Morphological Analysis of the Glorious Qur'an: A Comparative Survey of Three Corpora

Yasser Muhammad Naguib Sabtan

Department of Languages and Translation

Dhofar University, Oman

&

Faculty of Languages and Translation

Al-Azhar University, Egypt

Abstract

Some attempts have been made in the academic community to carry out an automatic morphological analysis of the Qur'anic text. Among the well-known endeavors in this regard is the morphological annotation of the Quranic Arabic Corpus (QAC) which was carried out in Leeds University, UK. In addition, researchers in the University of Haifa had previously implemented a computational system for the morphological analysis of the Qur'an. More recently, a new Quranic corpus has been built in Mohammed I University in Morocco. To the best of our knowledge, these are the only three studies to produce a morphologically analyzed part-of-speech tagged Qur'an encoded as a structured linguistic database. This paper surveys the morphological analysis in the above-mentioned annotation projects and compares between them to test the quality of their analysis using five criteria related to display of the text in the corpus, word segmentation, morphological disambiguation, part of speech (POS) tag set and manual verification. The paper concludes that the QAC of Leeds and the Quranic corpus of Morocco surpass the Quranic corpus of Haifa with regard to most of these criteria. Furthermore, some additional POS tags for derivative nouns are suggested in a step to reach a more fine-grained tag set that could be proposed for POS tagging of Qur'anic Arabic.

Keywords: Arabic morphological analysis, Arabic POS tagging, corpus annotation, corpus linguistics, the Glorious Qur'an

Cite as: Sabtan, Y. M. N. (2017). Morphological Analysis of the Glorious Qur'an: A Comparative Survey of Three Corpora. *Arab World English Journal*, 8 (4). DOI: <https://dx.doi.org/10.24093/awej/vol8no4.7>