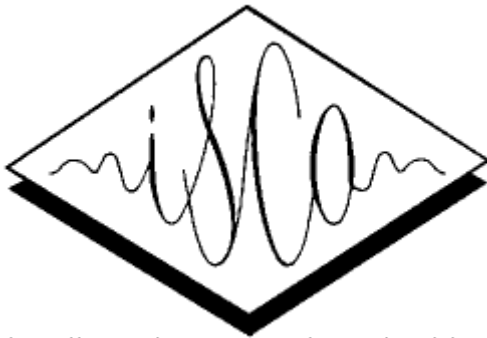**EUROSPEECH 2003 - INTERSPEECH 2003**

8th European Conference on Speech Communication and Technology

**Geneva, Switzerland
September 1-4, 2003**

Collecting Machine-Translation-Aided Bilingual Dialogues for Corpus-Based Speech Translation

Toshiyuki Takezawa, Genichiro Kikui

ATR-SLT, Japan

A huge bilingual corpus of English and Japanese is being built at ATR Spoken Language Translation Research Laboratories in order to enhance speech translation technology, so that people can use a portable translation system for traveling abroad, dining and shopping, as well as hotel situations. As a part of these corpus construction activities, we have been collecting dialogue data using an experimental translation system between English and Japanese. The purpose of this data collection is to study the communication behaviors and linguistic expressions preferred in front of such systems. We use human typists to transcribe the users' utterances and input them into a machine translation system between English and Japanese instead of using speech recognition systems. In this paper, we present an overview of our activities and discussions based on the basic characteristics.

Full Paper

Bibliographic reference. Takezawa, Toshiyuki / Kikui, Genichiro (2003): "Collecting machine-translation-aided bilingual dialogues for corpus-based speech translation", In *EUROSPEECH-2003*, 2757-2760.