



TURKISH DICTATION SYSTEM FOR BROADCAST NEWS APPLICATIONS (ThuAmPO1)

★ Author(s) :

Ebru Arisoy

(Bogazici University, Turkey)

Levent M. Arslan

(Bogazici University, Turkey)

★ Abstract :

We have designed a Turkish dictation system for Broadcast news applications. Turkish is an agglutinative language with free word order. These characteristics of the language result in the vocabulary explosion, large number of out-of-vocabulary (OOV) words and the complexity of the N-gram language models in speech recognition when words are used as recognition units. Therefore, we proposed new recognition units. We parsed some of the words to smaller recognition units like stems, endings and morphemes, and introduced these smaller units and the unparsed words to the speech recognizer as lexicon entries. This way, we were able to overcome to the problem of large number of OOV words with a moderate vocabulary size and get better estimates for the N-gram language models. However, best recognition result was obtained using the word-based language model.