

PETER GRZYBEK
(University of Graz — Graz, Austria)

THE EMERGENCE OF STYLOMETRY: PROLEGOMENA TO THE HISTORY OF TERM AND CONCEPT

INTRODUCTION

This contribution intends to shed some light on the beginnings and the development of ‘stylometry’, which came into being primarily in the second half of the 19th century.¹ The emergence of this discipline is of course no isolated cultural phenomenon; rather, it is related to more general cultural contexts of 19th century society and culture. On the one hand, this cultural embedding concerns, on the one hand, the increased importance of and reflection on individuality in general, a specific style being understood as an indicator of individual expression. On the other hand, the specific application of empirical and mathematical methods, as a complementary approach to genuine philological problems deserves special attention in the context of 19th century academia, particularly because at that time we are concerned with a period, when cultural and natural sciences increasingly became to be seen separately from and in juxtaposition to each other — be that with regard to methods or objects —, empirical and quantitative approaches thus combining both alleged opposites. As a result, the reconstruction of the history of stylometry is likely to offer far-reaching conclusions, further than might be expected at first sight.

Although the emergence of stylometry, both as a concept and as a term, has repeatedly been the object of historiographical reconstructions, its history has remained scattered and is yet to be written. First and foremost, this is due to the

¹ The increased emphasis on individuality has been sufficiently discussed, cf., e.g., for the British context: Camlot 2008. Therefore, more or less arbitrarily chosen quotations may suffice here, e.g.: “[T]he essence of style is individuality” – Earle 1890: 293; „Style, in fact, is the vehicle of character” – Mallock 1892: 447); “[A] distinctive style is ... almost as inevitable as a distinctive handwriting” – “Literary Style”, *Spectator* (Oct. 1., 1892: 445); “All pure literature is the revelation of the man” – Burroughs 1899: 400.

fact that those few and sporadic works which have touched upon and dealt with the historical dimension of stylometry, have approached the issue from rather limited perspectives, focusing on selected phenomena only. It is no wonder, therefore, that the beginnings and development of stylometry have been seen at different times, and that they have been linked with different persons who are credited for having given birth to it.

Still in 1986, to give but one example, the supplement to the *Oxford English Dictionary (OED)* (1986) referred to British philosopher and historian Robin G. Collingwood (1889–1943) as allegedly being responsible for the first occurrence of the term ‘stylometry’ in his 1945 book *The Idea of Nature*; likewise, the associated adjective ‘stylometric’ was assumed to have appeared about a decade before. But as a quick glance at Collingwood’s book and its chapter about Plato’s theory of forms shows, Collingwood explicitly related the emergence of the term ‘stylometry’ to Polish scholar Wintency Lutosławski’s and his monograph on *The Origin and Growth of Plato’s Logic*,² classifying it “as a continuation and elaboration of researches set on foot by Lewis Campbell in 1867.”³ As a consequence, Binongo and Smith, in their 1996 article on ‘stylometry’,⁴ corrected the view expressed in the *OED* and suggested two additions, a decade later, pointing out that Wellek and Warren, in their *Theory of Literature*,⁵ had already referred to Scottish philologist Lewis Campbell and Lutosławski, both of whom had made use of statistical methods in establishing the chronological order of Plato’s dialogues. Whereas Wellek and Warren, however, explicitly referred to Lutosławski as being responsible for having coined the term ‘stylometry’, in 1897, Binongo and Smith weaken this argument, assuming “that even in 1897, neither the term nor the discipline, stylometry, was new.”⁶ In their opinion, Lutosławski himself, albeit not ignoring earlier quantitative approaches to stylistic analyses, did not claim to have fathered stylometry, referring to Campbell as one of his forerunners.

As these few examples show, the history of stylometry has indeed remained rather obscure till today. One of the reasons for the overall uncertain state of affairs is the fact that the concept of stylometry had of course some forerunners before the term ‘stylometry’ itself became established, and that it has been, in the course of its existence and development, differently defined, or interpreted, and submitted to various kinds of practical application. Another reason has to be seen

² Lutosławski 1897a.

³ Collingwood 1945: 58.

⁴ Binongo–Smith 1996.

⁵ Wellek–Warren 1956: 65.

⁶ Binongo–Smith 1996: 448.

in the circumstance that the emergence and development of stylometry has been everything else but uni-linear; rather, as will be shown in the discussion below, it has been a combination and overlapping of different lines, converging only partly, at particular points of time or periods. Finally, many relevant works from the second half of the 19th century have simply been forgotten and are unreasonably ignored nowadays.

It goes without saying that a full account of the history of stylometry would go beyond the scope of this contribution. Rather, subsequent to the initial discussion of some basic concepts, including a number of methodological remarks about style and stylometry, the overall objective of this contribution will be to outline the major historical lines of influence which, in the beginning, seem to have evolved independently from each other, and first went along unrelated parallels, before they increasingly converged into one common field of research, evolving into the multi-faceted discipline it is today.

STYLE AND STYLOMETRY: SOME DEFINITION(S)

In the most general understanding of this term, stylometry, as we understand it today, is that discipline which, for the stylistic study of texts, integrates and relies upon statistical procedures to achieve its goal. This goal can be seen in the detection of stylistic specifics, characteristic of the given text(s) under study; from a different perspective, and in terms of a (complementary) alternative, such stylistic text specifics may then be projected from the textual level to the author(s) of the text(s) under study and interpreted in terms of some author-specific style. Which texts are chosen to achieve the goal, which features are considered to be relevant and submitted to statistical analysis, and which statistical procedures are applied, depends on many factors, starting from the concrete research questions to the advancement of the given discipline at a given point of time.

In this sense, the concrete (pragmatically or theoretically motivated) research problems, submitted to stylometric analyses, can be seen as derived or secondary goals. Traditionally, it has become common to reduce the field of stylometry with its broad spectrum of options outlined above to two major fields of interest. These can be seen as pragmatic applications, i.e., related to some concrete interest with a background of a more theoretical framework, namely, 'attributional' and 'chronological' aspects. In the first case, we are basically concerned with problems of authorship — i.e. authorship detection, identification, or attribution (what may in turn be related with secondary objectives as, e.g. in context of forensic linguistics,

literary history, juridical issues, etc.) —, in the second case, with the establishment of a chronological order of texts (by a given author).

It is self-evident that any reduction of the vast field of stylometry, as outlined above, to some more narrowly defined research area will result in a different account of the discipline and its history. But even if common agreement can be achieved as to the general definition of stylometry, various problems and issues may come into play, including different understandings of basic terms and concepts — be that the notion of ‘style’ itself, or the necessarily immanent notions of ‘text’ or of ‘statistics’, mentioned above. In fact a number of distinctions may be made in this context, all of which can, in principle, be sub-summarized under the general umbrella term of stylometry. However, depending on the understanding chosen, deliberately or not, the resulting understanding of stylometry will differ. This is not the place to discuss at length all possible implications, or interpretations, of each of these concepts, which have been more or less important in the history of approaching related questions. Yet, it seems worthwhile, for the purpose of orientation, to at least mention some cornerstones, in order to define and stake out the vast field of stylometry.

First, it is the notion of ‘style’ itself, which can and, in fact, has been differently defined in the history of this term’s existence. Even when concentrating on questions of definition, rather than evaluation — i.e., focusing on the question who or what *is* style, rather than who or what *has* (good or bad) style — a number of different approaches to style can be distinguished: in addition to what one might term some kind of “all-or-nothing positions” — ranging from the assumption that style does not exist at all or, if so, it cannot be defined, to an understanding of style as an integral component of any semiotic, or structuring, activity — style has mainly been defined along one of the following major three lines: (a) as choice, (b) as norm, or (c) as deviation from some norm. Whatever decision or perspective is taken in this respect, style has traditionally been interpreted to be something, or to denote something “individual”. In this respect, one should be careful, however, if individuality refers to an individual text or to an individual author. Moreover, account should be taken of the fact that style may of course not only be understood to characterize some specific individual phenomenon. By way of an alternative, one may also assume style to characterize a whole group, or a set of elements, distributed in space and/or time; this is the case, for example, when all texts by a given author, or all texts by a given author, belonging to one and the same genre, or all texts written at a particular time, etc., are assumed to be characterized by some kind of “individual” style. In this case, more than one element is grouped into a set by the underlying assumption of stylistic homogeneity, e.g., between texts

or authors. In this sense, one may also speak of inter-individual stylistics, juxtaposing it to individual stylistics, as has been the case, for example with functional styles, characterizing particular discourse types of communication.

But not only may the definition (or implicit understanding) of 'style' vary, also different understandings of the term 'text' may come into play. Again, ignoring difficulties in precisely defining what distinguishes a 'text' from a 'non-text', a 'text' may either be understood as some kind of verbal utterance, only, or it may be understood in a broader (semiotic) understanding of the term; in this case, musical 'texts', film 'texts', etc., would be included into a more general understanding of text and, as a consequence, might be submitted to stylometric analyses. Such inter-medial stylistics across texts from different semiotic systems may enrich insight into the semiotic spectrum.

Finally, the usage and application of statistical procedures may differ in stylometric analyses: restricting them to descriptive procedures — what has been particularly the case in the beginnings of stylometry, when test statistics in our understanding were not yet, or not yet fully, available — any (comparative) conclusion as to stylistic specifics cannot go beyond intuition, subjective graphical impression, etc. But only the application of inferential statistics, necessarily requiring model building, allows for hypothesis formulation and testing. In this framework, stylometry can be seen to be based on a probabilistic understanding of style as (the result of) a specific selection process: the repeated selection from a set of alternatives $A=\{a,b,\dots,n\}$ results in the probability distribution $P\{A\} = P(a), P(b), \dots, P(n)$, to each of the alternatives being attributed its frequency of occurrence in the given text. Each text T can thus be represented by a set of measurable statistical characteristics, $T=\{C_1,C_2,\dots,C_n\}$ with C_i as the defined text characteristics. In this framework, a text can be analyzed in various aspects and from various perspectives — for example as a process (e.g., by way of sequential analysis, focusing the linear development of the characteristics C_i), or as a product (e.g., with regard to a frequency model resulting from the repeated occurrence of the various C_i) —, and can thus be described as a multidimensional space with the values C_i for its individual vectors.

Without a doubt, the above-mentioned list of approaches to style is far from being exhaustive: not only may some of the points mentioned be combined with each other, but the list as such may also be enlarged and extended. Anyway, it should become evident from the foregoing remarks that depending on any decision as to the definition of style, text, and statistics, an account of the emergence and history of stylometry, as a concept and as a discipline, will take different shapes, and provide a more or less encompassing picture: a different number of works with

different orientations will be included or excluded, and they will be differently evaluated with regard to their relevance, be that from a contemporary perspective, or in the course of the further development of the discipline, or from today's point of view.

THE HISTORY OF STYLOMETRY: TERM VS. DISCIPLINE

If the idea were to write a full and comprehensive history of stylometry, as a discipline, it would likely be necessary to take into account and to cover all aspects of the broad spectrum of this approach mentioned above. Furthermore, “stylometric research” (in whatever sense) may well have arisen before the term came up, but also the term, once having been invented, may well have denoted more or less of what later became agreed upon as stylometry. In any case, one should be cautious in clearly distinguishing *term* from *concept*: it is one question to reconstruct when a particular term came into being (and what was related to it), but another question, as to how a particular stylometric concept (or discipline), emerged (and how it was defined). Unfortunately, these two aspects have not always been kept apart clearly enough in those works which have attempted to shed some light on the historical dimension of this issue.

As a matter of fact, a broader understanding of ‘stylometry’ will result in a longer history. In other words: depending on the option(s) chosen, it might be quite difficult to distinguish any kind of statistical analysis of texts from stylometric analyses in a closer understanding of this word, and one might end up many centuries ago. This is the case, for example, when Williams, in the introduction to his *Style and Vocabulary*, refers to the Masoretes (ca. 600–1000 CE), who counted the number of letters and the number of words in each Book of the Hebrew Old Testament, as well as the number of repetitions of certain words.⁷ And this is also the case, when Italian humanist Lorenzo Valla, with his 1439 proof that the *Donation of Constantine* [*Constitutum Constantini*, or *Donatio Constantini*] was a forgery, is considered to be an important source of stylometry, since it includes techniques of analyzing texts for evidence of authenticity, authorial identity, and other questions, based on analyses of occurrence or otherwise of specific words.

Despite such attempts to localize the beginnings of stylometry (in a broader understanding) “as early as possible”, usually the 19th century is agreed upon as that period when this discipline had its crucial offshoots. However, those (few)

⁷ Williams 1970: 2.

historiographic accounts, which are based on this general agreement, usually tend to focus on one particular work by a chosen scholar as the starting point for the reconstruction of some research “line”. The existence of one such line, however, is but an illusion, if one takes a closer look at stylometry’s emergence and development. In fact, the development of stylometry has been anything but uni-linear: there has been more than one line, the different lines either running parallel or overlapping, at different stages of their development. The complexity of these interweaving lines has been hitherto largely ignored, however; as a result, we have but a piecemeal understanding of the history of stylometry. Moreover, even attempts focusing on the reconstruction of one “line” have only provided results which tend to be full of gaps, leapfrogs, and inconsistencies — too many works from this field of academic historiography have been forgotten and are ignored until today.

It goes without saying that no complete historiography of stylometry can be provided in this article. Yet, the aim is to demonstrate, on the basis of selected examples, how multi-faceted the emergence and evolution of this discipline has been, and how thus far neglected missing links may contribute to a broader, and more reliable, even of mosaic picture of the history of stylometry as a whole.

ORIGINS, DEVELOPMENTS AND FILIATIONS IN 19TH-CENTURY STYLOMETRY

As to origins of stylometry in the 19th century, one might side with scholars like Holmes,⁸ who place Augustus de Morgan in first place in stylometry’s emergence. In fact, in an 1851 private letter to his friend, Reverend W. Heald, the English logician suggested that questions of authorship might be settled by determining if one text “does not deal in longer words” than another. This suggestion became public when de Morgan’s letter was published in 1882, and Mendenhall himself pointed out that it was just this quotation which served as a starting point for his own empirical studies in this direction.⁹ As a mathematician, Mendenhall was well familiar with contemporary spectral analysis in physics, and in analogy to this he proposed to go beyond mere averages of word length, as suggested by de Morgan, and to analyze a composition by forming what he suggested might be called “a ‘word-spectrum’, or ‘characteristic curve’, which shall be a graphic representation

⁸ Holmes 1998.

⁹ Mendenhall 1887: 237.

of an arrangement of words according to their length and to the relative frequency of their occurrence".¹⁰

Mendenhall was convinced that "personal peculiarities ... will, *in the long-run*, recur with such regularity that short words, long words, and words of medium length, will occur with definite relative frequencies".¹¹ In this context, Mendenhall suggested that this method might be applied to cases of disputed authorship, a perspective which he took up some 15 years later,¹² admitting that this had been in view a long time before. But it is clear that the issue of disputed authorship was not the impetus for his research; rather, as Mendenhall¹³ put it himself, for him this turned out to be rather an "application of the doctrine of chance".

In the field of stylometry, Mendenhall's works have similarly been appreciated in the field of quantitative linguistics; yet, they remain to be interpreted in terms of some isolated phenomenon. Practically the same holds true for Lucius A. Sherman: the Canadian scholar from the University of Nebraska conducted and initiated a number of important sentence-length studies, which were particularly honored almost a century later by quantitative linguist Gabriel Altmann, who baptized this as the Sherman Law of sentence length frequency distribution.

Although both authors, Mendenhall and Sherman, have been treated and appreciated independently of each other, it may well be the case — although we have no direct evidence in favor of this assumption — that an early reaction to Mendenhall's 1887 study, which was published in the journal *Science* still in March 1887, and which was authored by mathematician Henry T. Eddy, prompted Sherman¹⁴ to publish his results on sentence length in English; according to his own words, his article based on research, started much earlier.¹⁵ In his overall positive short reaction — there were some more reactions to Mendenhall's article in 1889, e.g., those signed by "A.B.M.", "H.A. Parker", and "M". —, Eddy had pointed out that personal style, as depicted in Mendenhall's "characteristic curves" is likely to be "principally controlled by the language in which the composition is written", and that word length (alone) might not be adequate to characterize an author's peculiarity.¹⁶ Assuming that "there are other characteristics of writers ..., in which their personal peculiarities differ more widely, and which are therefore

¹⁰ Ibid. 238.

¹¹ Ibid. 239.

¹² Mendenhall 1901: 101.

¹³ Mendenhall 1904: 373

¹⁴ Sherman 1888.

¹⁵ Cf. Sherman 1892: 337.

¹⁶ Eddy 1887.

more characteristic than the habitual selection and use of long or short words”, Eddy suggested “that the length of the sentences employed by a writer is such a peculiarity”, and that “the simultaneous application of several of these enumerations would, in any case of disputed authorship, afford decisive tests such as could not be obtained from any one of them singly”.¹⁷

Whereas we thus do not reliably know what initially prompted Sherman’s interest in sentence length, it seems likely that Eddy’s short reaction was the impetus for his publication and thus represents some kind of missing link between Mendenhall’s influential work on word length and the following work on sentence length. It should not go unmentioned, in this context, that Sherman’s own work served as a starting point for further research in this direction, be that by students of his, like George W. Gerwig or Carson Hildreth¹⁸ (1897), or by his Nebraska colleague, mathematician Robert E. Moritz, whose valuable works still remain to be (re)discovered and integrated into the history of stylometry and quantitative linguistics.

In any case, Mendenhall and Sherman played an important role in the history of quantitative stylistics; taking additionally into consideration all those works which arose in the direct and indirect context, their role turns out to be even more important than has hitherto been assumed. Still, crediting Mendenhall and Sherman, or de Morgan, for having initiated stylometric studies at large, i.e., in the broader understanding of the word, would do injustice to yet another branch of early stylometry, namely to the quantitative stylistics line in the English tradition of Shakespeare research.¹⁹

Works from this tradition are predominantly related to the well-known *New Skakspere Society*, which was founded in 1873, with Frederick James Furnivall (one of the three editors of the OED) as its initiator. The major objective of this society was the promotion of quantitative techniques in the study of disputed authorship

¹⁷ Ibid. 297.

¹⁸ Gerwig and Hildreth published their papers in the *Nebraska University Studies*, as had Sherman before. Other students of Sherman’s who participated in these studies, were Hugh C. Laughlin and Louise Pound. Two papers by Laughlin on “The co-ordinate stage in language development” and “On the principle of predicate suppression” have never been published. Louise Pound’s paper “The Romaunt of the Rose: additional evidence that it is Chaucer’s”, representing what she terms a “sentence-length test” to study authorship of spurious Chaucer texts, was published in the renowned *Modern Language Notes* in 1896. After having finished her dissertation in Heidelberg 1901, Pound was to become one of the most prominent woman academicians in linguistics and folkloristics in the United States in the 1920s.

¹⁹ Cf. Grieve 2005: 4ff.

and chronology of the Shakespearian canon. In this respect, it is particularly the society's *Transactions* (1874ff.) which would contain numerous quantitative studies in this direction, authored by researchers such as Frederick Gard Fleay, John Kells Ingram, and others.

These works are important and indispensable elements of a history of stylometry, and played a crucial role in this discipline's development. But again, these works, were no isolated phenomena but had their forerunners, too, and with interrelations to other developmental lines, at a later point of history.

As to the pre-history of this branch, one might, for example, point to Edward Malone²⁰ who, as early as in 1787, shed doubt on the assumption that Shakespeare wrote any of the three parts of *Henry VI*. Analyzing meter and rhyme in these works, Malone noted the frequent use of end-stopped and infrequent use of double endings and rhyming lines, i.e., metrical properties not characteristic of Shakespeare's plays. Likewise, Henry Weber, in 1812, in his *The Works of Beaumont and Fletcher*, analyzed the number of feminine endings in *The Two Noble Kinsmen*, and, on the basis of this analysis, attributed parts of this text to Shakespeare, other parts to John Fletcher. Almost the same conclusion was drawn by James Spedding²¹ on the basis of his quantitative analysis of feminine endings of *Henry VIII*. Interestingly enough, Spedding's text was published in the well-known *Gentleman's Magazine and Historical Review* exactly one year before de Morgan suggested quantitative analyses, as was mentioned above.

All these works thus not only preceded de Morgan's famous quotation, but also paved the way for the foundation of stylometric Shakespeare studies in the 1870s. Yet, irrespective of the importance of these studies for the development of stylometric analyses, they remained more or less "autonomous" for a long time. In this sense, it may be rather misleading when Williams, speaking of a 'school' which had its height in the foundation of the *New Shakspeare Society*, lists names not only of members belonging to that society, but also of other early proponents of "stylometrics" (Lewis Campbell, Moritz W. Drobisch, Wincenty Lutoslawski), who had no relation at all, neither to the society and its members, nor to their works.²²

Rather, on the broad and multifarious road leading to the discipline of stylometry, the above-mentioned "Shakespeare line" — whose importance must by no means be underestimated — led along an almost independent and parallel

²⁰ Malone 1887.

²¹ Spedding 1850.

²² Williams 1970: 2.

way to yet another line, which might appropriately be termed the “Plato line”. In fact, this “Plato line” was, as can be shown, composed of two filiations, which one might term the Scottish and the German.

The beginnings of this line are basically related to Scottish philosopher Lewis Campbell. In his 1867 book *The Sophistes and Politicus of Plato*, Campbell strived for a periodization of Plato’s *Dialogues*, by way of analyzing rare words occurring in a given text and thus assumed to be characteristic of it.²³ However, this book remained unnoticed even among Plato researchers for more than two decades, when Campbell related his own research to German Plato studies in the very same direction.

The beginnings of these studies in the early 1880s are related to classical philologist Wilhelm Dittenberger. Concentrating rather on the analysis of frequent words in Plato’s works, Dittenberger, completely independent of Campbell and without knowing of Campbell’s work, started in the very same direction as Campbell before him.²⁴ Dittenberger’s approach did not remain uncontradicted, among German philologists (e.g., by Eduard Zeller), but it also gave rise to a number of studies from scholars who followed his tracks (e.g., Arthur Frederking, Hermann Hoefler, Martin Schanz, and others).

Yet, the common traits of Campbell’s and Dittenberger’s results, and their similar methodological orientations, went generally unnoticed, until they were discussed years later, in Campbell’s²⁵ review of Constantin Ritter’s²⁶ *Untersuchungen über Plato*. This short and neglected review, published in the internationally renowned journal *Classical Review*, might be another missing link in the history of stylometry, the more since Campbell not only relates his own earlier research to the contemporary German studies, but also refers to the British Shakespeare tradition, summarizing all three lines under the term ‘quantitative criticism’.

But Campbell’s name and his work continued to remain unknown in the German academic world. As Ritter pointed out much later,²⁷ it was Polish scholar Wintency Lutosławski who acquainted German scholars with Campbell’s work — and this means, not before 1895. In this year, Lutosławski²⁸ published an article in the renowned German journal *Archiv für Geschichte der Philosophie*, in which he discussed at length results of both the German and the Scottish “Plato lines”,

²³ Campbell 1867.

²⁴ Dittenberger 1881.

²⁵ Campbell 1889.

²⁶ Ritter 1888.

²⁷ Ritter 1903a: 247.

²⁸ Lutosławski 1895a.

thus combining their results, but without introducing yet the term ‘stylometry’.²⁹ Two years later, Lutosławski³⁰ published his book *The Origin and Growth of Plato’s Logic* in English, dedicating it to Campbell, on the occasion of the 30th anniversary of his forgotten book. And in the same year, a short review of his presentation of this book (which was only to be published later that year) in the Oxford Philological Society, in May 1897, was published in the *Classical Review*.³¹ This article, entitled “On stylometry”, thus is in fact the first prominent mention of this term, and Lutosławski himself declared it to be the “first public explanation of the method of stylometry”;³² as to ‘stylometry’ as a term, this article therefore is of historical importance. In fact, Lutosławski was soon internationally credited for his works on stylometry, as can be particularly seen from Tannery’s³³ or Lyon’s³⁴ French surveys.

In his article, Lutosławski presented stylometry, defining it as “measuring stylistical affinities”, as “a new and powerful instrument of historical research”; moreover, Lutosławski proposed the following “law of stylistical affinity”³⁵ (which he understood as a general psychological law):

Of two samples of text of the same author and of the same size, that is nearer in time to a third, which shares with it the greater number of units of affinity.³⁶

As compared to this definition, Lutosławski phrased this “law” slightly differently in his book, published later that year:

Of two works of the same author and of the same size, that is nearer in time to a third, which shares with it the greater number of peculiarities, provided that their different importance is taken into account, and that the number of observed peculiarities is sufficient to determine the stylistic character of all three works.³⁷

²⁹ A Polish summary of this article appeared in the same year in Polish; cf. Lutosławski 1895b.

³⁰ Lutosławski 1897a.

³¹ Lutosławski 1897b.

³² Ibid. 286.

³³ Tannery 1899.

³⁴ Lyon 1902.

³⁵ Lutosławski 1897b: 284f.

³⁶ Ibid. 284.

³⁷ Lutosławski 1897a: 152.

Also in the same year, on June 18, 1897, Lutosławski made a presentation in French at the Académie des Inscriptions et Belles-Lettres, which was published in 1898 under the title “Principes de stylométrie appliqués à la chronologie des œuvres de Platon” in *Revue des études grecques*. Here, he formulated the stylometric law as follows:

De deux échantillons de textes de la même longueur comparés à un troisième sous le rapport de style, celui qui présente une affinité stylistique décidément plus grande avec l'étalon de comparaison, lui sera plus proche quant à la date de la composition, pourvu qu'un nombre suffisant de stylèmes ait été étudié et inclus dans la calculation des affinités.³⁸

We can thus be sure that the term ‘stylometry’ came into being in 1897. An additional argument in favor of this assumption is the fact that one year before, on May 16, 1896, Lutosławski had delivered a presentation entitled “Sur une nouvelle méthode de déterminer la chronologie des dialogues de Platon” at the Institut de France devant l'Académie des sciences morales et politiques. In that lecture, which was later published in the *Compte rendu des séances et travaux de l'Académie des sciences morales et politiques*, he had not yet used the term ‘stylometry’, but had generally spoken of “la statistique du style”.³⁹

Interestingly enough, in his extensive German article published in *Zeitschrift für Philosophie und philosophische Kritik*,⁴⁰ Lutosławski would more clearly distinguish between linguostatistics (Sprachstatistik), on the one hand, and stylometry, on the other: according to his opinion, traditional linguostatistics used to find exclusive features (ausschließliche Merkmale) in order to unite particular works into groups; it did not deal with random phenomena, and did not make a difference between a language feature (Sprachmerkmal) and a stylistic feature, or ‘styleme’.⁴¹

As Lutosławski points out in his German-language contribution “Stylometrisches”, published one year later in the journal *Zeitschrift für Philosophie und philosophische Kritik*,⁴² the relevance of a styleme for issues of chronological order must never be seen in isolation, but in context of the whole system under study:

³⁸ Lutosławski 1898a: 65.

³⁹ Lutosławski 1896: 25.

⁴⁰ Lutosławski 1897c.

⁴¹ *Ibid.* 218.

⁴² Lutosławski 1898b.

Ein einzelnes Stylem, oder selbst einige Styleme, berechtigen nie zu chronologischen Schlüssen. Der Styl eines Autors beruht in jeder Periode seiner schriftstellerischen Thätigkeit auf sehr vielen Stylemen, wovon in jedem Werk nur ein Teil zur Verwendung gelangt. Werke, die mehr Styleme miteinander gemein haben, sind nur dann zeitlich wahrscheinlich einander näher gestellt, wenn die Gesamtzahl der untersuchten Styleme hinreichend ist, um die Styleigentümlichkeit zu bestimmen.⁴³

Thus objecting to a deterministic and arguing in favor of a probabilistic understanding of ‘law’ — in a similar way as Frederking had done before him⁴⁴ —, Lutosławski arrives at the following definition of ‘stylometry’ in German:

Von zwei Werken desselben Schriftstellers und derselben Größe ist dasjenige der Zeit nach einem dritten näher, welches mit ihm die grössere Zahl stilistischer Eigentümlichkeiten teilt, vorausgesetzt, dass deren verschiedene Wichtigkeit in Rechnung gezogen wird, und dass die Zahl der beobachteten Eigenheiten ausreichend ist, den stilistischen Charakter aller drei Werke zu bestimmen.⁴⁵

It deserves to be mentioned that with regard to the methodology of his approach, Lutosławski⁴⁶ refers to his educational background at the University of Tartu (Dorpat), as he had already done in the preface to his earlier book on *Plato’s Logic*.⁴⁷ Lutosławski particularly mentions his teachers in natural sciences, internationally renowned experts in chemistry, physics, physiology, mineralogy, astronomy and mathematics, scholars such as Wilhelm Ostwald, Gustav Bunge, Arthur von Oettingen, Carl Schmidt, Andreas Lindstedt, Johannes Lemberg. They all, in addition to philosopher Gustav Teichmüller and linguist Baudouin de Courtenay, had been Lutosławski’s teachers, and he explicitly pointed out that without their training, it would never have been possible to establish his theory of stylometry. In fact, as early as 1885, in his master’s thesis “*Aesthetisches Studium. Über das phonetische Element in der Poesie*”⁴⁸ — some results of which were published in the

⁴³ Ibid. 38; in a similar way, Zeller (1887: 218) had criticized the concentration on one stylistic trait only a decade before Lutosławski.

⁴⁴ Frederking 1882: 526.

⁴⁵ Lutosławski 1897c: 205.

⁴⁶ Lutosławski 1898b: 41.

⁴⁷ Lutosławski 1897a: ix.

⁴⁸ Lutosławski’s thesis from 1885 has recently been critically edited by Pawłowski et al 2008. For further information, cf. also Pawłowski 2004.

internationally renowned journals *Internationale Zeitschrift für Sprachwissenschaft*⁴⁹ and *Zeitschrift für Völkerpsychologie und Sprachwissenschaft*⁵⁰, Lutosławski had studied phonological issues in (poetic) texts, based on methodological analogies from the field of physical chemistry.

It would definitely exceed the framework of this text, to shed more light on this background, which remains to be discussed at appropriate length in some other context. The same holds true for the further development of stylometry, which cannot be discussed here in detail. Yet, some specific details deserve mention, which concern the methodological transfer from classical or antique authors to modern writers, and which thus are characteristic of the conversion of the different stylometric lines outlined above.

One of these studies is Campbell's contribution "The Growth of A Thinker's Mind",⁵¹ in which he analyzes, among others, Lord Tennyson's *Morte d'Arthur / Passing of Arthur / Idylls of the King*, written and published under different titles at different times in Tennyson's life. This study continues to be characterized by the concentration on matters of chronological order, as is Zeller's⁵² study "Sprachstatistisches": suggesting to first test stylometric methods with contemporary authors, the chronology of whose works is well known, before applying it to classical authors, Zeller concentrated on the analysis of 19th century texts by German writer David Friedrich Strauß. In a critical debate with both of Zeller's methods, Ritter, devoting two studies to Goethe texts from different periods of his life, showed that the analysis of isolated phenomena is no appropriate basis for any kind of statistical conclusions.⁵³ An early synchronous study is Polish scholar Władysław Ćwik's stylometric analysis of the texts by Juliusz Słowacki, a Polish Romantic writer.⁵⁴ Finally, Russian Nikolay A. Morozov's study "Лингвистические спектры"⁵⁵ represents an extension from classical Plato studies to modern Russian authors like N. V. Gogol', A. S. Puškin, L. N. Tolstoy, and I.S. Turgenyev.

One can say that from this time onwards, stylometric approaches went different ways in different academic societies; and at different periods of their development, they became established to different degrees. It is not the objective of this text, which is intended to concentrate on the rise and early development of the concept

⁴⁹ Lutosławski 1885.

⁵⁰ Lutosławski 1887.

⁵¹ Campbell 1898.

⁵² Zeller 1897.

⁵³ Ritter 1903a,b.

⁵⁴ Ćwik 1909.

⁵⁵ Morozov 1915.

and the term of stylometry, to pursue these lines in detail. In any case, it is self-evident from the foregoing discussion that in the field of stylometry, we have more questions than answers, not only from a contemporary methodological perspective, but also with regard to the history of this discipline. It should have become obvious that much work and patience is still needed to arrive at a reliable historical depiction of this important concept which, due to its combination of linguistic, literary and mathematical/statistical methods will contribute to a better understanding of, or maybe even open new insights into the development of history of science, in general.

LITERATURE

- BINONGO J. N. G. – SMITH, M. 1996. Stylometry. *Notes & Queries* 43, № 4, 448–452.
- BURROUGHS J. 1899. The vital touch in literature. *The Atlantic Monthly* 83, 399–406.
- CAMLOT J. 2008. *Style and the Nineteenth-Century British Critic: Sincere Mannerisms*. Ashgate.
- CAMPBELL L. 1867. *The Sophistes and Politicus of Plato*. Oxford.
- CAMPBELL L. 1889. [Rev.:] Ritter (1888). *Classical Review* 3, 28–29.
- CAMPBELL L. 1898. The Growth of A Thinker's Mind. A Study in Platonic Chronology. *The Fortnightly Review* 63, № 1, 36–56.
- COLLINGWOOD R. 1945. *The Idea of Nature*. Oxford.
- ĆWIK W. 1909. *Badania stylometryczne nad językiem Juliusza Słowackiego*. Lwów.
- DITTENBERGER W. 1881. Sprachliche Kriterien für die Chronologie der Platonischen Dialoge. *Hermes: Zeitschrift für Klassische Philologie* 16, 321–345.
- EARLE J. 1890. *English Prose: Its Elements, History and Usage*. London.
- EDDY H. 1887. The characteristic curves of composition. *Science* 9, № 216, 297.
- FREDERKING A. 1882. Sprachliche Kriterien für die Chronologie der platonischen Dialoge. *Neue Jahrbücher für Philologie* 125, 534–541.
- GRIEVE J. W. 2005. *Quantitative authorship attribution: a history and an evaluation of techniques*. B.A., Simon Fraser University.
- HOLMES D.I. 1998. The Evolution of Stylometry in Humanities Scholarship. *Literary and Linguistic Computing* 13, № 3, 111–117.
- LUTOSŁAWSKI W. 1885. Über die Konstanz des phonetischen Charakters. *Internationale Zeitschrift für allgemeine Sprachwissenschaft*, 2, 393.

- LUTOSŁAWSKI W. 1887. Über das phonetische Element in der Poesie. *Zeitschrift für Völkerpsychologie und Sprachwissenschaft* 17, 215–220.
- LUTOSŁAWSKI W. 1895a. Über die Echtheit, Reihenfolge und logische Theorien von Platos drei ersten Tetralogien. *Archiv für Geschichte der Philosophie* 8, 67–114.
- LUTOSŁAWSKI W. 1895b. [Résumé] O trzech pierwszych tetralogiach dzieł Platona (Sur les trois premières tétralogies de Platon). *Bulletin international de l'Académie de Cracovie* [= *Anzeiger der Akademie der Wissenschaften in Krakau*], 268–277.
- LUTOSŁAWSKI W. 1896. 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.
- LUTOSŁAWSKI W. 1897a. *The Origin and Growth of Plato's Logic*. London etc.
- LUTOSŁAWSKI W. 1897b. On stylometry. *Classical Review* 11, 284–286.
- LUTOSŁAWSKI W. 1897c. W. Lutosławskis Theorie der Stylometrie auf die Platonische Frage angewendet. *Zeitschrift für Philosophie und philosophische Kritik* 110, 171–219.
- LUTOSŁAWSKI W. 1898a. Principes de stylométrie. *Revue des études grecques* 11, 61–81.
- LUTOSŁAWSKI W. 1898b. Stylometrisches. *Zeitschrift für Philosophie und philosophische Kritik* 112, 34–41.
- LYON G. 1902. Platon et la stylométrie. *Revue de synthèse historique* 4, 1–17.
- MALLOCK W. 1892. Le style c'est l'homme. *New Review* 6, 441–454.
- MALONE E. 1787. *A dissertation on parts I, II and III of Henry VI tending to show that those plays were not written originally by Shakespeare*.
- MENDENHALL T. C. 1887. The Characteristic Curves of Composition. *Science* 9, (214), 237–246.
- MENDENHALL T. C. 1901. A Mechanical Solution of a Literary Problem. *Popular Science Monthly* 60, 97–105.
- MENDENHALL T. C. 1904. Shorter Articles and Discussion: Characteristic curves of composition. *Popular Science Monthly* 65, № 4, 373–377.
- MOROZOV N. A. [Морозов Н. А.] 1915. Лингвистические спектры. Средство для отличения плагиатов от истинных произведений того или другого известного автора. *Известия отделения русского языка и словесности Императорской академии наук XX*, № 1–4, 95–127.
- PAWŁOWSKI A. 2004. Travail de maîtrise de Wincenty Lutosławski: «Aesthetisches Studium. Über das phonetische Element in der Poesie». Description matérielle et analyse du contenu. *Organon* 33, 121–139.
- PAWŁOWSKI A. – HOFBAUER, H. – TWORKE, A. 2008. Kritische Edition des

- Werks *Aesthetisches Studium. Das phonetische Element in der Poesie* von Wincenty Lutosławski. *Organon* 37, № 40, 177–255.
- RITTER C. 1888. *Untersuchungen über Plato. Die Echtheit und Chronologie der Platonischen Schriften nebst Anhang; Gedankengang und Grundanschauungen von Platos Theaetet*. Stuttgart.
- RITTER C. 1903a. Die Sprachstatistik in Anwendung auf Platon und Goethe. *Neue Jahrbücher für das Klassische Altertum* 4/5, 241–261, 313–325.
- RITTER C. 1903b. Die Sprachstatistik in Anwendung auf Goethes Prosa. *Euphorion. Zeitschrift für Literaturgeschichte* 10, 558–578.
- SHERMAN L. A. 1888. Some Observations upon the Sentence-Length in English Prose. *The University of Nebraska Studies* 1, № 11, 119–130.
- SHERMAN L. A. 1892. On Certain Facts and Principles in the Development of Form in Literature. *University Studies (University of Nebraska, Lincoln Campus)* 1, № 4, 337–366.
- SPEDDING J. 1850. Who wrote Shakespeare's *Henry VIII*? *The Gentleman's Magazine*, 115–123.
- TANNERY P. 1899. La stylométrie: ses origines et son présent. *Revue philosophique de la France et de l'étranger* 3, № 47, 159–169.
- WELLEK R. – WARREN, A 1949. *Theory of Literature*. New York.
- WILLIAMS C. B. 1970. *Style and Vocabulary: Numerical Studies*. London.
- ZELLER E. 1887. Über die Unterscheidung einer doppelten Gestalt der Ideenlehre in den platonischen Schriften. *Sitzungsberichte der Preussischen Akademie der Wissenschaften* [Berlin] 13, 192–220.
- ZELLER E. 1897. Die deutsche Litteratur über die sokratische, platonische und aristotelische Philosophie. *Archiv für Geschichte der Philosophie* 10, 557–595.