[CICLING 2010: 11th International Conference on Intelligent Text Processing and Computational Linguistics, March 21-27, 2010, Iaşi, Romania]

Abstract

Raquel Justo, Alicia Pérez, M. Inés Torres and Francisco Casacuberta. Hierarchical finite-state models for speech translation using categorization of phrases

In this work a hierarchical translation model is formally defined and integrated in a speech translation system. As it is well known, the relations between two languages are better arranged in terms of phrases than in terms of running words. Nevertheless phrase-based models may suffer from data sparsity at training time. The aim of this work is to improve current speech translation systems by integrating categorization within the translation model. The categories are sets of phrases, being the latter either linguistically or statistically motivated. Both category and translation and acoustic models are finite-state models. In what temporal cost concerns, finite-state models count on efficient algorithms. Regarding the spatial cost, all the models where integrated on-the-fly at decoding time, allowing an efficient use of the memory.