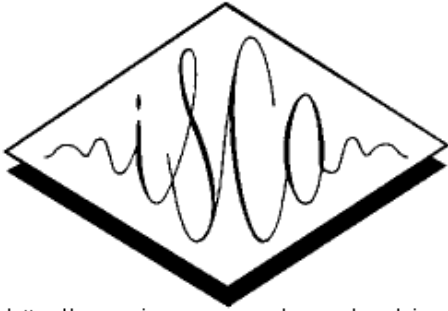**Sixth European Conference on Speech Communication and Technology (EUROSPEECH'99)**

**Budapest, Hungary
September 5-9, 1999**

Automatically Deriving Categories for Translation

Sergio Barrachina, Juan Miguel Vilar

Unidad Predepartamental de Informatica, Universidad Jaume I, Castell   n (Spain)

An adequate approach to speech translation for small to medium sized tasks is the use of subsequential trans-ducers - a finite state model - as language model for a speech recognizer. These transducers can be automatically trained from sample corpora. The use of manually defined categories improves the training of the subsequential transducers when the available data are scarce. These categories depend on the source and target languages we want to translate. We introduce an automatic approach to derive categories that can be used in training subsequential transducers. This approach extends monolingual word clustering methods to the bilingual case using alignments obtained from statistical models. Experimental results indicate that the models trained with these categories have lower translation errors.

Full Paper (PDF)   Gnu-Zipped Postscript

Bibliographic reference. Barrachina, Sergio / Vilar, Juan Miguel (1999): "Automatically deriving categories for translation", In *EUROSPEECH'99*, 2415-2418.