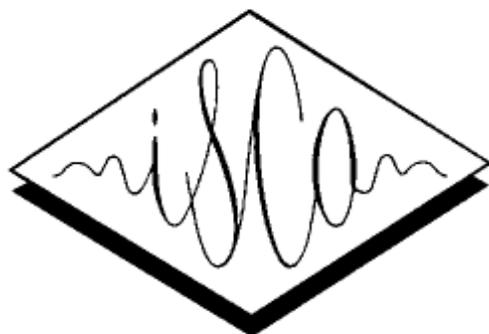


ISCA Archive

<http://www.isca-speech.org/archive>

**INTER_SPEECH
2008
9th Annual
Conference of
the
International
Speech
Communication
Association**

**Brisbane, Australia
September 22-26,
2008**



**Strategies for Building a Farsi-English SMT System from
Limited Resources**

Andreas Kathol, Jing Zheng

SRI International, USA

One of the recent tasks for machine translation research has been development of translation capabilities in a time frame as short as 100 days. Such a task requires developers to consider what can be done with relatively small amounts of data in a small time frame. This inherently limits the type and complexity of the effort to be devoted to this task. In this paper we will focus on the kinds of improvements for a Farsi-to-English translation system achieved by means of algorithmic changes, adding raw, domain-unspecific resources, and unsupervised morphological segmentation. The cumulative effect of these measures has been an improvement in BLEU scores of about 25% relative on an internal test set.

[Full Paper](#)

Bibliographic reference. Kathol, Andreas / Zheng, Jing (2008): "Strategies for building a Farsi-English SMT system from limited resources", In *INTER_SPEECH-2008*, 2731-2734.