Journal of Current Scientific Research

What Can a Stylometric Analysis Contribute to the Diary af Saint Faustina Kowalska?

Raúl Isea[1,*]

[1]Fundación IDEA. Hoyo de la Puerta, Baruta, Venezuela

Abstract

Saint Maria Faustina Kowalska's Diary records the words of Christ revealed to her between 1934 and 1938. The frequency of the words used by God and Sor Faustina could be statistically determined by stylometric analysis since each has its own linguistic footprint. To do this, we analyze various parts of the Diary in three different editions (Polish, Spanish, and English). The results show a linguistic style characteristic of Sister Faustina and another to God. So this paper confirms that there was a conversation between the two.

Letter

Open Access &

Peer-Reviewed Article

DOI: 10.14302/issn.2766-8681.jcsr-24-5075

Corresponding author:

Raúl Isea, Fundación IDEA. Hoyo de la Puerta, Baruta, Venezuela .

Keywords:

Dialog; Diary; Linguistic footprint; St. Faustina; Stylometry

Received: April 15, 2024

Accepted: May 03, 2024

Published: May 11, 2024

Citation:

Raúl Isea (2024) What Can a Stylometric Analysis Contribute to the Diary af Saint Faustina Kowalska?. Journal of Current Scientific Research - 2(2):15-22. https://doi.org/10.14302/issn.2766-8681.jcsr-24-5075

Introduction

Maria Faustina Kowalska (1905–1938) was the third daughter of 10 brothers in the family of Mariana Babel and Stanisław Kowalski, born in Głogowiec, Poland. Her spiritual vocation was expressed from the time she was seven years old (Diary, 1404) until she arrived at the final convent of the Sisters of the Mother of God of Mercy on August 1, 1925 (Diary, 1404)

In this convent, she wrote her diary at the express command of Jesus Christ and in obedience to her confessors, Blessed Michał Sopoćko (1888–1975) and Jesuit Józef Andrasz SJ (1891–1963), from 1934 until three months before she death in 1938. At that time, she was the secretary of the Good Lord (Diary, 1695), and although her education was not high, this did not prevent her from rewriting the Gospel of Mercy in the 20th century, as Saint John Paul II expressed [12].

Sister Faustina is not a theologian, but she explained to us that her mission is to bring and announce mercy to the whole world, above all to sinners, and to promote devotion to Divine Mercy. Mercy God. Knowledge of His mercy is gained through faith and reason, but above all through mystical experience. She was canonized by Saint John Paul II in the Holy Year 2000 in

his homily on April 30 of the same year. Saint John Paul II said, "Christ entrusted to him the message of mercy; he transmitted it to everyone so that they might learn to know better the true face of God." [13].

Diary

The Diary is a faithful record of the words of the merciful God, with which he awarded Sister Faustina the title of Secretary of Mercy. (Diary, 1605). It is characterized by the absence of corrections, amendments, blank pages, and grammatical errors. It consists of 477 pages divided into six notebooks, written on

both sides and without page numbers. In it are reflected all his thoughts, and she emphasizes the words that the Lord dictated to her.

St. Faustina ordered the publication of her Diary after her death to "console souls" (Diary, 1390). These notebooks remain in the possession of the congregation. It was originally transcribed by Sister Javiera Olszamowską, but it was not a literal transcription, which led to errors. Later, Father Isidor Borkiewicz, together with Sister Beata Piekut, rewrote it from the original and published it in 1981, serving as the basis for all translations.

Stylometry

Stylometry is summarized as a statistical method based on the analysis of each author's word frequency [17; 6; 14]. The term was coined by the Polish writer Wincenty [10] when he dated Plato's Dialogues (Lutosławski, 1898) and was originally based on the chi-square method for analyzing relationships between word length and the frequency of their appearance [11].

More recently, it has been used to determine the authorship of certain unknown or questionable works, such as a comedy originally attributed to Miguel Bermúdez in the National Library of the West. Spain, but this method can detect that it belongs to Lope de Vega. [3].

Another example is Ulysses, which was published in 1922 by Irishman James Joyce, but stylistic studies concluded that it was written by five other people not mentioned in the work. [15].

Based on the above, the work analyzes some sentences attributed to the person of Jesus Christ and others to the mystic of Krakow to show that they do not reproduce the same linguistic footprint, and therefore, there was a dialogue between the two persons, which removes the doubt that the passages said by Jesus were the product of the imagination of the Saint.

Methodology

A stylometric study of the Diary written by Santa Maria Faustina Kowalska is carried out in three different versions, corresponding to the Spanish, English, and Polish editions. The fourth edition in Spanish of the Marian Fathers of the Immaculate Conception of the Blessed Virgin Mary was used in 2001. The English version is from the 2005 edition of Marian Press Stockbridge (a translation of the 1981 Polish version). The Polish version is published by Zgromadzenie Sióstr Matki Bożej Miłosierdzia (card. Franciszek Macharski metropolitan Krakowski).

This method is described by both [4] and [7, 8] and is roughly summarized below. For each language of the Diary, a corpus was built from the sayings of Sister Faustine and the Lord (presented in Table 1). On that basis, determine the frequency of occurrence of words, then determine the delta function, which is a linguistic measure capable of distinguishing the authorship of text according to the definition proposed by [2].

Some articles have questioned the use of the delta function, but Burrows has shown that it is an excellent technique for determining the authorship of a work [16]. This function simply determines the frequency variation of the most frequently occurring words in the text using z-scores [16]. Remember that the z-score is a measure of the relative frequency difference minus the value from the mean divided by the standard deviation [19].

In the next step, a distance-based matrix derived from the frequency of occurrence of the words was calculated. [18]. Distance calculations are generally of the Manhattan, Euclidean, and other types [18].

The results can be displayed using a dendrogram, a type of tree in which similar distances are grouped into a single conglomerate or cluster. In the context of this article, the latter means that sentences written by the same person should be grouped into the same branch or node [5]. All calculations were performed using the Python programming language.

Table 1. Nomenclature of the passages of the Diary of Sister Faustina used for stylometric analysis in the Spanish, English, and Polish versions of the journal

Spanish version		**English version**		**Polish version**	
Nomenclature	Paragraph	Nomenclature	Paragraph	Nomenclature	Paragraph
JSF1	1777	J1	154	J1	229
JSF2	300	J2	156	J2	229
JSF3	1327	J3	165	J3	229
JSF4	1320	J4	169	J4	232
JSF5	299	J5	177	J5	282
JSF6	1032	J6	178	J6	639
JSF7	687	J7	186	J7	687
JSF8	367	J8	219	J8	699
SF1	450	J9	229	J9	739
SF2	529	SF1	154	J10	714
SF3	843	SF2	156	SF1	229
SF4	462	SF3	165	SF2	230
SF5	118	SF4	170	SF3	231
SF6	119	SF5	177	SF4	233
SF7	303	SF6	178	SF5	236
SF8	136	SF7	188	SF6	254
SF9	477	SF8	207	SF7	655
SF10	208	SF9	220	SF8	677
SF11	364	SF10	230	SF9	707
SF12	78	SF11	233	SF10	713
SF13	78			SF11	720
SF14	225			SF12	724
				SF13	731
				SF14	741

Results

Based on the above, the work analyzes some sentences attributed to the person of Jesus Christ and others to the mystic of Krakow to show that they do not reproduce the same footprint. Figure 1 shows the result of the normalization of the frequency of words obtained from the Spanish-translated Diary, where it can be observed that Sister Faustina more often uses the words "*de Dios*", "*lo que*", "*a Dios*",

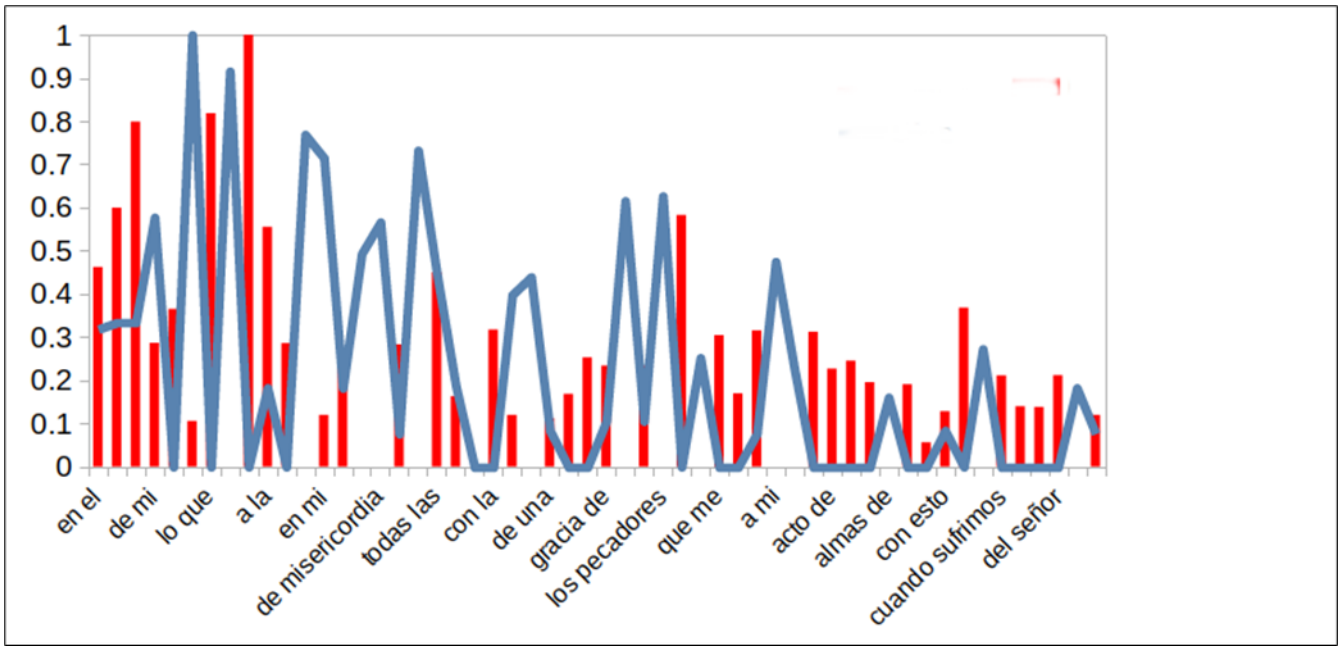

Figure 1. Normalized frequency of words obtained from the Diary translated into Spanish (bigram equal to 2) grouped in the texts of Sor Faustina (red color) and God (blue color).

"*a la*", rather than the phrases used by God. Instead, the words used by God, which differ from those used by St. Faustina, are "*de mi*", "*en mi*", "*de este*", "*la misericordia*", "*mi corazón*", "*mi misericordia*" among others. (figura 1). From this, you can appreciate the difference in the use of language between the two people.

Figure 2 shows the results of normalizing the different frequencies of words obtained in the English translation of the journal, in which Sister Faustine often uses the following words: "*of god*", "*I will*",

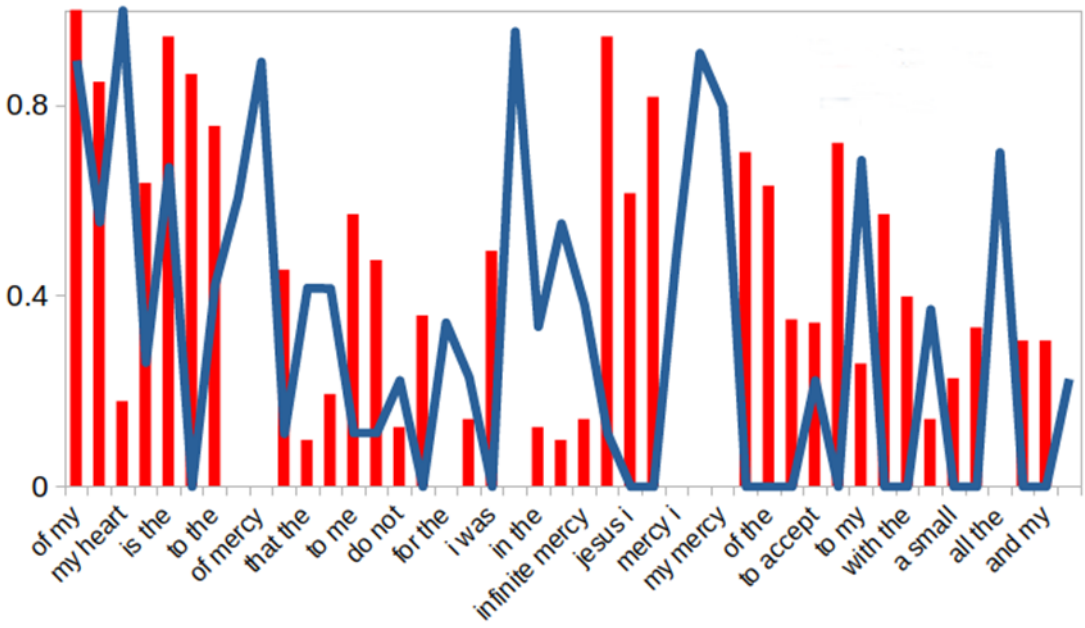

Figure 2. Normalized frequency of words found in the journal's English texts (bigram equals 2) grouped in the Sister Faustine (red color) and God texts (blue)

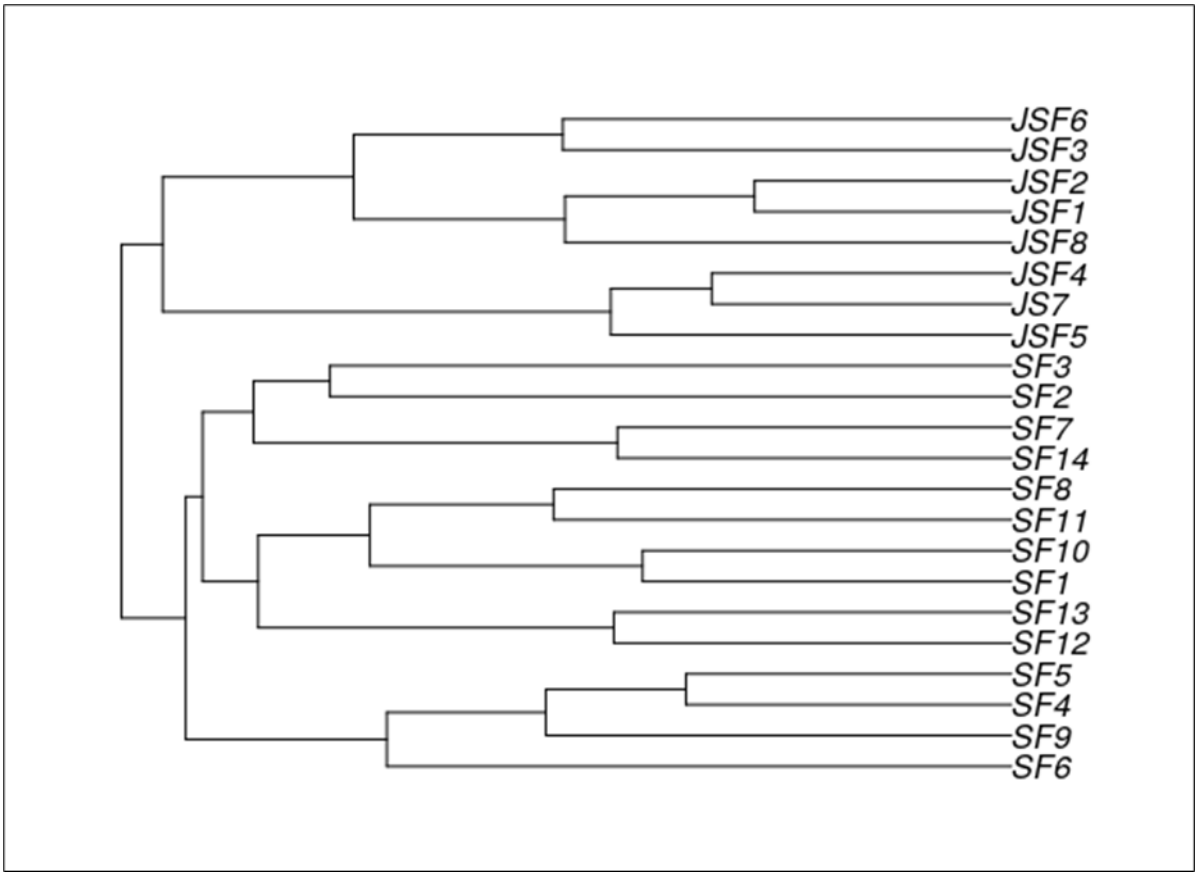

Figure 3. Rectangular dendrogram obtained from stylometric analysis obtained from the Spanish translation of Sister Faustine's Diary, visualized using iTOL [9].

"*is the*", "*that is*", "*to me*", "*it is*", "*me like*", "*my soul*", "*to god*", among others. On the other hand, the expressions used by God are"*my heart*", "*i desire*", "*of mercy*", "*in my*", "*my daughter*", "*my mercy*", "*all that*" and other words. (no research on this studio since it discovered the idiom).

Figure 3 shows a rectangular dendogram obtained from the Spanish version of the Diary of Sister Faustina. The first thing that draws attention is that there are two well-defined groups (conglomerates), the first consisting of all the sentences spoken by Jesus (they begin with the letters JSF), while the second corresponds to the words of Sister Faustina (inician con las letras SF). Thus, two different linguistic prints corresponding to a dialogue between two people are observed.

Figure 4 shows a dendrogram obtained from the English translation of the Diary of Sister Faustina, where two large conglomerates are observed, which in turn are divided into two different subgroups. The first of them consists of two other subgroups: the sentences of Jesus (J2, J1, J8, J5, J4, J7) and those spoken by Sister Faustina (SF8 and SF9). The second large conglomerate is in turn divided into two subgroups. The first of them corresponds to a division of the sentences spoken by Sister Faustina (SF11, SF12) and another branch, which in turn is the division between SF7 and two sentences of Jesus (J3, J6), while the rest is formed by the phrases of Sister Faustina.

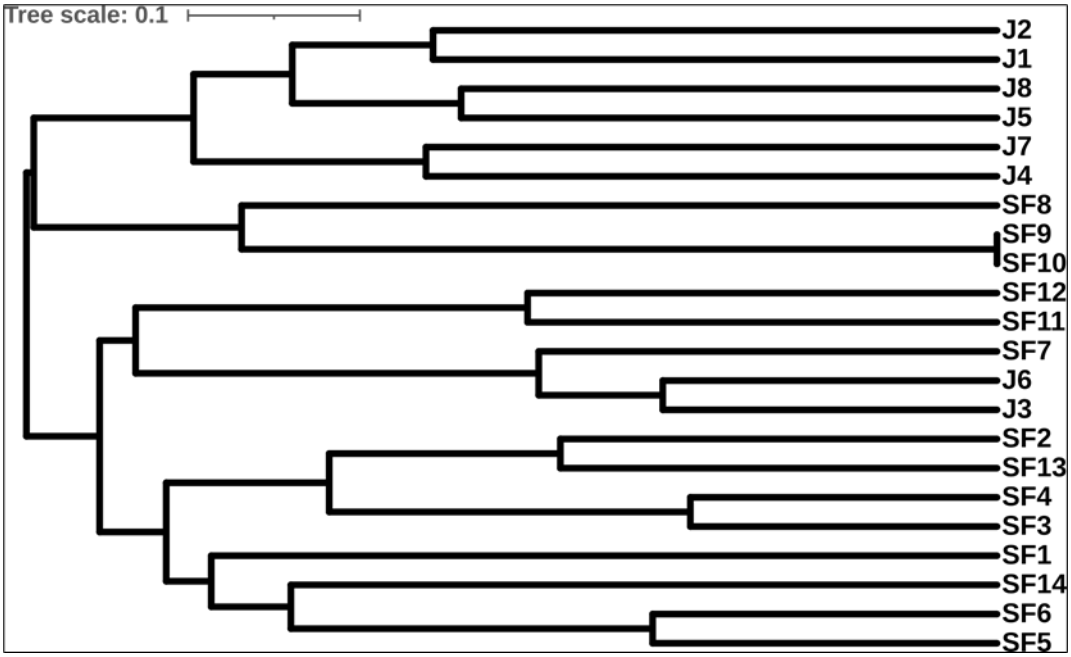

Figure 4. Rectangular dendrogram obtained from stylometric analysis obtained from the English translation of Sister Faustine's Diary, visualized using iTOL [9].

Figure 5 shows the dendrogram obtained from the Polish version. In it, you can see a sentence spoken by Sor Faustina (SF2) with a linguistic style different from the rest of her sentences. Then there are three large conglomerates, where the first corresponds to the words spoken by Sister Faustina (SF13, SF3, SF1, SF6, SF9, SF8, SF12, SF7, SF4). The second conglomerate is made up of two subgroups: one composed of the sentences of Sister Faustina and one formed with the words of God. The last great conglomerate consists exclusively of sentences spoken by Jesus.

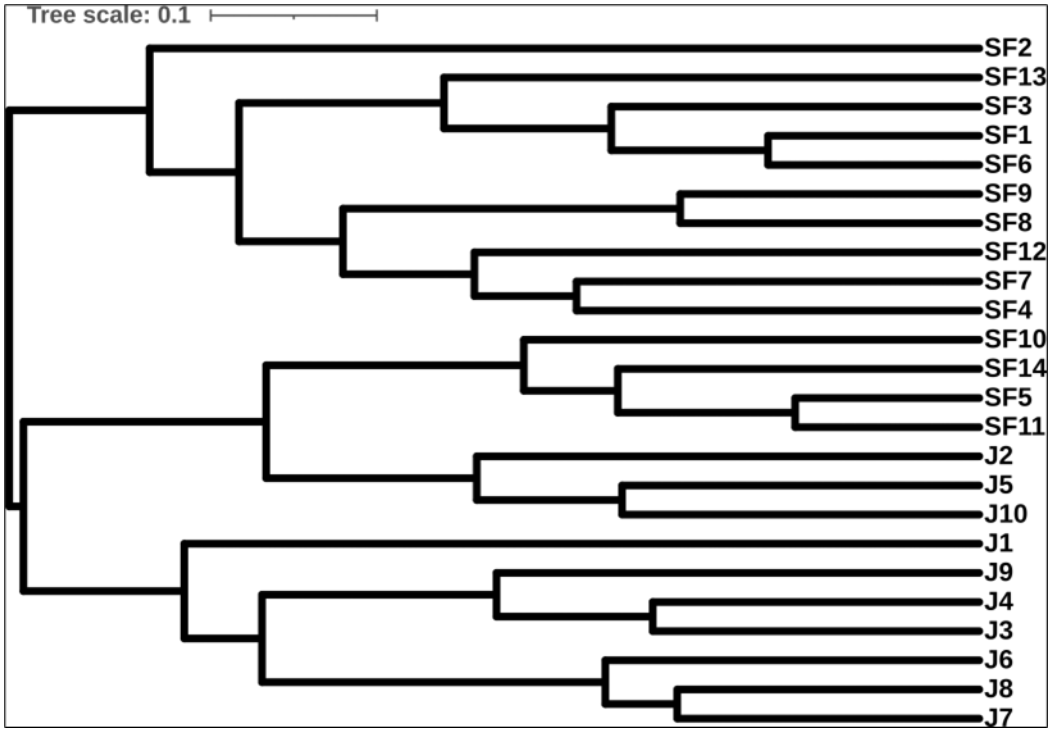

Figure 5. Rectangular dendrogram obtained from stylometric analysis obtained from the Polish translation of Sister Faustine's Diary, visualized using iTOL [9].

Conclusion

The goal of the work was to find the linguistic footprints present in the conversation between Sister Faustina and Christ using a stylometric methodology. There are differences in the frequency of the appearance of words between the two.

According to the dendrogram derived from the Spanish, English, and Polish versions, the linguistic styles did not merge into a single branch, a clear sign that there was dialogue between the two people. On this basis, we conclude that the Diary is not a product of St. Faustina's imagination, but it recreates the enduring dialogue between her and God.

References

1. Bensalem, I., Rosso, P. y Chikhi, S. (2019). On the use of character n-grams as the only intrinsic evidence of plagiarism, *Language Resources and Evaluation*. 53(3), 363–396. https://doi.org/10.1007/s10579-019-09444-w
2. Burrows, J.F. (2002). Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship. *Literary and Linguistic Computing*, 17(3), 267-287. https://doi.org/10.1093/llc/17.3.267
3. Bvmc (2014, January 23) "Descubierta una comedia inédita de Lope de Vega" Available at https://blog.cervantesvirtual.com/descubierta-una-comedia-inedita-de-lope-de-vega/
4. Daelemans, W. (2013). Explanation in Computational Stylometry. In: Gelbukh, A. (eds) Computational Linguistics and Intelligent Text Processing. CICLing 2013. Lecture Notes in Computer Science, vol 7817. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37256-8_37
5. Eder, M. (2017), Visualization in stylometry: Cluster analysis using networks, *Digital Scholarship in the Humanities*, 32(1), April 2017, 50–64. https://doi.org/10.1093/llc/fqv061
6. Fuller, S. y O'Sullivan, J. (2017). Structure over Style: Collaborative Authorship and the Revival of Literary Capitalism. *Digital Humanities Quarterly*, 11 (1). http://dx.doi.org/10.17613/M6BH0D
7. Isea, R. (2023a). A stylometric analysis of two sentimental letters by Albert Einstein and Simón Bolívar. *Journal of Language Research*, 1(1), 15-20.
8. Isea, R. (2023b). Stylometric analysis of Lafond's letter that never wrote Simón Bolívar to San Martín. *Journal of Current Scientific Research*, 2(1), 24-30. https://doi.org/10.14302/issn.2766-8681.jcsr-23-4727
9. Letunic, I. y Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, 49(W1), W293–W296. https://doi.org/10.1093/nar/gkab301
10. Lutoslawski, W. (1898). Principes de stylométrie appliqués à la chronologie des œuvres de Platon. *Revue des Études Grecques*. 11(41), 61–81. https://doi.org/10.3406/reg.1898.5847.
11. Mendenhall, T. (1887). The characteristic curves of composition. *Science*, 214, 237-249.
12. Pablo, J. II (1998, s.f.). Carta Encíclica Dives in Misericordia. *Editorial Editrice Vaticana*. https://www.vatican.va/content/john-paul-ii/es/encyclicals/documents/hf_jp-ii_enc_30111980_dives-in-misericordia.html
13. Pablo, J. II (2000, s.f.). Homilía del Santa Padre Juan Pablo II. https://www.vatican.va/content/

john-paul-ii/es/homilies/2000/documents/hf_jp-ii_hom_20000430_faustina.html (recuperado 30 de Agosto)

14. Schaalje, G.B., and others (2011). Extended nearest shrunken centroid classification: A new method for open-set authorship attribution of texts of varying sizes, *Literary and Linguistic Computing*, Volume 26, Issue 1, April 2011, Pages 71–88, https://doi.org/10.1093/llc/fqq029
15. Schoenbaum, S. (2018). Internal evidence and Elizabethan dramatic authorship; an essay in literary history and method, p. 196. Northwestern University Press (15 Octubre 2018) ISBN 0810138662
16. Škorić, M., Stanković, R., Ikonić, N. M., Byszuk, J., Eder, M. (2022). Parallel Stylometric Document Embeddings with Deep Learning Based Language Models in Literary Authorship Attribution. *Mathematics*, 10(5):838. https://doi.org/10.3390/math10050838
17. Stamatatos, E. (2009). A survey of modern authorship attribution methods. *JASIST*, 60 (3), 538–556. https://doi.org/10.1002/asi.21001
18. Stanikunas, D., Mandravickaite, J. y Krilavicius (2017). Comparison of distance and similarity measures for stylometric analysis of Lithuanian texts. CEUR Workshop proceedings [electronic resource]: ICYRIME 2017 : proceedings of the symposium for young researchers in informatics, mathematics and engineering, Kaunas, Lithuania, April 28, 2017. Aachen : CEUR-WS, 2017, Vol. 1852
19. Stefan, E., et al (2017). Understanding and explaining Delta measures for authorship attribution, *Digital Scholarship in the Humanities*, 32(suppl_2), ii4–ii16. https://doi.org/10.1093/llc/fqx023