическая кластеризация текстов Б. Васильева. 200 mfw

При объеме частотного словаря в 1000 наиболее частотных слов группировка другая: «Откровение» образует отдельную ветвь с другими рукопи-

сями, но «Рыцарь окровавленной богини» перемещается к дописанным романам (рис. 2).

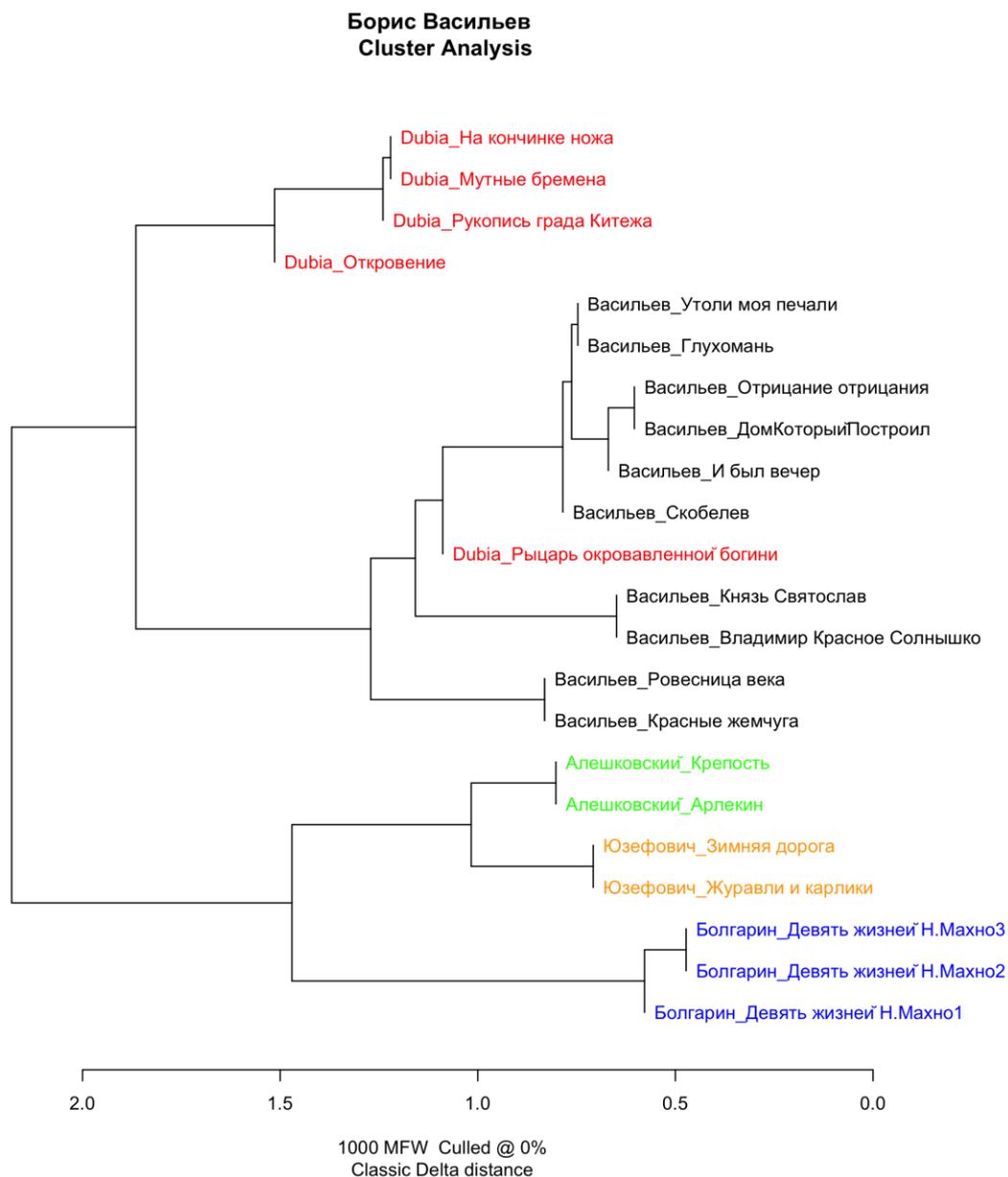


Рис. 2. Иерархическая кластеризация текстов Б. Васильева. 1000 mfw

Таким образом, Delta Берроуза подтверждает факт, что обнаруженные тексты были написаны Б.Л. Васильевым; характерно, что «Откровение» стилистически отличается от остальных текстов. Так при помощи количественных методов еще раз подтверждается читательское впечатление и наблюдение, сделанное посредством традиционного литературоведческого анализа. Тем не менее следует подчеркнуть, что при всей отмеченной разнице «Откровение» находится в орбите текстов Васильева.

Еще раз напомним, что Delta Берроуза служит лишь для формирования атрибуционной гипотезы, далее – слово за текстологами и историками лите-

ратуры. Однако, согласно результатам замеров, гипотеза формулируется следующая: все новообнаруженные тексты принадлежат Б.Л. Васильеву.

Литература

В поисках потерянного автора: Этюды атрибуции / М.А. Марусенко [и др.]. СПб.: Филологический факультет СПбГУ, 2001. 211 с.

Великанова Н.П., Орехов Б.В. Цифровая текстология: атрибуция текста на примере романа М.А. Шолохова «Тихий Дон» // Мир Шолохова. Научно-просветительский общенациональный журнал. 2019. № 1. С. 70–82.

Двинятин Ф.Н., Ковалев Б.В. Романы Владимира Набокова в контексте русской межвоенной прозы: стилеметрический аспект // Новый филологический вестник. 2024. № 1(68). С. 203–216.

Ковалев Б. В. Рождение третьего автора: стилеметрический анализ рассказов Онорио Бустоса Домека // Литература двух Америк. 2024. № 16. С. 120–146.

Метод Дельты Берроуза для определения авторства анонимных и псевдоанонимных литературных произведений на русском языке / Н.К. Мамаев [и др.] // Proceedings of the R. Piotrowski's Readings in Language Engineering and Applied Linguistics. СПб.: RWTH Aachen University, 2018. С. 1–14.

Об авторстве «писем Берии из заточения» / В.В. Петров [и др.] // Вестник СПбГУ. Право. 2019. Т. 10. № 3. С. 586–605.

AbdulRazzaq A.A., Mustafa T.K. Burrows Delta method fitness for Arabic text authorship stylometric detection // International Journal of Computer Science and Mobile Computing. 2014. № 6. P. 69–78.

Burrows J. Delta: a measure of stylistic difference and a guide to likely authorship // Literary and Linguistic Computing. 2002. № 3(17). P. 267–287.

Calvo Tello J. Delta inside Valle-Inclán: stylometric classification of periods and groups of his novels // Romanische Studien. 2019. № 6. P. 151–163.

Eder M. Style-Markers in authorship attribution. A cross-language study of the authorial fingerprint // Studies in Polish Linguistics. 2011. № 6. P. 99–114.

Eder M., Rybicki J., Kestemont M. Stylometry with R: a package for computational text analysis // R Journal. 2016. № 1(8). P. 107–121.

Gladwin A., Lavin M., Look D. Stylometry and collaborative authorship: Eddy, Lovecraft, and The Loved Dead // Digital Scholarship in the Humanities. 2017. № 1(32). P. 123–140.

Iosifyan M., Vlasov I. And Quiet Flows the Don: the Sholokhov-Kryukov authorship debate // Digital Scholarship in the Humanities. 2020. № 2(35). P. 307–318.

B. V. Kovalev

St. Petersburg State University

St. Petersburg, Russia

New Texts of B. L. Vasiliev: Attribution Using Burrows's Delta Method

Keywords: Burrows's Delta; attribution; digital textology; B.L. Vasiliev; stylometry.

The article provides a stylistic analysis of manuscripts found on the hard drive of the remarkable Russian writer B. L. Vasiliev. The main tool was Burrows's Delta as the most common and reliable method for assessing the stylistic similarity of texts. The comparative material for the experiment included novels written by P.

M. Aleshkovsky; three parts of the novel "Nine Lives of Nestor Makhno" (I.Ya. Bolgarin and V.V. Smirnov), texts of L.A. Yuzefovich, as well as works written by B.L. Vasiliev himself, published during the author's lifetime. The results of the experiments allow us to formulate a hypothesis, which textual critics and literary historians will have to confirm or refute: the found manuscripts were created by B.L. Vasiliev.

References

AbdulRazzaq A.A., Mustafa T.K. Burrows Delta method fitness for Arabic text authorship stylometric detection. *International Journal of Computer Science and Mobile Computing*, 2014, no. 6, pp. 69–78. (in English)

Burrows J. Delta: a measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing*, 2002, no. 3(17), pp. 267–287. (in English)

Calvo Tello J. Delta inside Valle-Inclán: stylometric classification of periods and groups of his novels. *Romanische Studien*, 2019. no. 6, pp. 151–163. (In Spanish)

Dvinyatin, F.N., Kovalev B.V. Romany Vladimira Nabokova v kontekste russkoi mezhoennoi prozy: stilemetricheskii aspekt [Novels of Vladimir Nabokov in the Context of Russian Interbellum Prose: Stylometric Aspect]. *Novyi filologicheskii vestnik*, 2024, no. 1, pp. 203–216. (in Russian)

Eder M. Style-Markers in authorship attribution. A cross-language study of the authorial fingerprint. *Studies in Polish Linguistics*, 2011, no. 6, pp. 99–114. (in English)

Eder M., Rybicki J., Kestemont M. Stylometry with R: a package for computational text analysis. *R Journal*, 2016, no. 1(8), pp. 107–121. (in English)

Gladwin A., Lavin M., Look D. Stylometry and collaborative authorship: Eddy, Lovecraft, and The Loved Dead. *Digital Scholarship in the Humanities*, 2017, no.1, vol. 32, pp. 123–140. (in English)

Iosifyan M., Vlasov I. And Quiet Flows the Don: the Sholokhov-Kryukov authorship debate. *Digital Scholarship in the Humanities*, 2020, no. 2, vol. 35, pp. 307–318. (in English)

Kovalev B.V. The Birth of the Third Author: Stylometric Analysis of the Stories of Honorio Bustos Domecq. *Literature of the Americas*, 2024, no. 16, pp. 120–146. (in Russia)

Mamayev N.K., Marusenko M.A., Piotrovskaya K.R., Ronzhin A.L. Metod Del'ty Berrouza dlya opredeleniya avtorstva anonimnykh i psevdoanonimnykh literaturnykh proizvedeniy na russkom yazyke [Burrows's Delta for Authorship Attribution of Russian Literary Texts]. Proceedings of the R. Piotrowski's Readings in Language Engineering and Applied Linguistics. St. Petersburg, RWTH Aachen University, 2018, pp. 1–14. (in Russia)

Marusenko M.A., Bessonov B.L., Bogdanova L. M., Anikin M.A, Miasoedova N.E. In search of the lost author: Attribution studies. Sankt-Peterburg, Filologicheskii fakul'tet SPbU Publ., 2001. (in Russia)

Petrov. V.V., Marusenko M.A, Piotrowskaya K.R, Mañas I.N., Mamayev N.K. Ob avtorstve 'pisem Berii iz zatocheniia' [On the authorship of 'Beria's Letters from imprisonment']. *Vestnik SPbGU Pravo*, 2019, no. 3. pp. 586–605. (in Russia)

Velikanova N.P., Orekhov B.V. Cifrovaia tekstologiya: atribuciia teksta na primere romana M.A. Sholohova *Tihii Don* [Digital Textual Criticism: Text Attribution on the Example of the Novel by M.A. Sholokhov *Quiet Don*]. *Mir Sholokhova*, 2019, no. 1, pp. 70–82. (in Russia)