Building an English-Iraqi Arabic Machine Translation System for Spoken Utterances with Limited Resources

Jason Riesa (1), Behrang Mohit (2), Kevin Knight (3), Daniel Marcu (3)

(1) Johns Hopkins University, USA; (2) University of Pittsburgh, USA; (3) University of Southern California, USA

This paper presents an English-Iraqi Arabic speech-to-speech statistical machine translation system using limited resources. In it, we explore the constraints involved, how we endeavored to mitigate such problems as a non-standard orthography and a highly inflected grammar, and discuss leveraging existing plentiful resources for Modern Standard Arabic to assist in this task. These combined techniques yield a reduction in unknown words at translation time by over 40% and a +3.65 increase in BLEU score over a previous state-of-the-art system using the same parallel training corpus of spoken utterances.

Full Paper