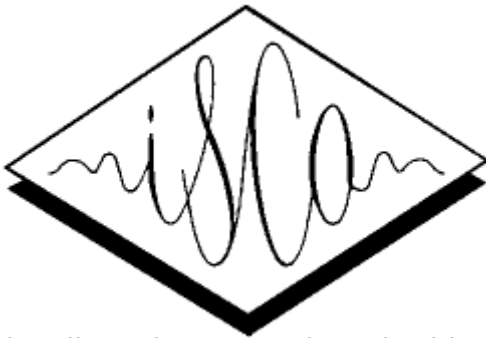**INTERSPEECH 2007**
8th Annual Conference of the International Speech Communication Association

Antwerp, Belgium
August 27-31, 2007

Context Dependent Word Modeling for Statistical Machine Translation Using Part-of-Speech Tags

Ruhi Sarikaya, Yonggang Deng, Yuqing Gao

IBM T.J. Watson Research Center, USA

Word based translation models in particular and phrase based translation models in general assume that a word in any context is equivalent to the same word in any other context. Yet, this is not always true. The words in a sentence are not generated independently. The usage of each word is strongly affected by its immediate neighboring words. The state-of-the-art machine translation (MT) methods use words and phrases as basic modeling units. This paper introduces Context Dependent Words (CDWs) as the new basic translation units. The context classes are defined using Part-of-Speech (POS) tags. Experimental results using CDW based language models demonstrate encouraging improvements in the translation quality for the translation of dialectal Arabic to English. Analysis of the results reveals that improvements are mainly in fluency.

Full Paper

Bibliographic reference. Sarikaya, Ruhi / Deng, Yonggang / Gao, Yuqing (2007): "Context dependent word modeling for statistical machine translation using part-of-speech tags", In *INTERSPEECH-2007*, 2821-2824.