

Abstract

Iman M Mahfouz

Variation in Arabic transliteration of English words on Twitter: A corpus-based approach

In recent years, Arab Internet users in general and Egyptians in particular have started transliterating English words in Arabic characters in their interactions on social media. It has been noticed that some of the transliterated items display significant spelling variation. Using a corpus of 666,000 tweets written by Arab users collected automatically in June 2017, the present study examines variation in spelling for a number of 168 transliterated items. Special emphasis is given to the mapping of short vowels, the representation of sounds not existing in Arabic, initial consonant clusters, as well as compound nouns. Drawing on research on loanword adaptation and using corpus-based techniques, the research attempts to account systematically for these variations with the aim of determining what constitutes a preferred variant. The paper thus demonstrates the potential of corpus-based methods in the study of new trends in social media from the perspective of contact linguistics. The study also sheds light on the difficulties involved in the automatic retrieval of Arabic tweets in general.