

Stylometric Analysis of the Quran and Hadith based on the FWP and BWP

H. Sayoud

EDT, USTHB University

www.sayoud.net

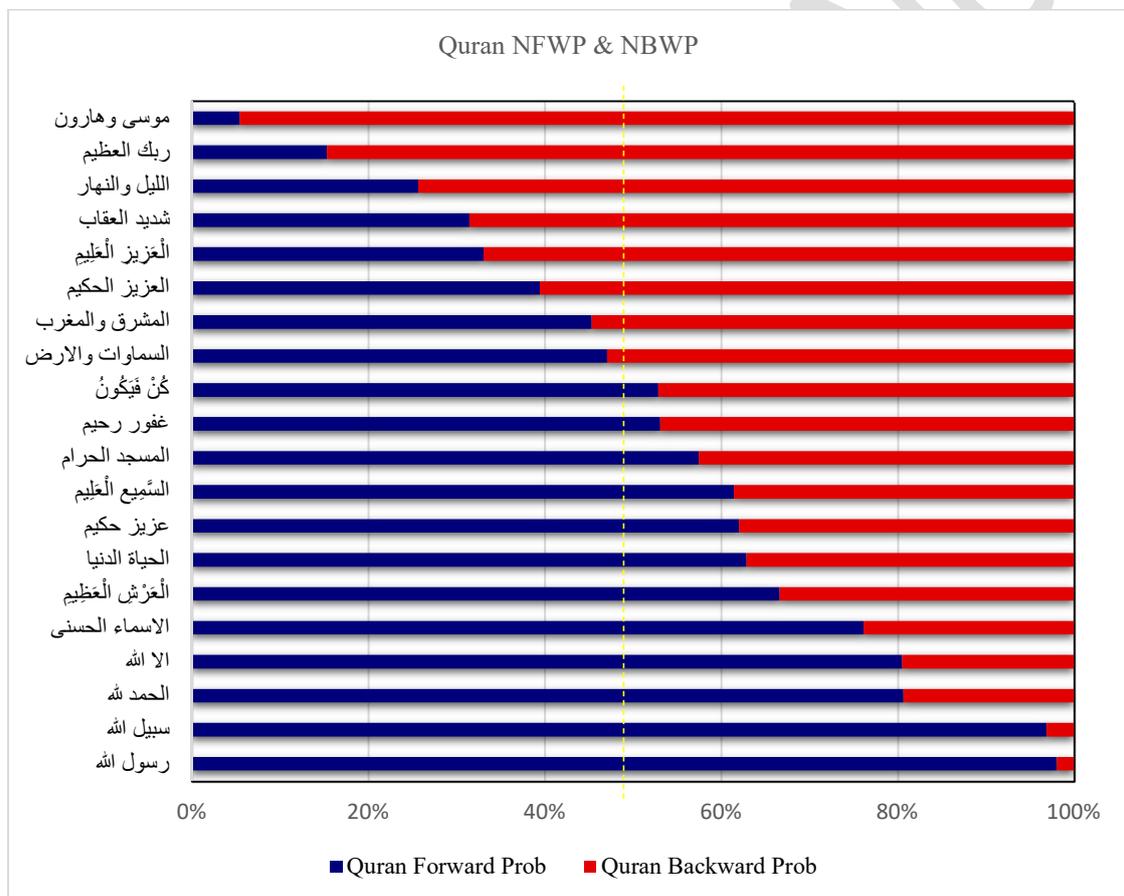


Fig 1: Graphical representation of the normalized probabilities NFWP and NBWP in the Quran

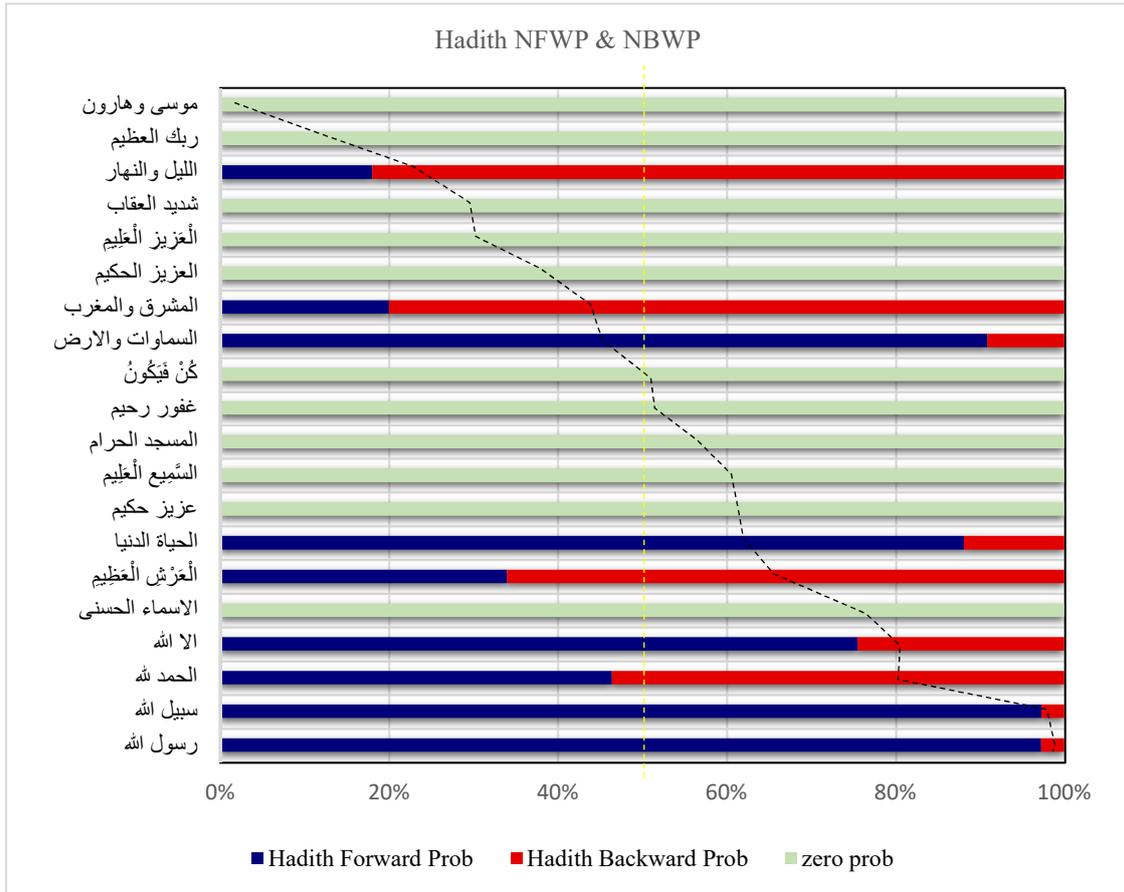


Fig 2: Graphical representation of the normalized probabilities NFWP and NBWP in the Hadith. For the case of zero probabilities, we plotted a light green bar chart for representing such cases. The black dashed line represents the shape of the intersection line between the NFWP and NBWP bars in the Quran, for a comparison purpose.

According to figure 1, there are 4 bigrams from the Quran that have balanced normalized probabilities (i.e. $NFWP \cong NBWP$), namely:

غفور رحيم, كُنْ فَيَكُونُ, السماوات والأرض, المشرق والمغرب

While for the Hadith (see figure 2), only 1 bigram has a balanced normalized probability, namely:

الحمد لله

As for the intersection line, which represents a borderline curve between the NFWP and NBWP bars (dashed black curve) in figure 2, we notice that the Quran's borderline is completely different from the Hadith's one, except for the 1st and 2nd bigrams (i.e. رسول الله, سبيل الله) and maybe also the 4th and 18th bigrams. Hence, among 20 analyzed bigrams, only 4 over 20 bigrams have almost similar NFWP and NBWP (i.e. 20% of the whole bigrams), while 16 among 20 bigrams, namely 80% of the whole bigrams, present normalized probabilities that are completely different between the 2 books.

The assessment based on visual-analytics has been made thanks to the normalized parameters NFWP and NBWP, where we, visually noticed that the bar charts displayed for the 2 books are completely different (i.e. visual analytical assessment), and where 80% of the word bigrams presented different normalized probabilities.

Hence, it appears that the two investigated books possess two different Author's styles, which confirms once again the previous results stating that they should probably belong to two different Authors.