Stylometric Comparison between the Quran and Hadith based on Successive Function Words: Could the Quran be written by the Prophet?

Halim Sayoud
EDT, Faculty of Electronics
USTHB University, Algiers
http://sayoud.net
halim.sayoud@gmail.com

Abstract

Usually, when one writes a text document, several function words are put within that text to make a logical link between words and to give more explanation to the major idea of the paragraph. However, it is not very common to see two Successive Function Words (SFW) put together in a sentence, even if it is not really incorrect. For instance, instead of writing "I left my room after our discussion", one can say in Arabic: "I left my room from-after our discussion", which is correct in this language. In this work, those successive function words are investigated in the Quran, Hadith and five other Arabic religious books, in order to see whether the corresponding styles are similar (i.e. Stylometric comparison between the different books). Results of this investigation show that the use of SFW is very discriminative between the Quran and Hadith, which implies that the Quran and Hadith come from two different Authors. Moreover, the discrimination is even noticeable between the Quran and several other books investigated in this work. This result suggests that the Quran Author style is beyond the traditional human religious styles that are usually employed in the Arabic literature (ancient or contemporary). Hence, the only explanation we have is: The Quran should probably come from God (Allah).

Keywords: Quran style, Hadith style, Arabic Language, Text Mining, Stylometry, Successive Function Words, Quran, Successive Function Words (SFW).

1. Introduction

The Muslim community believes that the holy Quran is the divine book of Allah (i.e. God), which was only sent down to the Prophet during the revelation. However, some doubtful persons claim that it could be only the invention of the Prophet. To give a fair response to this question: Who is the real Author of the Quran? It is important to handle the problem with care and to use only scientific approaches to avoid any kind of subjectivity. Fortunately, we also possess a copy of the Hadith, which represents the original statements and speech of the Prophet.

This precious book makes the Stylometric comparison between the Quran and the Prophet style possible. Consequently, the comparison between the Quran and Hadith could bring a key information by giving a fair decision on the acceptance or the rejection of that claim, hopefully.

This challenge represents a great motivation to try discovering the truth, which was difficult to show centuries before. Hence, we have decided to continue our previous series of experiments on the discrimination between the Quran style and the Hadith style (Sayoud, 2012-a) (Sayoud, 2012-b) (Sayoud, 2014) (Sayoud, 2015-a) (Sayoud, 2015-b) (Ouamour, 2016) (Sayoud, 2018-a) (Sayoud, 2018-b), with a new set of features that is based on successive function words. The details of this investigation are given in the next sections.

2. Used function words

Function words are very used in Arabic (El-Zohairy, 2008) (García-Barrero, 2012) (Almujaiw, 2017), however successive pairs of function words cannot be translated directly to English, because of the complex translation process and the different morphologies in Arabic and in English. So, for instance, the succession of the following function words: من بعد (rough translation: From After) cannot be translated into English by keeping both prepositions, which may lead to an ambiguous meaning (i.e. From after), while the correct and most suitable translation is simply: After.

That is, 10 different types of Successive Function Words (SFW) have been selected from the Quran, Hadith and five other religious books by computing their corresponding frequencies, in a purpose of authorship discrimination (Sayoud 2012-a) (Tambouratzis, 2004). Those successive words are represented in table 1 and their corresponding frequencies are given in table 2 and 3.

Table 1: The used Arabic successive words (SFW)

prep +prep	Translation
و إِذ	And Then-When-Since
و مما	And From-what
و لقد	And Already
أَوَ مَنْ	And-do-you-think-that Who
من بعد	After / From After
ذَلِكَ لِمَنْ	That (is) for (whom/who)
ذَلِكَ بِأَنَّ	That (is) because
وَ لَئِنْ	And If
مَا لَا	What (is) Not
مِنْ قَبْل	Before / From Before

In Table 1, one can see the 10 SFWs that were selected and used as features in this investigation. For instance, if we consider the fifth one (i.e. من بعد), it can be translated into: "From After", which is equivalent to "After" in Arabic. Similarly, if we consider the last one (i.e. مِنْ قَبْل), it can be translated into: "From Before", which is equivalent to "Before" in Arabic. So, in Arabic using "Before" or "From Before" is equivalent, and the authors can choose any one of them.

3. Analysis of the books based on the SFW
3.1 Comparison between the Quran and Hadith

The main goal of this investigation is to see whether the two religious books Quran (Nasr, 2016) (Ibrahim, 1997) and Hadith (Islahi, 1989) come from the same author or not, which could shed some light on the authorship origin of the holy Quran. The occurrence frequencies of the quoted pairs of function words, between the Quran and Hadith, are given in Table 2, where one can see that for the Quran the frequencies values are relatively high, while for the Hadith, the frequencies are very low and sometimes null. This difference can be better noticed in Figure 1.
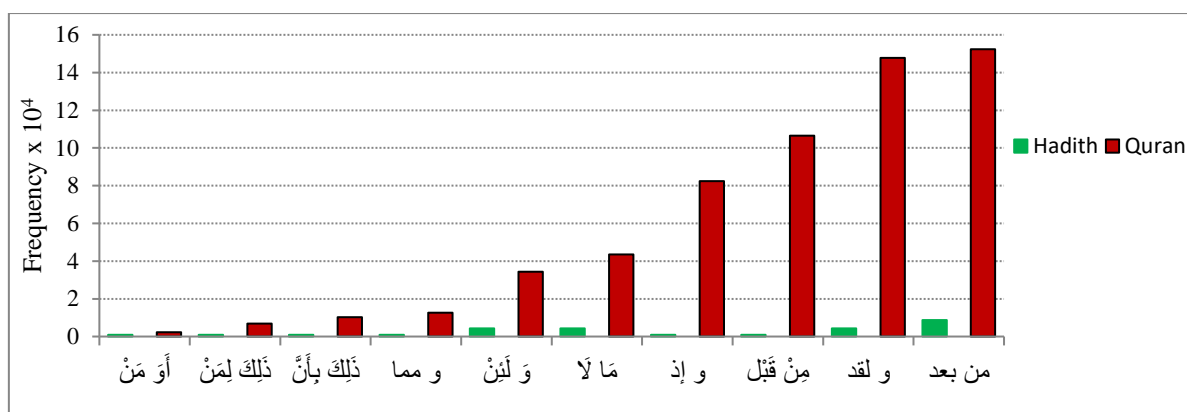
Figure 1: Graphical representation of the frequencies of the Arabic successive words (SFW)

In Table 2 and Figure 1, one can notice that the level of SFW is much higher in the Quran than in the Hadith. More particularly, for the following SFW: من بعد - و لَقد - مِنْ قَبْل - و إذ, where one notes a huge difference in their values between the two books. Hence, we notice that the use of SFW is very different between the Quran and Hadith, showing the simple and easy style of the Hadith while for the Quran the style appears to be more sophisticated and richer by using much more successive function words.

3.2 Comparison of seven different books: Quran, Hadith and five other Arabic books

In this experiment, we try to compare the SFW frequencies of seven different books belonging to different authors. The occurrence frequencies of the quoted pairs of function words, between the different books: Quran, Hadith and 5 other mixed Arabic books written by 5 religious scholars (i.e. Abdalkafy, Alghazali, Alqaradawi, Alqarni and Amrokhaled) (Hadjadj, 2018), are given in Table 2. One can see that for the Quran, the frequencies values are relatively high, while for all other books, the frequencies are very low and sometimes null. This difference is easily noticeable in Figure 2.

Table 2: Frequencies of the Arabic successive words (SFW) in the seven books.

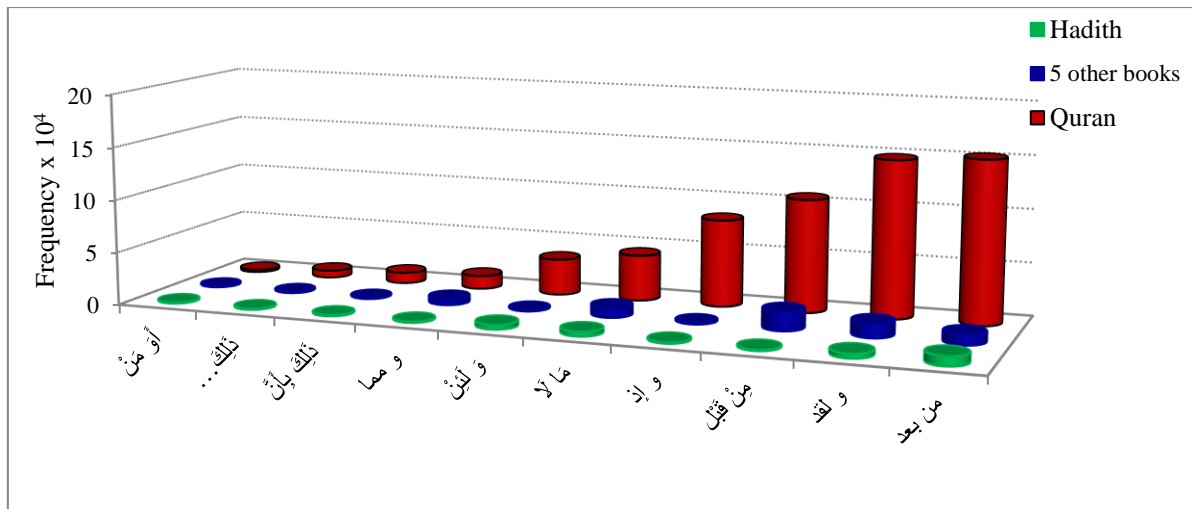| SFW	Frequency in the Quran $\times 10^{-4}$	Frequency in the Hadith $\times 10^{-4}$	Medium Frequency in the 5 other books $\times 10^{-4}$
و إذ	8.24	0.00	0.00
و مما	1.26	0.00	0.59
و لقد	14.77	0.43	1.32
أَوَ مَنْ	0.23	0.00	0.00
من بعد	15.23	0.87	0.92
ذَلِكَ لِمَنْ	0.69	0.00	0.00
ذَلِكَ بِأَنَّ	1.03	0.00	0.00
وَ لَئِنْ	3.43	0.43	0.00
مَا لَا	4.35	0.43	0.92
مِنْ قَبْل	10.65	0.00	1.76
Mean value	**5.99**	**0.22**	**0.55**

Figure 2: Graphical representation of the frequencies of the Arabic successive words (SFW).

In Table 2 and Figure 2, one can see a global comparison between the holy Quran, the Hadith and five other Arabic religious books. As previously, one can notice that the level of SFW is much higher in the Quran than in the other books. More particularly, for the following successive function words: من بعد - و لقد - مِنْ قَبْل - و إذ, where there is a great difference in the SFW values.

As an overall consequence, one can conclude that the Quran style uses much more frequently the combination of function words than all other books (i.e. the Hadith and the other religious books). In other words, the SFW frequency is clearly higher in the holy Quran than in all other books that have been analyzed in this investigation, which involves a particular and distinct style for the holy Quran.

4. Discussion and conclusion

By observing Figure 1, one can notice that the utilization of SFW in the Hadith is very rare: the mean frequency is only $0.22 \cdot 10^{-4}$, while in the case of the holy Quran, the frequency is much higher: mean frequency of about $6 \cdot 10^{-4}$, which is 27 times higher than the previous one. Similarly, by observing Figure 2, one can notice that the utilization of SFW in the six religious books (i.e. the Hadith and five other books) is very rare too. In fact, the mean frequency does not exceed $0.55 \cdot 10^{-4}$ for those books, while in the case of the holy Quran, the frequency is much higher: mean frequency of about $6 \cdot 10^{-4}$, which is 11 times higher than the mean frequency of the five religious books and 27 times higher than the frequency of the Hadith.

This noticeable difference in the use of SFW has two important interpretations:
- Firstly, the two writing styles of the Quran and Hadith appear different, with regards to the use of SFW.
- Secondly, the writing style of the Quran is even different from all the human religious books that have been investigated, with regards to the SFW.

In conclusion, these results show that the holy Quran's style is not only different from the Hadith style, and then different from the Prophet style, but also different from all the Arabic styles that have been analyzed in this work (*i.e. five Arabic religious books*). This discovery suggests that the Quran Author style is beyond the traditional human religious styles that are usually employed in the Arabic literature (*by ancient authors or contemporary authors*).

Now, concerning the key question: Could the Quran be written or invented by the Prophet?
We can state that, according to this investigation, and other works previously published in Sayoud H. (2012-a), the doubtful hypothesis claiming that the Quran could be a simple invention of the Prophet is scientifically impossible. See the following verse [10-38] of the holy Quran: ﴿ أَمْ يَقُولُونَ افْتَرَاهُ ۖ قُلْ فَأْتُوا بِسُورَةٍ مِّثْلِهِ وَادْعُوا مَنِ اسْتَطَعْتُم مِّن دُونِ اللَّهِ إِن كُنتُمْ صَادِقِينَ ﴾.

This verse can be translated into: "*And yet, they [who are bent on denying the truth] assert, "[Muhammad] has invented it!" Say [unto them]: "Produce, then, a surah of similar merit; and [to this end] call to your aid whomever you can, other than God, if what you say is true!*"
Indeed, and from a scientific point of view, the holy Quran could not have been written or invented by the Prophet (PBUH).

References

Almujaiw S. (2017). Grammatical construction of function words between old and modern written Arabic: A corpus-based analysis. Corpus Linguistics and Linguistic Theory 13(2) May 2017.

El-Zohairy N. (2008). A Dictionary of Function Words in Arabic. Liban Publisher, Lebanon, 2008.

García-Barrero D., Feria M. and Turell M.T. (2012). Using function words and punctuation marks in Arabic forensic authorship attribution. IAFL conference Proceedings, Porto 2012.

Hadjadj H. & Sayoud H., (2018). Authorship Attribution of Seven Arabic Religious Books-A Fusion Approach. Conference paper: 6ème Conférence Internationale en Automatique & Traitement de Signal (ATS-2018), Morrocco.

Ibrahim I.A. (1997). A brief illustrated guide to understanding Islam. Library of Congress, Catalog Card Number: 97-67654, Published by Darussalam, Publishers and Distributors, Houston, Texas, USA, 1997. Web version: http://www.islam-guide.com/contents-wide.htm.

Islahi A.A. (1989), 1989. Fundamentals of Hadith Interpretation – an English translation of "Mabadi Tadabbur-e-Hadith" by T. M. Hashmi. Lahore: Al-Mawrid. www.monthly-renaissance.com/DownloadContainer.aspx?id=71.

Nasr S. H. (2016), Encyclopædia Britannica Online. http://www.britannica.com/eb/article-68890/Quran, Last access in 2016.

Ouamour S., Khennouf S., Bourib S., Hadjadj H. and Sayoud H. (2016). Effect of the Text Size on Stylometry -Application on Arabic Religious Texts. International Conference on Computer Science Applied Mathematics and Applications, ICCSAMA'2016. pp 215-228. http://www.trinidata.at/iccsama2016/.

Sayoud H. (2012-a). Author discrimination between the Holy Quran and Prophet's statements. Literary and Linguistic Computing 2012, pp 427-444. Vol. 27, No. 4, 2012.

Sayoud H. (2012-b). Authorship Classification of two Old Arabic Religious Books Based on a Hierarchical Clustering. LRE-Rel: Language Resources and Evaluation for Religious Texts. 22 May, Lütfi Kirdar Convention & Exhibition Centre Istanbul, Turkey, pp 65-70.

Sayoud H. (2014). Automatic Authorship Classification of Two Ancient Books: Quran and Hadith. The 11th ACS/IEEE International Conference on Computer Systems and Applications (AICCSA'2014), November 10-13, 2014. pp 666-671. http://cse.qu.edu.qa/aiccsa2014/.

Sayoud H. (2015-a). A Visual Analytics based Investigation on the Authorship of the Holy Quran. . The 6th International Conference on Information Visualization Theory and Applications (IVAPP'2015), March 11-14, 2015, pp 177-181. http://www.ivapp.visigrapp.org/?y=2015.

Sayoud H. (2015-b). Segmental Analysis Based Authorship Discrimination between the Holy Quran and Prophet's Statements. Digital Studies Journal. 2015: 2014-2015, Open Issue. ISSN: 1918-3666. Website: https://www.digitalstudies.org/articles/10.16995/dscn.44/

Sayoud H. (2018-a). Visual Analytics Based Authorship Discrimination Using Gaussian Mixture Models and Self Organising Maps: Application on Quran and Hadith. International Conference on Industrial, Engineering & Other Applications of Applied Intelligent Systems, pp 158-164, June 25-28, 2018 Montreal, Proc. by Springer. https://www.springer.com/us/book/9783319920573.

Sayoud H. and Hadjadj H. (2018-b). Fusion Based Authorship Attribution -Application of Comparison between the Quran and Hadith. 6th International Conference on Arabic Language Processing, 2017, pp 191-200. Proc. by Springer. http://www.springer.com/us/book/9783319734996.

Tambouratzis G., Markantonatou S., Hairetakis N., Vassiliou M., Carayannis G. & Tambouratzis D. (2004). Discriminating the Registers and Styles in the Modern Greek Language-Part 1: Diglossia in Stylistic Analysis. Literary and Linguistic Computing, 19(2), 197-220.

Biodata

Pr Halim Sayoud is Full Professor at the USTHB University. He is the head of the EDT research team. Pr Halim Sayoud published about 100 scientific research papers in conferences proceeding or international journals and is also the Editor-in-Chief of the HDSKD international journal. He is particularly interested in the following research fields: Speaker Recognition, NLP, Stylometry, Text categorization, Ancient documents analysis and Artificial Intelligence. Official website: http://sayoud.net
Personal website: http://scholarpage.org/sayoud.html