

DRAFT

Tue Jul 25 20:03 1995 :: icassp96.janus.dvi

JANUS II — ADVANCES IN SPONTANEOUS SPEECH TRANSLATION

*szyna MFinke T.Kemp A.MNair A.Lawie L.Muyfield M.Mier I.Rojina
T.Sloboda A.Wübel P.Zahn T.Zeppenfeld*

INTERACTIVE SYSTEMS LABORATORIES
at Carnegie Mellon University, USA
and University of Karlsruhe, Germany

ABSTRACT

JANUS II is a research system to design and test components of speech to speech translation systems as well as a research prototype for such a system. We will focus on two aspects of the system: 1) new features and performance of the speech and 2) t

DRAFT

Tue Jul 25 20:03 1995 :: icassp96.janus.dvi

3. RECOGNITION ENGINE

The recognizer used in the current JANUS II prototype system is a CDHMM based recognizer. The exact configuration varies from task to task. For the last *VERBMOBI* evaluation on German scheduling dialog processing on a 7 frame file

DRAFT

The two parsers have clear strengths and weaknesses.

GLR* tries to match input utterances to an interlingua specification, so although words can be skipped with a penalty, the parser is less robust over disfluent input. Input that is parsed, though, is generated in the target language using syntactic constraints; this means that translations through GLR* are more likely to be complete grammatical sentences than those translated through PHOENIX which parses and generates only at the speech act level.

GLR* tends to break down when parsing long utterances that are highly disfluent, or that significantly deviate from the grammar. In many such cases, GLR* succeeds in parsing only a small fragment of the entire utterance, and important input segments end up being skipped. PHOENIX is significantly better in analyzing such utterances. Because PHOENIX is a chart parser that is capable of skipping over input segments that do not correspond to any top level semantic concept, it can recover from out of domain segments in the input, and "restart" itself on the in-domain segment that follows. However, pre-breaking input to GLR* based on occurrences of human noise and parsing the shorter sub-utterances separately significantly reduced this problem.

Pre-breaking benefits PHOENIX only slightly, mainly in better resolution of time expression attachment ambiguities. At the current time, PHOENIX uses only very simple disambiguation heuristics, whereas a parse quality metric helps to decide between possible parses in GLR*.

Computational requirements of GLR*, which is implemented in lisp, are far greater than those of PHOENIX implemented in C. PHOENIX is also much faster, taking 16 ms per parse compared to GLR*'s 1-2 minutes.

Because the two parsing architectures perform on different types of utterances, they may be compared in a way that takes advantage of the strengths of each.

5. SPEECH TRANSLATION RESULTS

As the goal of the translation in JANUS is to preserve the content of an utterance, the recognition (SR) and translation (MT) and end-to-end quality need to be a good indicator of how well the meaning is preserved. The results chosen for evaluation, good, ok, and bad are:

Transcription: tuesday morning

if an important semantic concept is missed during recognition or translation, the translation is judged as bad.

bad (SR): you say

bad (MT): tuesday

if the meaning is preserved

somehow funny, it is

there is still

ok (SR)

ok (MT)

a 100% correct

the

8

8