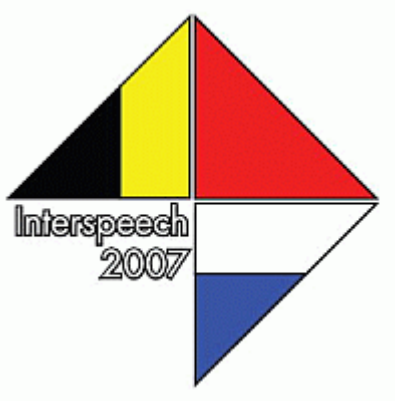
INTERSPEECH 2007
8th Annual Conference of the International Speech Communication Association



Antwerp, Belgium
August 27-31, 2007

Improving Speech Translation with Automatic Boundary Prediction

Evgeny Matusov (1), Dustin Hillard (2), Mathew Magimai-Doss (3), Dilek Hakkani-T□r (3), Mari Ostendorf (2), Hermann Ney (1)

(1) RWTH Aachen University, Germany
(2) University of Washington, USA
(3) ICSI, USA

This paper investigates the influence of automatic sentence boundary and sub-sentence punctuation prediction on machine translation (MT) of automatically recognized speech. We use prosodic and lexical cues to determine sentence boundaries, and successfully combine two complementary approaches to sentence boundary prediction. We also introduce a new feature for segmentation prediction that directly considers the assumptions of the phrase translation model. In addition, we show how automatically predicted commas can be used to constrain reordering in MT search. We evaluate the presented methods using a state-of-the-art phrase-based statistical MT system on two large vocabulary tasks. We find that careful optimization of the segmentation parameters directly for translation quality improves the translation results in comparison to independent optimization for segmentation quality of the predicted source language sentence boundaries.

Full Paper

Bibliographic reference. Matusov, Evgeny / Hillard, Dustin / Magimai-Doss, Mathew / Hakkani-T□r, Dilek / Ostendorf, Mari / Ney, Hermann (2007): "Improving speech translation with automatic boundary prediction", In *INTERSPEECH-2007*, 2449-2452.