

# From Classics to Digital Philology: On the Origin and Growth of Stylometry\*

*Boris V. Kovalev*

St. Petersburg State University,  
7–9, Universitetskaya nab., St. Petersburg, 199034, Russian Federation;  
<https://orcid.org/0000-0002-1904-1844>, [b.v.kovalev@spbu.ru](mailto:b.v.kovalev@spbu.ru)

**For citation:** Kovalev B. V. From Classics to Digital Philology: On the Origin and Growth of Stylometry. *Philologia Classica* 2024, 19 (2), 347–360. <https://doi.org/10.21638/spbu20.2024.211>

This article is devoted to the history of stylometry and its development at an early stage. Stylometry is an applied philological discipline that considers style as a set of quantitative parameters. Stylometry arose on the material of classical studies. Using the example of the most reliable and authoritative stylometric method to date — Burrows’ Delta — the advantages and disadvantages of this type of analysis are examined. The genesis of the term “stylometry” is established. In the seminal book of G. Martynenko “Fundamentals of Stylometry” it is indicated that it was coined by the German classical philologist W. Dittenberger. The study reveals that the term “stylometry” actually existed in the 19<sup>th</sup> century in the meaning of ‘the art of measuring columns’, and it is not used in Dittenberger’s works. This term was introduced by W. Lutoslawski, who tried to solve the problem of periodisation of Plato’s dialogues. It turns out that “stylometry” was first used in a new meaning on May 21, 1897, during a report by W. Lutoslawski at the Oxford Philological Society. In Russia, the term first appears in a review in 1898 in the form *стилометрия*, and in Morozov’s 1915 article in the form *стилеметрия*, which became widespread in the Soviet academic community.

*Keywords:* stylometry, Lutoslawski, Dittenberger, linguopoetics, stylistics, history of linguistics, textual criticism, Plato.

## 1. Modern stylometry: affordances and limitations

Nowadays, there are several contradictory trends in academic conversations concerning the author of the term “stylometry”. Some authors maintain that the creator of this term is the German philologist W. Dittenberger;<sup>1</sup> others believe that it was the Polish philologist and philosopher W. Lutoslawski who coined it.<sup>2</sup> Sometimes, Lutoslawski is not being identified as the author of the term<sup>3</sup> or not mentioned at all.<sup>4</sup>

In light of this, the goals of our research are multiple. First, to analyze the status and significance of two contenders for the creation of this term: W. Dittenberger and W. Lu-

---

\* This paper is part of the research project 95434615 “Literary texts and their language vs quantitative, corpus and computer methods: mutual testing (Nabokov and comparative material)”, funded by Saint Petersburg State University.

<sup>1</sup> See Martynenko 1988; Zhuravleva 2012; Tuldava 2005; Martynenko 2014; Martynenko, Grebennikov 2018.

<sup>2</sup> Brandwood 1990; Grzybek 2014; Orekhov 2020.

<sup>3</sup> Williams 1970; Herdan 1966.

<sup>4</sup> Holmes 2018.

toslawski. Second, to analyze the roots of stylometry as a discipline formed on the basis of classical studies, and third and last, to trace the history of the appearance of the term “stylometry” and some specifics of its perception.

G. Martynenko defines modern stylometry as an applied philological discipline that deals with measuring stylistic characteristics for the purpose of systematization (attribution, taxonomy, periodization, dating, etc.) of texts and their parts.<sup>5</sup> We would add that stylometry is a discipline that considers style as a set of quantitative parameters. Its main principle is that the text can be transferred from the text level to the level of a mathematical model, therefore, we can calculate statistically significant distributions and patterns, which indicate the individual stylistic characteristics of a particular author.

Stylometry originated in the second half of the 19<sup>th</sup> century. It should be emphasized that the main material for researchers was Plato’s dialogues.<sup>6</sup> Mathematically verified, based on a quantitative approach, stylometry grew from the need to analyze ancient Greek texts in a new, more efficient way. However, before describing the history of this term, it is necessary to make a general explanation of what stylometry is today.

Nowadays, the most authoritative stylometric method is Burrows’ Delta. Delta is just one method among others but it is the most popular and it allows to reveal advantages and disadvantages of modern stylometric instruments. It is based on the assumption that the distribution of the most frequent words (prepositions, conjunctions, pronouns) is an imprint of the individual author’s style. For each text in the analyzed corpus, a certain number of the most frequent words is taken, and then their frequencies are compared, presented as z-scores. Z-score is calculated using the following formula:

$$z(f_i(D_1)) = \frac{f_i(D_1) - \mu_i}{\sigma_i}, \quad (1)$$

where  $f_i(D_1)$  is the word frequency in text  $D_1$ ,  $\mu_i$  is the average word frequency in the sample, and  $\sigma_i$  is the standard deviation of this frequency. Correspondingly, Burrows’ Delta is the mean of the absolute differences between the z-scores for a set of word-variables in a given text-group and the z-scores for the same set of word-variables in a target text.<sup>7</sup>

In this way the distance between two units is calculated. The distance is greater for texts written by different authors than for the texts written by the same author. Delta is used to solve the tasks on classification, periodization and the attribution of texts. It is important to emphasize that Delta is actively used for analyzing ancient texts in Latin<sup>8</sup> and Ancient Greek.<sup>9</sup> We also should mention some works related to the study of Apuleius.<sup>10</sup> The stylometric approach is not the principal one in modern classical studies, but this direction is actively developing.<sup>11</sup>

Delta has proven its reliability and effectiveness in various languages. It was determined that the texts written under the pseudonym R. Galbraith belonged to J. K. Rowling

---

<sup>5</sup> Martynenko 1988, 54.

<sup>6</sup> Dittenberger 1881; Lutoslawski 1897d.

<sup>7</sup> Burrows 2002, 271.

<sup>8</sup> See Rybicki *et al.* 2011; Shumilin 2021.

<sup>9</sup> See Storey, Mimno 2020; Alieva 2022.

<sup>10</sup> Stover 2016; Stover, Koppel *et al.* 2016; Stover, Kestemont 2017; Nikolaev, Shumilin 2021.

<sup>11</sup> See also other works using different stylometric tools: Gianitsos *et al.* 2019; Martins *et al.* 2021.

which was later confirmed by the writer herself.<sup>12</sup> The classification of novels of Valle-Inclan<sup>13</sup> and translations of Nabokov<sup>14</sup> has been successfully carried out as well.

However, there is a number of restrictions. The compared texts should be of the same genre and include more than 5,000 words. In addition, there is no consensus on the volume of the frequency dictionary that should be used for measuring. Some researchers use from 100 to 500 most frequent words, researchers from Würzburg use up to 5,000 mfw. Additionally, the accuracy of the measurements depends on the number of texts in the corpus.

We have to admit that none of the proposed methods can determine the author of a particular text with 100% accuracy. The latest computer tools (such as Delta) are not trustworthy in the matter of attribution and do not have independent heuristic power. They help to form an attribution hypothesis, which then needs to be confirmed at an historical level. As before, the most reliable way to attribute a text to an author is to search for historical documents and carry out an analysis from which one can draw a conclusion about the true author.

In essence, nothing has changed since the origin of stylometry. Methodological atomicity is dominant. Statistical calculations do not have independent evidentiary power, but serve only as an additional way to describe the text. However, due to the ability to cover a large number of texts (which is impossible to collect using traditional analysis), Delta makes it possible to identify certain stylistic trends, solving the problems of classification and periodization of texts. Delta is not enough to attribute some texts, but it can be used to describe the texts and to reveal important tendencies with its computing power. However, it has been known since the time of Lutoslawski: "Style statistics, like all statistics, require great numbers."<sup>15</sup>

Regarding this, the question of the genesis of stylometry as a phenomenon and as a term seems to be especially relevant.

## 2. Stylometry as a phenomenon

Stylometry arises in the second half of 19<sup>th</sup> century against the active use of basic mathematical methods in the humanities and natural sciences. In 1851, the mathematician A. de Morgan suggested that the average length of words in a text may be the feature of the author's style. He proposed testing the technique on the Epistles of the Holy Apostle Paul and the texts of Herodotus. His ideas resonated with the physicist T. Mendenhall, who developed them on the material of Dickens and Thackeray in 1887, and on the dramas of Shakespeare, Bacon and Marlowe in 1901.<sup>16</sup> He concluded that Shakespeare and Marlowe most often used four-letter words. In the 1870s, quantitative research was carried out within the framework of the New Shakespeare Society by Furnival, Ingram, Conrad.<sup>17</sup>

Classical studies formed the basis for the significant development of stylometry. The primary research material for measuring the author's style employing quantitative methods was Plato's dialogues. Philologists (mainly German and Polish) dealt with the

---

<sup>12</sup> Juola 2015.

<sup>13</sup> Calvo Tello 2019.

<sup>14</sup> Orekhov 2021.

<sup>15</sup> Lutoslawski 1897d, 142.

<sup>16</sup> Mendenhall 1887; 1901.

<sup>17</sup> Williams 1970, 3.

issues of dating and periodization. One of the earliest works belonged to Lewis Campbell (1867). His method was to identify rare words that serve as markers of a particular period. Droste (1886) analyzed forms of adjectives ending in *--ειδής* and *-ώδης*.<sup>18</sup> Dittenberger (1881) found different relative frequencies of comparative particles and synonymous tense conjunctions in different dialogues.<sup>19</sup> Research supplied by the material amplification and methodology improvement was continued by Ritter (1888). He counted the specific words used in the text, and he also took into account the possibility of using synonyms in their places. Calculating the predilections of certain authors for the choice of synonyms, Ritter makes conclusions about individual author's style.

This type of quantitative research peaked in the 1880s–1900s. However, despite the chronological density, this process is characterized by atomism and inconsistency: often some researchers do not cite (and, apparently, do not know) the work of others. At the same time, the methods of calculation differ significantly. There was no single school of stylometry neither Shakespearean nor classical. A detailed analysis of dozens of works and methods is given in the Lutoslawski's book, which sets out the main provisions in detail and his own method of analysis.<sup>20</sup> Briefly in French, the same ideas are conveyed in the article “Principes de stylométrie appliqués a la chronologie des œuvres de Platon”.<sup>21</sup> Brandwood summarizes all stylometric studies of Plato's dialogues, including articles written in the twentieth century, in his book.<sup>22</sup>

Grishunin thinks that the only echo of these works in the Russian Empire was N. A. Morozov's article on “linguistic spectra”.<sup>23</sup> This is not entirely true. We should take into account the methodologically similar book dedicated to Orphic hymns, published in Warsaw.<sup>24</sup> The study is rather of a poetic nature, the term “stylometry” is not used, but the quantitative component plays a huge role in the work. There are calculations at the levels of meter, phonetics, morphology and syntax, so we can assert that this comprehensive analysis of the linguistic specifics of Orphic hymns can be called stylometric. It is interesting that Novosadsky refers to calculations and tables of German classical philologists, but there are no bibliographic intersections with Lutoslawski's fundamental book. Of course, this discrepancy can be explained by differences in material. However, as Lutoslawski has repeatedly demonstrated, even in the works devoted solely to Plato, the achievements of other philologists were often ignored.

There is no doubt that Morozov follows Plato's researchers. He is familiar with Lutoslawski's book and his earlier articles written in Polish. Through Lutoslawski, Morozov refers to Gompertz, Dittenberger, Droste and uses the term “stylometry” (*стилеметрия*). Apparently, it is Morozov to whom we owe this translation option and it is from his article that the term migrates to Vinogradov and subsequent authors.

Despite all the methodological imperfections of Morozov's “linguistic spectra”,<sup>25</sup> Morozov's concept turned out to be extremely important in an historical perspective for

---

<sup>18</sup> Cf. Lutoslawski 1897a, 112.

<sup>19</sup> Dittenberger 1881.

<sup>20</sup> See Lutoslawski 1897a.

<sup>21</sup> Lutoslawski 1898. This article is cited, in particular, by B. V. Orekhov, arguing that the term “stylometry” was invented by Lutoslawski: Orekhov 2020, 286.

<sup>22</sup> Brandwood 1990.

<sup>23</sup> Grishunin 1960, 150.

<sup>24</sup> Novosadsky 1900.

<sup>25</sup> They were indicated in: Markov 1916; Sezeman 1918.

several reasons. Firstly, he managed to transfer stylometric studies from the sphere of classical philology and apply them to the Russian material.

Secondly, Morozov narrows down the circle of calculations: he proves that counting the frequency of significant words usage does not make sense, since they turn out to be thematically marked.<sup>26</sup> Consequently, according to Morozov, such words cannot serve as a marker of the author's style. He further argues that it is necessary to focus on the studying of the functional parts of speech. It is important to mention that the idea of analyzing functional parts of speech is very productive for stylometry. The study of Mosteller and Wallace is based on the analysis of function words: using the technique of naive Bayes classifiers on the material of 30 function words, they determined the authorship of 12 anonymous articles from the Federalist Papers (1964). Mosteller and Wallace's study became the most significant in Western stylometry of the 1960s, now viewed as a classical one.

### 3. Towards the origin of the term

In the book *Fundamentals of Stylometry* G. Martynenko states: "The term 'stylometry' was invented by the German philologist W. Dittenberger, who at the end of the last century tried to solve the problem of attribution and dating of Plato's dialogues."<sup>27</sup> This statement begins the substantive part of the book, which has become a reference work for a whole generation of Russian linguists. As evidence, the link to Dittenberger's article is provided: Sprachliche Kriterien für die Chronologie der Platonischen Dialoge. Hermes 16. Berlin. 1880. S. 321–345. However, there are two problems: Dittenberger's article dates from 1881; it does not contain the term "stylometry" or any of its periphrases.

Tuldava also refers to Martynenko, repeating the dating error: "The term 'stylometry' (stylometrics) was also used by the German researcher W. Dittenberger (1880) who made an attempt to solve the task of attribution and chronology of the dialogues of Plato <...> (cf. Martynenko 1988, 5)."<sup>28</sup> Dittenberger's influence on the development of stylometry is undeniable, but, seemingly, it turned out to be especially significant for Soviet researchers.

Also, according to Sommer's dictionary, in German *Stylometrie* was an equivalent to German word *Säulenmesskunst* that was used to denote the art of measuring columns.<sup>29</sup> The same meaning is reflected in the later edition of Kaltschmidt's dictionary.<sup>30</sup> During the 19<sup>th</sup> century, this word was used in the same meaning in English (*stylometry*), French (*stylométrie*) and Italian (*stilometria*).<sup>31</sup> The concept existed before linguistic research, but it had a different meaning. Dittenberger did not even use the Germanized analogue of this concept. Consequently, for our current study it is necessary to consider the re-creation of the term and its adaptation to a new sphere, rather than trying to find out about its creation.

Another possible creator of the term is the Polish philologist, chemist and philosopher W. Lutoslawski<sup>32</sup> (1863–1954), a graduate of the University of Dorpat (Tartu).

---

<sup>26</sup> Morozov 1915.

<sup>27</sup> Martynenko 1988, 3.

<sup>28</sup> Tuldava 2005, 371.

<sup>29</sup> Sommer 1814, 472.

<sup>30</sup> Kaltschmidt 1870, 791.

<sup>31</sup> Diezmann 1836.

<sup>32</sup> A relatively detailed biography is given in a recent work: Ziemacki 2022.

A review of bibliographies of journals devoted to Greek studies from 1891 to 1900,<sup>33</sup> demonstrates that the term “stylometry” appears in the title of the following material for the first time: Lutoslawski, W. On Stylometry. (Abstract of a paper read at the Oxford philological Soc.) // *Class. Review*, 1897, no. 6, p. 284–286.

It summarizes the main points of the report given by Lutoslawski to the Oxford Philological Society on May 21, 1897:

Mr Lutostawski, after a short survey of earlier investigations on Plato’s style, explained his own method of measuring stylistical affinities, which he calls stylometry. Stylometry is a new science, which investigates samples of text as to their style, as paleography investigates the external peculiarities of manuscripts.<sup>34</sup>

Then he describes the differences between stylometry and the German Sprachstatistik and announces the publication of his book in October. Moreover, Lutoslawski states that this report provides the first public explanation of the method of stylometry.

The note also presents the results of Lutoslawski’s calculations:

1. *Gorgias* is later than *Meno*, *Euthydemus*, *Protagoras* and all Socratic dialogues. 2. *Cratylus*, *Symposium*, *Phaedo* form a group later than the *Gorgias* and were written probably in the order here mentioned. 3. *Republic* Bks. II.–X. have been written in a few years, and are later than the *Phaedo*. The composition of this work has not been interrupted by other labours; only Bk. I. is very much earlier, probably written between *Gorgias* and *Cratylus*. 4. *Phaedrus* is written about 379 B. C. and after the *Republic*. The concluding passage, in which educational activity is esteemed above literary activity, is explained by the circumstance that Plato dedicated himself after the *Phaedrus* solely to his oral teaching, and interrupted his literary activity for about twelve years. 5. *Theaetetus* and *Parmenides* follow after a long interval, probably after 368 B. C. 6. *Sophist* and *Politicus* are later than *Parmenides*; *Philebus* is later than the *Sophist*, and perhaps later than the *Politicus*. 7. *Timaeus* and *Critias* are later than the *Sophist*, and probably later than *Politicus* and *Philebus*. 8. The *Laws* are later than the *Sophist*, probably later than *Politicus* and *Philebus* and written contemporaneously with the, *Timaeus* and *Critias*.<sup>35</sup>

Meyer’s article was published almost simultaneously. The article is a translation of some chapters of the book “The Origin and Growth of Plato’s Logic”, with advance notice from Lutoslawski himself, who “with the kind permission of the English publisher introduces German readers to the important results of one of the chapters of this work before its publication”.<sup>36</sup> A short afterword by Lutoslawski is published after the article. The author states: “The above statement of the principles of stylometry is the first publication on this subject (except for a brief report of my Oxford lecture in the *Classical Review*, July 1897, vol. XI, pp. 284–286, and report of the Paris lecture in *Comptes rendus des séances de l’Académie des Inscriptions et Belles-Lettres*, juin 1897, pp. 284–286, 311–314).”<sup>37</sup> The detailed indication of the output data is evidence that short notes in English and French

<sup>33</sup> Ruelle 1893; 1894; 1896; 1897; 1898; 1900; 1901.

<sup>34</sup> Lutoslawski 1897c, 284.

<sup>35</sup> *Ibid.*, 285.

<sup>36</sup> Meyer 1897, 171.

<sup>37</sup> Lutoslawski 1897b, 219.

were published earlier, but the first voluminous article with a detailed presentation of the basic principles of stylometry was published in German.

However, in the issue of the French magazine *La Quinzaine: revue littéraire, artistique et scientifique*, dated July 1, in the “News of Science and Literature” section, a short note *La stylométrie* appears.<sup>38</sup> This three-paragraph note covers the contents of the report made by Lutoslawski at the Academy. Considering the brevity of the note and the speed of its publication compared to other more extensive materials, it is assumed to be the first appearance of stylometry in a new meaning in print.

In the fall of 1897, the book “The Origin and Growth of Plato’s Logic” was published in England. An entire section of the third chapter “Theory of stylometry” is devoted to the principles of the stylometric method.<sup>39</sup>

In a French-language article of 1898, written on the basis of a report at the French Academy of Inscriptions and Beaux-Letters of June 18, 1897, Lutoslawski once again introduces the concept of stylometry, as well as the concept of styleme: a synonym that can be replaced by another word, but is used or even created with the aim of giving a more original turn to the expression of thought. Stylemes are divided into four classes: accidental (*accidentels*; the most frequent, occurring once in the text), repeated (*répétés*; several occurrences), important (*important*; more than one occurrence per twelve pages) and very important (*très important*; most frequent, more than one occurrence per two pages).<sup>40</sup> To calculate stylistic affinity, it is proposed to count a random styleme as one unit, a repeated one as two, an important one as three, and a very important one as four. Thus, it is possible to compare dialogues of unknown date with each other and, based on the similarity of assessment, establish their chronology, which was done by Lutoslawski.

After the publication of the book on Plato at the end of 1897 and the program article of 1898, the number of articles with the term “stylometry” in the title increased.<sup>41</sup>

Tannery, analyzing Plato’s research, states in an article of 1899 that “seventeen years ago Dittenberger proposed to determine the chronology of Plato’s dialogues to follow a method that consists in compiling statistics of certain features of style that do not have philosophical significance, but take into account, for example, the more or less frequent use of equivalent expressions or words”.<sup>42</sup> The methodological, essential base of stylometry was largely established by Dittenberger, but Tannery emphasizes:

In a recently published work, W. Lutoslawski gave this method a new name — stylometry; having carefully analyzed all the previous works on Plato’s style, and shown how unsatisfactory or insufficient they might be, he attempted to formulate the laws to be observed and the rules to be followed in order to obtain in the future sound conclusions and reliable results.<sup>43</sup>

The famous mathematician, editor and researcher of Diophantus, science historian Tannery recognizes Lutoslawski as the creator of this new term.

Thus, we can refute Martynenko’s assertion about Dittenberger’s authorship and we can also correct the data of Orekhov, who indicates that the term first appeared in an ar-

---

<sup>38</sup> La Quinzaine 1897, 471.

<sup>39</sup> Lutoslawski 1897d, 140–193.

<sup>40</sup> Lutoslawski 1898.

<sup>41</sup> See Covotti 1898; Holzner 1899; Tannery 1899; Heikel 1900.

<sup>42</sup> Tannery 1899, 159.

<sup>43</sup> Ibid., 160.

ticle in 1898.<sup>44</sup> The first mention of the term “stylometry” in a linguistic context actually took place on May 21, 1897, as part of Lutoslawski’s report at the Oxford Philological Society. However, those publications where this term appears were published almost simultaneously in the summer of 1897. The precise listing of the imprint of the English and French publications in the afterword to the German-language article indicates that Meyer — Lutoslawski’s article was published later than the summary reports. The publication of the report materials in the *Classical Review* appeared in July (no. 6), similarly in the *Comptes rendus des séances de l’Académie des Inscriptions et Belles-Lettres* (the last report in the 3<sup>rd</sup> issue dates back to June 25). Lutoslawski does not refer to the note in *La Quinzaine* dated July 1, 1897, but there is a possibility that it appeared a little earlier than the materials mentioned: this particular periodical is assumed to be the place of the first appearance of *la stylométrie* in a new meaning. However, in this case, the important issue is linguistic primacy: stylometry is included in philology in English.<sup>45</sup>

Lutoslawski’s introduction of the new term is also emphasized in Brandwood’s work on the chronology of Plato’s dialogues. The structure of the book partly coincides with the historical and analytical section of Lutoslawski’s monograph: various methods and approaches to the periodization of Plato’s texts are examined thoroughly. The approaches of both Dittenberger and Lutoslawski himself are analyzed in detail. Brandwood does not find the term “stylometry” in Dittenberger’s works and concludes that the word is obviously coined by Lutoslawski.<sup>46</sup>

Brandwood sees the reason for the subsequent rejection and some oblivion of Lutoslawski in the arrogance of the Polish philologist, which is observed in the conclusions of his book:

The future science of stylometry [a word apparently coined by him] may improve our methods beyond the limits of imagination, but our chief conclusions can only be confirmed, never contradicted by further research... and now that the method of stylistic calculation has been shown on a small example of five hundred peculiarities, it will be very easy to apply it on a much larger scale, and to settle all the minor difficulties left to future inquirers.<sup>47</sup>

According to Brandwood, this tone encouraged “future inquirers” to abandon Lutoslawski’s method. Moreover, his opponents made no distinction between the stylistic method as a whole and Lutoslawski’s particular version of it, doubtless because to the greater part of the scholarly world the method was introduced by Lutoslawski. Many interpreted his own method as the one employable to the stylistic inquirers in general: “When Lutoslawski fell into disrepute, the stylistic method fell with him. It would hardly be too much to say that it has not yet fully recovered even today from the suspicion which its identification with his ‘stylometry’ incurred.”<sup>48</sup> The “appropriation” of stylometry by Lutoslawski turns out to be not only essential, but also formal and terminological: as shown above, the term existed in the European languages, but, apparently, it was Lutoslawski who first applied it in linguistics.

---

<sup>44</sup> Orekhov 2020, 286.

<sup>45</sup> In lectures and publications of 1895 and 1896 (Lutoslawski 1896; Grzybek 2014), the term stylometry does not yet appear.

<sup>46</sup> Brandwood 1990, 130.

<sup>47</sup> Lutoslawski 1897d, 193.

<sup>48</sup> Brandwood 1990, 130.



The term has been assimilated into the different languages: “stylometry” appears in the Oxford English Dictionary (OED),<sup>49</sup> “stylométrie” in the French dictionary.<sup>50</sup> Interestingly, according to the OED, the term was first mentioned in the book *The Idea of Nature* by the British philosopher Collingwood.<sup>51</sup> However, he rightly traces the origins of the term to Lutoslawski’s 1897 English-language book on Plato.<sup>52</sup> Perhaps, 1945 can be considered a date of “rebirth” of the term stylometry for the English language.

As in the West, in Russia the significance of Lutoslawski was gradually forgotten. The first mention of Lutoslawski’s “new science” in the Russian Empire was in 1898. An attentive reader of the *Philological Review* (Vol. XIV. Book I. 1898. Dept. 2. P. 104) could see such a certificate in the section “Review of Journals. Criticism and Bibliography”: W. Lutoslawski, On Stylometrie, pp. 284–286. Извлечение из прочитанной авторомъ въ Оксфордѣ лекціи о стилометріи примѣнительно къ вопросу о хронологіи Платоновскихъ діалоговъ.<sup>53</sup>

“Stylometric calculation of words” based on Plato’s material and stylometry are critically mentioned without reference to Lutoslawski in “Theoretical Philosophy” (1899).<sup>54</sup> Lutoslawski is then mentioned in the work of Morozov, who already uses the word “стилеметрия”. Morozov’s article marks the beginning of stylometry in Russia. Most linguists and textual critics who use quantitative methods inevitably refer to this work. Lutoslawski is actively quoted by Morozov. Moreover, he often relies on Lutoslawski’s retelling, rather than on the original texts of Campbell or Droste, and provides tables with calculations specifically from the book.<sup>55</sup>

In the review of Sezeman (at that time an active researcher of Plato) on Morozov’s article, Lutoslawski is not mentioned as the inventor of the term and is listed along with other researchers.<sup>56</sup> However, I. Kolubovsky, reviewing the new edition of “Theaetetus” in 1937, writes: “Conclusions from purely formal observations of the style of the ancient Greek author (the so-called stylometry, developed by Lutoslawski in his fundamental work “The Origin and Growth of Plato’s Logic”), as is known, in many cases coincide with philosophical analyzes of dialogues on the merits”.<sup>57</sup>

In the second half of the twentieth century, articles and monographs published by philologists who work with quantitative methods and stylometry treat Lutoslawski in two ways. He is either not mentioned,<sup>58</sup> or is listed among the first stylometricians: Dittenberger, Ritter, Gompertz and Campbell.<sup>59</sup> In most cases, no one mentions that Lutoslawski coined the term.

Apparently, a lacuna can be explained by the fact that soviet researchers had no access to the original English- and French-language texts of Lutoslawski. In American and British stylometry, for obvious reasons, the early stylometric works aimed at studying

---

<sup>49</sup> Oxford English Dictionary 2024.

<sup>50</sup> Dictionnaire de français Larousse 2024.

<sup>51</sup> Collingwood 1945, 58.

<sup>52</sup> The confusion associated with the appearance of the term “stylometry” in English lexicography is described in: Grzybek 2014, 61.

<sup>53</sup> *Filologicheskoe obozrenie* 1898, 104.

<sup>54</sup> Solov’ev 1988, 820.

<sup>55</sup> Morozov 1915.

<sup>56</sup> Sezeman 1918, 71.

<sup>57</sup> Kolubovsky 1937, 140.

<sup>58</sup> See Grishunin 1960; Martynenko 1988; Holmes 1998; Gurova 2016.

<sup>59</sup> See Vinogradov 1959; Herdan 1966; Williams 1970; Sineleva 2000.

Shakespeare's text are more relevant. This may explain the insufficient attention paid to Lutoslawski.

Finally, another hypothesis about the oblivion of Lutoslawski in Russia is related to ideology. Lutoslawski was a consistent supporter of Polish messianism, which is reflected in his texts.<sup>60</sup> This may also explain the arrogant tone, which, according to Brandwood, scared away many philologists not only from Lutoslawski, but also from stylometry as a young science. It cannot be ruled out that even with the presence of his texts in one or another archive, Lutoslawski's ideological and philosophical views became an obstacle.

#### 4. Conclusion

Most likely, the term "stylometry" was introduced into the philology by W. Lutoslawski in 1897. Philologists who studied the Platonic texts incline towards similar conclusions. However, it is worth considering that throughout the 19<sup>th</sup> century, the word was used in the sense of 'the art of measuring columns' and it was not an unambiguous novelty.

Apparently, the first mention of this term in a philological context and in a new meaning is Lutoslawski's report at the Oxford Philological Society on May 21, 1897, followed by a speech at the French Academy of Inscriptions and Fine Arts on June 16, 1897. The next one is a short note in the magazine *La Quinzaine* based on the June report, published July 1, 1897, in the Science News section; a more detailed note based on the report "On Stylometry" appeared in the *Classical Review* in July 1897; a French-language report "La loi de stylométrie" was also published in July, and an extensive German-language article by W. Lutoslawski's *Theorie der Stylometrie auf die Platonische frage angewendet* was printed in the summer of 1897.

In Russia, stylometry (*стилометрия*) was first mentioned in the *Philological Review* (*Filologicheskoe obozrenie*) in 1898, and in the form *стилеметрия* by Morozov in 1915. This version became widespread in Soviet philology.

Stylometry, as a term and as a direction, is a bridge between classical studies and mathematical linguistics. Created by classical philologists for the sake of new opportunities for studying and describing texts and solving problems of attribution, stylometry has become one of the most widespread, developed and authoritative disciplines that combine qualitative and quantitative analysis. Nowadays in classical studies researchers still use it but it is obvious that stylometric methods do not predominate.

Current stylometric studies are characterized by methodological atomicity. This fragmentation and multi-methodology is revealed at the very early stage of the discipline's existence. The most promising direction today seems to be the development of stylometry towards comparative stylistic research. The tools of descriptive statistics and the computing power of modern digital methods make it possible to compare the stylistic profiles of an unprecedented number of authors. This would be simply impossible to study using traditional linguistic-stylistic methods.

The covenant of "complexity", the need to combine qualitative and quantitative methods, remains immutable. We can only quote Vinogradov: "the study of style should be comprehensive and systematic"<sup>61</sup> and also recall the prudent words

<sup>60</sup> Lutoslawski 1922; Lutoslawski 2015.

<sup>61</sup> Vinogradov 1961, 198.

of Holmes: “in the context of authorship attribution, stylistic evidence must be weighed in the balance along with that provided by more conventional studies made by literary scholars”.<sup>62</sup>

## References

- Alieva O. V. Testing Burrows' Delta on Ancient Greek Authors. *ΣΧΟΛΗ* 2022, 2, 693–705 (in Russian).
- Brandwood L. *The Chronology of Plato's dialogues*. Cambridge, Cambridge University Press, 1990.
- Burrows J. 'Delta': a Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing* 2002, 3, 267–287.
- Calvo Tello J. Delta inside Valle-Inclán: Stylometric Classification of Periods and Groups of his Novels. *Romanische Studien* 2019, 6, 151–163.
- Collingwood R. *The Idea of Nature*. Oxford, Oxford University Press, 1945.
- Covotti A. Teoria della stilometria applicata ai dialoghi platonici da W. Lutoslawski. *Rendiconti dell' Accad. d. Lincei* 1898, V, 9–10, 396–413; VI, 12, 534–550.
- Diezmann J. A. (ed.). *Nouveau dictionnaire de poche des quatre langues principales de l'Europe*. Vol. 2. Leipzig, Baumgärtner's Buchhandlung, 1836.
- Dittenberger W. Sprachliche Kriterien Für Die Chronologie Der Platonischen Dialoge. *Hermes* 1881, 16, 321–45.
- Giantos T. E., Bolt T. J., Chaudhuri P., Dexter J. P. Stylometric Classification of Ancient Greek Literary Texts by Genre. In: B. Alex, S. Degaetano-Ortlieb, A. Kazantseva, N. Reiter, S. Szpakowicz (eds). *Proceedings of the 3<sup>rd</sup> Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature*. Minneapolis, 2019, 52–60.
- Grishunin A. L. Studying the Use of Language Doublets. *Voprosy tekstologii* 1960, 2, 146–196 (in Russian).
- Grzybek P. The Emergence of Stylometry: Prolegomena to the History of Term and Concept. In: K. Kroó, P. Torop (eds). *Text within Text — Culture within Culture*. Budapest — Tartu, L'Harmattan, 2014, 58–75.
- Gurova E. I. Methods of Attribution of Authorship in Modern Russian Philology. *Novyi filologicheskii vestnik* 2016, 3, 29–43 (in Russian).
- Heikel I. A. Bemerkungen zur Sprachstatistik und zur sogenannten Stylometrie. *Eranos* 1900, 4, 11–19.
- Herdan G. *The Advanced Theory of Language as Choice and Chance*. Berlin, Springer, 1966.
- Holmes D. I. The Evolution of Stylometry in Humanities Scholarship. *Literary and Linguistic Computing* 1998, 3, 111–117.
- Holzner E. Stilometrie (zu Platon). *Beil. zu Münchner allg. Zeitung* 1899, 44.
- Juola P. The Rowling Case: a Proposed Standard Protocol for Authorship Attribution. *Digital Scholarship in the Humanities* 2015, 30, 100–113.
- Kaltschmidt J. H. (ed.). *Neues und vollständiges Fremdwörterbuch zur Erklärung alles aus fremden Sprachen entlehnten Wörter und Ausdrücke*. Frankfurt, Salzwasser-Verlag, 2020 [1870].
- Kolubovskiy I. Ia. Review of the book: Plato. Theaetetus. Transl. by V. C. Serezhnikov. Moscow — Leningrad, 1936. *Front nauki i tekhniki* 1937, 2, 139–144 (in Russian).
- Lutoslawski W. La loi stylométrique. *Comptes rendus des séances de l'Académie des Inscriptions et Belles-Lettres* 1897a, 41, 3, 311–314.
- Lutoslawski W. Mesjanizm jako polski światopogląd narodowy. In: A. Wawrzynowicz. *Spór o mesjanizm*. Warsaw, Fundacja Augusta hr. Cieszkowskiego, 2015, 381–390.
- Lutoslawski W. Nachtrag zu der vorhergehenden Abhandlung. *Zeitschrift für Philosophie und philosophische Kritik* 1897b, 110, 217–219.
- Lutoslawski W. On Stylometry (Abstract of a paper read at the Oxford philological Soc.). *Classical Review* 1897c, 6, 284–286.
- Lutoslawski W. *Praca narodowa. Program polityki polskiej*. Vilnius, Księgarni Stowarzyszenia Nauczycielstwa Polskiego, 1922.
- Lutoslawski W. Principes de stylométrie appliqués a la chronologie des œuvres de Platon. *Revue Des Études Grecques* 1898, 11, 41, 61–81.

---

<sup>62</sup> Holmes 1998, 116.

- Lutoslawski W. *Sur une nouvelle méthode pour déterminer la chronologie des dialogues de Platon. Mémoire lu le 16 mai 1896, à l'institut de France devant l'Académie des sciences morales et politiques*. Paris, H. Welter, 1896.
- Lutoslawski W. *The Origin and Growth of Plato's logic*. London, Longmans, 1897d.
- Markov A. A. On One Application of the Statistical Method. *Izvestiia Imperatorskoi Akademii nauk* 1916, 10, 4, 239–242 (in Russian).
- Martins A., Grácio C., Teixeira C., Rodrigues Pimenta I., García Zapata J.L., Ferreira L. Historia Augusta Authorship: an Approach Based on Measurements of Complex Networks. *Applied Network Science* 2021, 6, 50. <https://rdcu.be/dTWNV> (accessed: 01.05.2024).
- Martynenko G. Ia. *Fundamentals of Stylometry*. Leningrad, LSU Publ., 1988 (in Russian).
- Martynenko G. Ia. Stylometry: Emergence and Formation in the Context of Interdisciplinary Interaction. In: I. S. Nikolaev (ed.). *Strukturnaia i prikladnaia lingvistika* 2014, 3–23 (in Russian).
- Martynenko G. Ia., Grebennikov A. O. *Fundamentals of Stylometry: Educational and Methodological Manual*. St Petersburg, Saint Petersburg University Press, 2018 (in Russian).
- Mendenhall T. C. A Mechanical Solution of a Literary Problem. *Popular Science Monthly* 1901, 9, 97–105.
- Mendenhall T. C. The Characteristic Curves of Composition. *Science* 1887, 9, 214 (Supplement), 237–249.
- Meyer P. W. Lutoslawski's Théorie der Stylometrie auf die platonische Frage angewendet. Auszug aus dem Werke. *Zeitschrift für Philosophie und philosophische Kritik* 1897, 110, 171–217.
- Morozov N. A. Linguistic Spectra as a Means for Distinguishing Plagiarism from the True Works of One or Another Famous Author and for Determining their Era. *Izvestiia otdela russkogo iazyka i slovesnosti Imperatorskoi Akademii nauk*, 1915, XX, 4, 93–134 (in Russian).
- Nikolaev D., Shumilin M. Some Considerations on the Attribution of the 'New Apuleius'. *Classical Quarterly* 2021, 71, 819–848.
- Novosadskii N. I. *Orphic hymns*. Warsaw, Tipografia Varshavskogo uchebnago okruga Publ., 1900 (in Russian).
- Orehov B. V. "The Iliad" by E. I. Kostrov and "The Iliad" by A. I. Lyubzhin: Stylometric Aspect. *Aristei* 2020, XXI, 282–296 (in Russian).
- Orehov B. V. Text and Translation by Vladimir Nabokov through the Prism of Stylometry. *Novyi filologicheskii vestnik* 2021, 3 (58), 200–213 (in Russian).
- Ruelle Ch.-Em. Bibliographie annuelle des Études Grecques (1891–92). *Revue des Études Grecques* 1893, 6–24, 410–478.
- Ruelle Ch.-Em. Bibliographie annuelle des Études Grecques (1892–93–94). *Revue des Études Grecques* 1894, 7–27/28, 402–456.
- Ruelle Ch.-Em. Bibliographie annuelle des Études Grecques (1894–94–95). *Revue des Études Grecques* 1896, 9–33, 110–167.
- Ruelle Ch.-Em. Bibliographie annuelle des Études Grecques (1894–95). *Revue des Études Grecques* 1897, 10–38, 193–20.
- Ruelle Ch.-Em. Bibliographie annuelle des Études Grecques (1895–96–97). *Revue des Études Grecques* 1898, 11–44, 439–517.
- Ruelle Ch.-Em. Bibliographie annuelle des Études Grecques (1897–98–99). *Revue des Études Grecques* 1900, 13–51, 45–121.
- Ruelle Ch.-Em. Bibliographie annuelle des Études Grecques (1898–99–1900). *Revue des Études Grecques* 1901, 14–57, 206–263.
- Rybicki J., Eder M. Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words? *Literary and Linguistic Computing* 2011, 26, 315–321.
- Sezeman V. "Linguistic spectra" of Mr Morozov and Plato's Question. *IORIaS* 1918, XXII, 2, 1918, 70–80 (in Russian).
- Shumilin M. V. Identifying Latin Authors through Maximum-Likelihood Dirichlet Inference: A Contribution to Model-Based Stylometry. *Shagi/Steps* 2021, 7 (1), 183–198.
- Sineleva A. V. Attribution of "A Romance with Cocaine": a linguistic-statistical study. Nizhny Novgorod, Nizhegorodsky gosuniversitet im. N. I. Lobachevskogo Publ., 2000 (in Russian).
- Solov'ev V. S. *Collected Works in 2 vols*. Vol. 1. Moscow, Mysl' Publ., 1988, 757–831 (in Russian).

- Sommer J. G. (ed.). *Neuestes wort- und sacherklärendes Verteutschungs-Wörterbuch aller jener aus fremden Sprachen entlehnten Wörter, Ausdrücke und Redensarten, welche die Teutschen bis jetzt, in Schriften und Büchern sowohl als in der Umgangssprache, noch immer für unentbehrlich und unersetzlich gehalten haben. Ein Handbuch für Geschäftsmänner, Zeitungsleser und alle gebildete Menschen überhaupt*. Prag, J. G. Calve'sche Buchhandlung, 1814.
- Storey G., Mimno. D. Like Two Pis in a Pod: Author Similarity Across Time in the Ancient Greek Corpus. *Journal of Cultural Analytics* 2020, 5 (2).
- Stover J. A. *A New Work by Apuleius: The Lost Third Book of the De Platone*. Oxford, Oxford University Press, 2016.
- Stover J. A., Koppel Y., Winter M., Kestemont M. Computational Authorship Verification Method Attributes a New Work to a Major 2<sup>nd</sup>-Century African Author. *Journal of the Association for Information Science and Technology* 2016, 67, 239–242.
- Stover J., Kestemont M. Reassessing the Apuleian Corpus: a Computational Approach to Authenticity. *Classical Quarterly* 2017, 66, 645–672.
- Tannery P. La stylométrie ses origines et son présent. *Revue Philosophique de La France et de l'Étranger* 1899, 47, 159–169.
- Tuldava J. Stylistics, Author Identification. In: R. Köhler, G. Altmann, R. Piotrowski (eds). *Handbooks of Linguistics and Communication Science*. Berlin — New York, De Gruyter, 2005, 368–387.
- Vinogradov V. V. *On the Language of Fiction*. Moscow, Uchpedgiz Publ., 1959 (in Russian).
- Vinogradov V. V. *The Problem of Authorship and the Theory of Styles*. Moscow, Goslitzdat, 1961 (in Russian).
- Williams C. B. *Style and Vocabulary: Numerical Studies*. London, Griffin, 1970.
- Zhuravleva N. N. Application of Quantitative Methods in Analyzing the Author's Style and Solving Problems of Attribution. *Vestnik Tiimenskogo gosudarstvennogo universiteta* 2012, 1, 150–155 (in Russian).
- Ziemacki J. *Wincenty Lutoslawski — droga myśli*. PhD thesis. Warsaw, 2022.

## От классической филологии к цифровой: о возникновении и развитии стилеметрии\*

Борис Вадимович Ковалев

Санкт-Петербургский государственный университет,  
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9;  
<https://orcid.org/0000-0002-1904-1844>, [b.v.kovalev@spbu.ru](mailto:b.v.kovalev@spbu.ru)

**Для цитирования:** Kovalev B. V. From Classics to Digital Philology: On the Origin and Growth of Stylometry. *Philologia Classica* 2024, 19 (2), 347–360. <https://doi.org/10.21638/spbu20.2024.211>

Статья посвящена возникновению стилеметрии и ее развитию на раннем этапе. Стилеметрия — прикладная филологическая дисциплина, рассматривающая стиль как набор определенных количественных параметров и занимающаяся измерением стилистических характеристик с целью систематизации текстов и их частей. Подчеркивается, что эта дисциплина возникла именно на материале классической филологии. На примере наиболее надежного и авторитетного на сегодняшний день стилеметрического метода Delta Берроуза, основанного на средней абсолютной разности между z-оценками некоторого количества наиболее частотных слов в контрольном и атрибутируемом текстах, анализируются преимущества и недостатки такого рода анализа. Устанавливается генезис термина «стилеметрия». В основополагающей книге Г. Я. Мартыненко «Основы стилеметрии» указывается, что термин придумал немецкий филолог-классик В. Диттенбергер. В ходе исследования выясняется, что термин «стилеметрия» существовал

\* Работа выполнена при поддержке СПбГУ, проект 95434615 «Литературные тексты и их язык vs количественные, корпусные и компьютерные методы: взаимное тестирование (Набоков и сопоставительный материал): 2024 г., этап 5».

на протяжении всего XIX в. в значении 'искусство измерения колонн', а в работах Диттенбергера он не используется. В филологию этот термин принес польский философ, химик и филолог В. Лютославский, пытавшийся решить задачу периодизации диалогов Платона. Выясняется, что впервые «стилеметрия» употребляется в новом значении 21 мая 1897 г. во время доклада В. Лютославского в Оксфордском филологическом обществе. На русском языке термин впервые встречается в обзоре 1898 г. в форме «стилометрия», а в статье Н. Морозова 1915 г. — в форме «стилеметрия», которая получила распространение в советском академическом сообществе в силу влиятельности этой статьи на позднейшие стилеметрические и текстологические исследования.

*Ключевые слова:* стилеметрия, Лютославский, Диттенбергер, лингвопоэтика, стилистика, история языкознания, текстология, Платон.

Received: 20.03.2024

Accepted: 26.09.2024