Stylometric analysis of Lafond's letter that never Wrote Simón Bolívar to San Martín

Raúl Isea[1,*]

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

Abstract

The goal is to analyze that Lafond's letter by stylometric methods, supposedly written by Simón Bolívar to General San Martín about the destiny of Ecuador. The Delta function was calculated after evaluating 16 letters from Simón Bolívar and including another 11 letters from San Martín. The reason for including San Martín's letter was to verify if the method used could distinguish between the two authors. A linguistic corpus was constructed using functional words, and a dendrogram was used to visualize the result. Finally, it is concluded that the letter to Lafond is false. Simón Bolívar never wrote this letter.

Introduction

Recently, several documents have appeared that have been collected at different times, and many of them have been auctioned or donated to various institutions. For example, the letter signed by Simón Bolívar for the victory at the Battle of Cúcuta in 1813 was auctioned by Doyle in 2016 for $23,750 [1]. Another example is the 1521 letter by Hernán Cortés that was sold for $32,500 in 2017, and two years later, it was discovered that the letter was stolen from the Historical Archives of the Nation of Mexico [2].

Other documents are also emerging that have subsequently been proven to be false, for economic reasons or even for public recognition. A famous example is Galileo's 1610 manuscript that was exhibited at the University of Michigan Library and later proved to be a forgery of Tobia Nicotra [3].

Therefore, it is necessary to develop computational methodologies that enable us to verify the authenticity of documents and to be able to rule out any counterfeiting with the help of information technologies. Thanks to this, it is possible to carry out linguistic studies using the tools generated by a discipline called stylometry, primarily dedicated to recognizing patterns in the written language [4-6].

Stylometry is a discipline that began to be established by the Polish writer Wincenty Lutoslawski (1863–1954) to determine the chronology of the Plato Dialogues [7], as well as the author's discovery of an unknown comedy that was in the National Library of Spain and was recently attributed to Lope de Vega instead of Miguel Bermúdez, as indicated in the work file [8]. Another example is the work of who have not been included as authors of the work [9].

Another example was the analysis of the Book of Mormon (wrote by Joseph Smith

and Sidney Rigdon), where it is indicated that the author was not Joseph Smith [10], but later work suggested that it was none of them [11]. In light of the foregoing, the present paper examines the letter Lafond allegedly written by Libertador Simón Bolívar to General José de San Martín, in which the fate of the Republic of Ecuador was being conferred [12].

Letter's Lafond

In 1940, a letter from Simón Bolívar to General José de San Martín that was included in the book "*San Martín y Bolívar en la entrevista de Guayaquil a la luz de nuevos documentos definitivos*" by Eduardo Colombres Mármol suggested that the recently formed Republic of Ecuador be incorporated into either Colombia or Peru [12]. This letter is known as Lafond's letter. We will take into consideration the letter written by Simón Bolívar to General San Martin on 1822, which is one of the letters that were included in that book. It is interesting to note that there is not the original letter, only a transcription of it, and there is neither evidence of it nor any reference to it in other records of Bolívar.

Furthermore, various works based on Eduardo Colombres' book have established the veracity of the letters between the two Hispano-American liberators [13,14]. Gabriel Pedro Mara Lafond de Lurcy (1801–1866), a Frenchman who served in the Peruvian navy, claimed to have met General San Martín on September 5, 1839, when he showed him his work relating to the memoirs of his travels to Paris, and then gave San Martín the letter to validate it, according to Lafond de Lurcy. According to José Pacífico Otero and Enrique de Gandía [14], other historians dispute this claim and contend that one of Bolivar's assistants actually delivered the letter.

Again, the validity of the Lafond letter is defended by the Peruvian historian Germán Leguía suggesting that Bolívar himself was lost to him, as did the Argentine historian Beatriz Bragoni who indicated its validity and even justified the exchange of letters between Lafond and General San Martín [14].

The Spanish version of the letter Lafond is attributed to Juan Bautista Alberdí (1810-1884) in his biography of General San Martín [15]. In fact, on May 28, 1844, and subsequently on June 5, the letter was referred to in *El Peruano* after the work entitled "*Bolívar y San Martín*," where the letter would have been validated by General San Martín himself. However, the Bolivarian historian Vicente Lecuna as well as Cristóbal Mendoza asserted, without any doubt, that this letter is false [14], as were others that were in a bad book, with the sole purpose of distorting the image of Simón Bolívar by cataloging him as ambitious. To argue that, they indicated that the letter speaks, for example, of nineteen thousand Spanish veterans in Peru and it is false because that number was reached in 1824. Likewise, this letter is false because it refers to a request for assistance from San Martín to Bolívar, who, knowing the Libertador's trajectory, would have responded to the same.

Due to this, statistical methods based on stylometric studies will be used to confirm the veracity of this letter in the wake of the controversy caused by Lafond's letter.

Methodology

The letters of Simón Bolívar were obtained from the Center for Digital Humanities Virtual Library Miguel de Cervantes, which was created in 2000 to be a reference center for the relevant works of the Spanish language, available freely and free of charge from the University of Alicante (details at www.cervantesvirtual.com). From it, we obtained 16 letters from Simón Bolívar as detailed in Table 1.

Table 1. The letters used in this study indicate who wrote the letter, the date and place from where it was written and to whom it is addressed.

Abbreviation	Who wrote the letter	Date	Place	To whom the letter is addressed
B942	Bolívar	1825, Sept 25	Oruro	Andrés de Santa Cruz
B943	Bolívar	1825, Sept 25	Oruro	F. de P. Santander
B977	Bolívar	1825, Oct 27	Potosí	Bartolomé Salom
B978	Bolívar	1825, Oct 27	Potosí	F. de P. Santander
B362	Bolívar	1830, May 26	Turbaco	General Sucre
B363	Bolívar	1830, May 26	Turbaco	Juan de Dios Amador
B364	Bolívar	1830, May 31	Turbaco	Juan de Dios Amador
B365	Bolívar	1830, Jun 17	Turbaco	Pedro Medrano
B366	Bolívar	1830, Jul 2	Cartagena	Mariana Carcelén de Sucre
B367	Bolívar	1830, Jul 31	Cartagena	Manuela Garaycoa de Calderón
B369	Bolívar	1830, Sep 1	Cartagena	Robert Wilson
B370	Bolívar	1830, Sep 2	Cartagena	Gabriel Camacho.
B371	Bolívar	1830, Oct 17	Soledad	Joaquín de Mier
B32	Bolívar	1807, Sep 14	Yare	Pedro Machado
B23	Bolívar	1804, Jan 29	Cádiz	Jph. Manuel Jaén
B31	Bolívar	1807, Sep 2	Yare	Pedro Machado
SM10	San Martín	1826, Dec 4	Bruselas	Guillermo Miller
SM11	San Martín	1837, May 13	Bruselas	Guillermo Miller
SM9	San Martín	1837, Nov 28	Bruselas	Guillermo Miller
SM8	San Martín	1837, Dec 3	París	Bernardo O'Higgins
SM7	San Martín	1837, Mar 26	París	Bernardo O'Higgins
SM6	San Martín	1832, Dic 22	París	Bernardo O'Higgins
SM5	San Martín	1832, Dic 22	París	Bernardo O'Higgins
SM4	San Martín	1830, Feb 18	Bruselas	Bernardo O'Higgins
SM3	San Martín	1839, Apr 5	Montevideo	Bernardo O'Higgins
SM2	San Martín	1837, Oct 20	Bruselas	Bernardo O'Higgins
SM1	San Martín	1823, Feb 20	Mendoza	Bernardo O'Higgins
Lafond	¿Bolívar?	1822, Jun 22	Quito	¿General San Martin?

This table indicates the date and place from which it was written and to whom it is addressed. The selection criteria for the letters were random so as not to mislead the information. However, the transcription of the above-mentioned letters with the originals was not verified, suggesting that they are genuine.

Furthermore, eleven letters attributed to General San Martín (see details in Table 1) were included to verify that the methodology developed in the paper was able to discern between the two authors. The Lafond letter was obtained from documents that the Congress of the Republic of Peru brought to the

public through its Office of Citizen Participation, available at www.congreso.gob.pe/participacion/ museo/congreso/mensajes/

carta_jose_san_martin_a_simon_bolivar_reclamando_anexion_guayaquil_colombia_3_marzo_1822.

The next step was to create the linguistic corpus using the words from the various letters mentioned above. This allowed us to determine the frequency of the words used by each author because each author's usage of adjectives, substantives, and verbs is consistent across all of their writings, which is how we know who wrote a given document. Python programming was used to conduct these studies.

Tokenizing is the stylometric term for the division of texts into groups of words. It was simple to determine the frequency of appearance of words in texts once the words had been separated and their lengths were known [16]. The current analysis is based on function words, or the distribution ratio of prepositions—a factor that is typically ignored when a document is forged. Recall that some of the propositions include, for instance, the words a, against, etc. Later, the delta function was established in accordance with Burrows' definition from 2002 [17], which is widely used as a linguistic measure that separates authorship from texts. Although it was previously thought to be just a quick way to establish who wrote a text, Burrows later confirmed that it is an excellent method for determining the authenticity of works [18].

The Delta function simply determines the frequency variation of the most common words in a text using a z-score [19] function. It should be borne in mind that z-scores are the difference between the relative frequency minus the average of the words in the corpus, divided by the standard deviation of such data [details in 19].

From these results, a matrix is constructed where the different frequencies obtained on each letter are compared, calculating the distance to Manhattan [20]. Such a matrix is usually visualized with the help of a dendrogram, that is, a hierarchical representation of the tree type, where the results are grouped into clusters, where texts that are related to each other, i.e., of the same author, appear in the same nodes of the branches.

Results

Figure 1 displays the results of the frequency of the words derived from Simón Bolívar's letters (shown as columns in this figure), while the results derived from Lafond's letter are shown as a black line. This graph demonstrates that there is no proportion of functional words, among which are highlighted "la," "a," "no," "los," "me," "lo," "este," "le," and "esta," indicating that the Lafond letter does not reproduce the style of the Liberator's Bolívar.

Figure 2 shows the result represented as a dendrogram (Fig. 2). This figure shows that the letters are forming clusters grouped by the same author; that is, it is presented how only the letters of Simón Bolívar or of General San Martín are grouped among themselves, noting that both authors have their own style when writing.

Moreover, the Lafond letter is not included in any cluster formed by other letters from any of the liberators of Hispano-American; that is, the letter does not reproduce the linguistic footprint of the Libertador Bolívar, much less the style of San Martín. Thirdly, the style of writing of San Martín has little variation, unlike the richness of style present in the letters of Bolívar, because they are grouped into different branches.
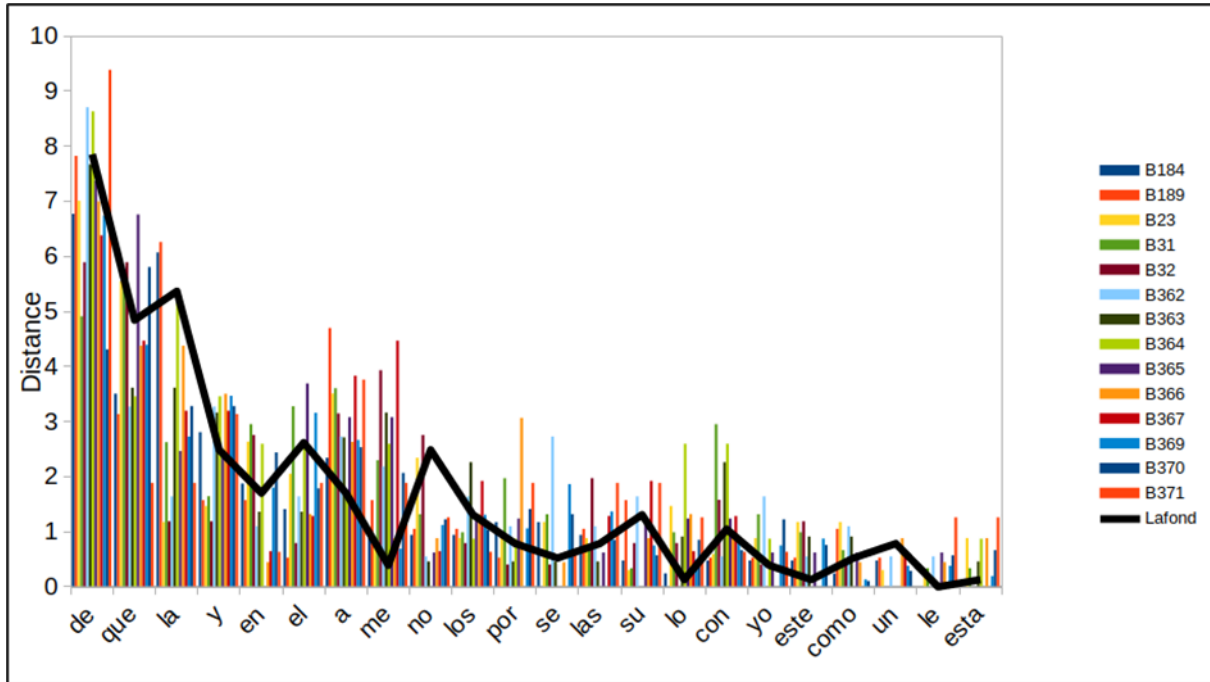
Figure 1. It shows the distances determined by the word frequencies of Simón Bolívar's letters (all of which begin with the letter B for easy identification), as well as the distance determined by the "evil letter" of Lafond (the result is shown as a black line).
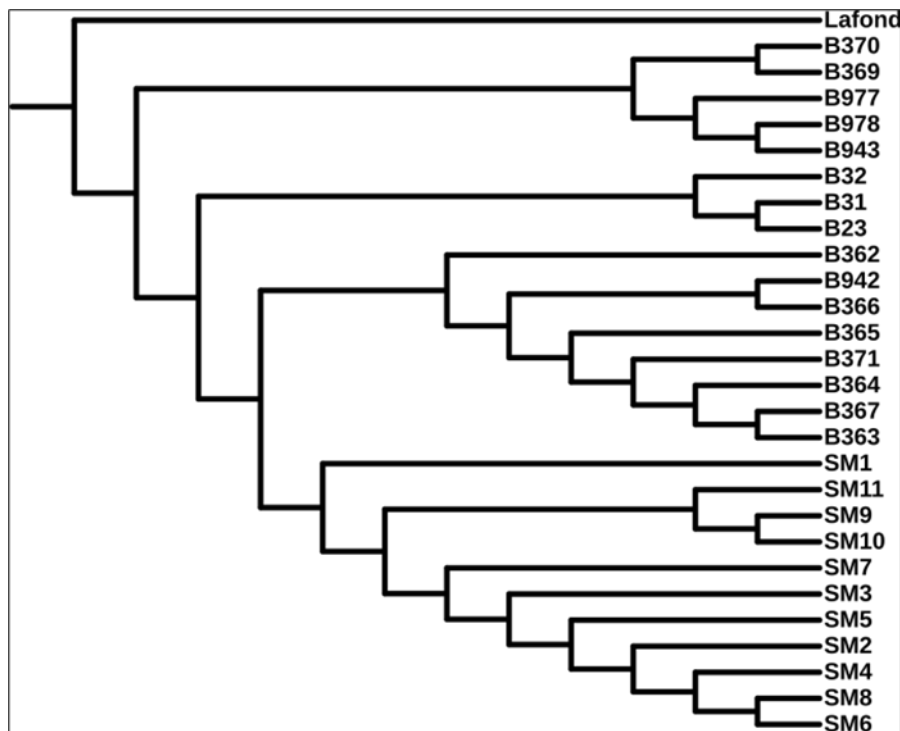

Figure 2. It shows the rectangular dendrogram that was produced by stylometric analysis of the letters written by Simón Bolívar, San Martin, and Lafond. San Martín's letter begin with SM, while Bolívar's letter begin with the B.

Conclusion

The goal of this paper is to find out the validity of Lafond's letter through stylometric analysis and to determine whether Simón Bolívar was really disposing of the future of the Republic of Ecuador with General San Martín. After analyzing the Delta function, it was found that Simón Bolívar never wrote the aforementioned letter to General San Martín. Therefore, it is confirmed that Lafond's letter is false and only sought to damage the reputation of Libertador Bolívar.

References

1. Redacción RPP (2016, November 22) "Carta que firmó Simón Bolívar en las guerras de Independencia fue subastada". Available at https://rpp.pe/mundo/actualidad/carta-que-firmo-simon-bolivar-en-las-guerras-de-independencia-fue-subastada-noticia-1011778 (Retrieved 2023, August 1)

2. Villagran, L (2020, October 2) "Allegedly stolen from Mexico, letters from Spanish conquest go for thousands at auction" Available at https://www.elpasotimes.com/story/news/2020/10/02/hernan-cortes-letters-stolen-mexico-sell-new-york-inah-search/3520423001/ (Retrieved 2023, August 1)

3. Feldman, E (2022, August 22) "Historian Discovers a Prized Galileo Manuscript Was Forged" Available at https://www.smithsonianmag.com/smart-news/historian-discovers-a-prized-galileo-manuscript-was-forged-180980615/ (Retrieved 2023, August 1)

4. Bensalem, I, Rosso Paolo and 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.

5. Stamatatos E (2009). "A survey of modern authorship attribution methods". JASIST. 60 (3): 538–556.

6. Fuller S, O'Sullivan J (2017). "Structure over Style: Collaborative Authorship and the Revival of Literary Capitalism". Digital Humanities Quarterly. 11 (1).

7. Lutoslawski W. (1898). "Principes de stylométrie appliqués à la chronologie des œuvres de Platon". Revue des Études Grecques. 11 (41): 61–81. doi:10.3406/reg.1898.5847.

8. 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/ (Retrieved 2023, August 1)

9. Schoenbaum S, Internal evidence and Elizabethan dramatic authorship; an essay in literary history and method, Editorial Northwestern University Press

10. Jockers ML, Witten DM, Criddle CS (2008) "Reassessing authorship of the Book of Mormon using delta and nearest shrunken centroid classification", Literaty and Lingustic Computing, 23(4):465-491.

11. Schaalje GB, Fields PJ, Roper M, Gregory L (2011) "Extended nearest shrunken centroid classification: a new method for open-set authorship attribution of texts of various sizes". Digital Scholarship in the Humanuties, 26(1):71-88

12. Colombres Mármol, Eduardo, San Martín y Bolívar en la Entrevista de Guayaquil a la luz de nuevos documentos definitivos, prólogo de Rómulo Carbia, Buenos Aires, Coni, 1940 [reseñado por José Torre Revello, Revista Historia de América, n° 10, 1940]

13. Bragoni, Beatriz (2016). "El intercambio epistolary entre San Martín y Lafond". Prismas, 20(1):47-

60

14. Perez Amuchástegui AJ (1962) "La carta de Lafond. Córdoba : Universidad Nacional de Córdoba, Instituto de Estudios Americanistas. Edicions Siglo XX, Buenos Aires.

15. Alberdí JM (1844) "Biografia del Jeneral San Martín" Editrial Imprenta de Ducessois , Quai Des Augustins.

16. Škorić M, Stanković R, Ikonić Nešić M, Byszuk J, Eder M. Parallel Stylometric Document Embeddings with Deep Learning Based Language Models in Literary Authorship Attribution. Mathematics. 2022; 10(5):838.

17. Burrows JF (2002) "Delta: A Measure of Stylistic Difference and a Guide to Likely Authorship". Literary and Linguistic Computing 17(3):267-287

18. Stefan E and others, Understanding and explaining Delta measures for authorship attribution, Digital Scholarship in the Humanities, Volume 32, Issue suppl_2, December 2017, Pages ii4–ii16

19. Patergianakis A, Limniotis K. Privacy Issues in Stylometric Methods. Cryptography. 2022; 6(2):17. https://doi.org/10.3390/cryptography6020017

20. Stanikunas D, Mandravickaite J and 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