Pronunciation Variation Metric for Four Dialects of Arabic

Mahmoud Abunasser

Abbas Benmamoun

Mark Hasegawa-Johnson

University of Illinois at Urbana Champaign

Pronunciation variation is one important and frequently used measure to quantitatively determine linguistic variation, also called linguistic distance. It is a prominent topic in the growing field of dialectometry which is concerned with quantifying linguistic similarities and differences and, often, link it to geographical distances between the areas where the relevant languages are spoken (Nerbornne and Kretzschmar, 2003). In this paper we will report on research that introduces new computational techniques to quantify the pronunciation variation between four Arabic dialects and varieties.

The varieties under consideration are local dialects of Egyptian Arabic (EA), Gulf Arabic (GA), Levantine Arabic (LA), and Moroccan Arabic (MA). The technique used to estimate the pronunciation variation is based on the word similarity metric between words with the same meaning.

Word similarity can be diagnosed at different levels of analysis. At the most Abstract level, it is defined based on whether the words are cognates. At an intermediate level, word similarity can be judged based on the minimum number of insertions, deletions, and substitutions required to transform the IPA transcription of one word to the other – This can be calculated based on a computational technique called Levenshtein distance or the minimum edit distance algorithm. In this research, we are investigating a deeper level of representation where word similarity is calculated based on articulatory gestures and phonetic features of the sequence of phones of each word. This is approached by developing a mathematical representation of sound that captures variation based on phonetic features and articulatory gestures. This presentation reports results from all aforementioned levels of analysis.

Preliminary results show that the Levantine variety is close to both Egyptian and Gulf while Moroccan turned out to be more distant from the former three varieties. The distance between Egyptian and Gulf was in the middle.