We discuss how to measure the reliability of recognized utterances based on a confidence measure, and applied it to a dialog speech translation system. In this study, we employ generalized word posterior probability (GWPP), a confidence measure for verifying recognized words, and expand it to measure the reliability of recognized utterances. We confirmed the performance improvement by applying the rejection technique to a dialog speech translation system from Japanese to English. We conducted two kinds of performance evaluation. One is a ranking evaluation of translation output by human evaluators. The other is to measure the machine output against human results by a paired-comparison method. Both of them yield significant improvements.

Full Paper