

Panel on Spoken Dialog Corpus Composition and Annotation for Research

Organizers: *Giuseppe DiFabrizio, Dilek Hakkani-Tür, Oliver Lemon, Mazin Gilbert, Alex Rudnicky*

The goal of this forum is to provide researchers from various institutes with the opportunity to comment on a proposed NSF-sponsored data collection plan for a spoken dialog corpus. The corpus is to be used for research in speech recognition, spoken language understanding, dialog management, machine learning, and language generation. Currently, there exists a corpus with over 600 dialog interactions, collected from users using the DiscoH system (from the IEEE SLT 2006 workshop) and the Conquest system (from ICSLP 2006) to obtain general information about conference services. These systems were created as part of a joint collaboration between CMU, ATT, Edinburgh and ICSI.

The workshop panel will host a number of invited researchers who have received a subset of the corpus and annotation. An open discussion will be held to obtain comments from workshop participants to help finalize the annotation guidelines.

If you would like to provide feedback with respect to the annotation plan and receive a sample of the dialog interactions then please send an email to info@discoh.org or go to the following web page:

<http://ramps.speech.cs.cmu.edu/survey/>

where you will find the materials for the survey. These consist of samples from two systems, a questionnaire and some background material on annotation.