RECENT ADVANCES IN JANUS: A SPEECH TRANSLATION SYSTEM

M. Woszczyna, N. Coccaro, A. Eisele, A. Lavie, A. McNair, T. Polzin, I. Rogina, C.P. Rose, T. Sloboda, M. Tomita, J. Tsutsumi, N. Aoki-Waibel, A. Waibel, W. Ward

Carnegie Mellon University
University of Karlsruhe

ABSTRACT

esent recent advances from our efforts in increasing cov-
robustness, generality and speed of JANUS, CMUs
to-speech translation system. JANUS is a speaker-
nt system which translates spoken utterances in
also in German into one of German, English or
system has been designed around the task
stration (CR). It has initially been built
h database of 12 read dialogs, encompass-
around 500 words. We have since been
ong several dimensions to improve
age and to move toward sponta-

UCTION

ibe recent improvements of
o speech translation system. Im-
ve been made mainly along the following
ons: 1.) better context-dependent modeling im-
proves performance in the speech recognition module,
2.) improved language models, smoothing, and word
equivalence classes improve coverage and robustness of
the sentences that the system accepts, 3.) an improved
N-best search reduces run-time from several minutes to
now real time, 4.) trigram and parser rescoring improves
selection of suitable hypotheses from the N-best list for
subsequent translation. On the machine translation side,
5.) a cleaner interlingua was designed and sy
and domain-specific analysis were separa
reusability of components and
lation, 6.) a semanti
semantic anal

Th

pendent segment weights.

Error rates using context dependent phonemes are lower by a factor 2 to 3 for English (1.5 to 2 for German) than using context independent phonemes. Results are shown in table 1.

language model	English PP	English WA	German PP	German WA
none	400.0	58.2	425.0	63.0
word-pairs	28.9	83.4	20.8	89.1
bigrams	16.2	92.6	18.3	93.7
smoothed bigrams	18.1	91.5	28.90	84.7
after resorting	—-	98.8		

Table 1: Word Accuracy for First Hypothesis

The performance on the RM-task at comparable perplexities is significantly better than for the CR-task, suggesting that the CR-task is somewhat more difficult.

2.2. Search

The search module of the recognizer builds a sorted l
of sentence hypotheses. Speed and memory req
have been dramatically improved: Tho
of hypotheses computed for each ut
from 6 to 100 hypotheses
computation was r
seconds.

This was a
N-be

When the standard GLR parser fails on all sentence candidates, this robust GLR parser is applied to the best sentence candidate.

3.2. The Interlingua

The output of the parser, known as "syntactic f-structure", is then fed into a mapper to produce an Interlingua representation. For the mapper, we use a software tool known as Transformation Kit [10]. An ping grammar with about 300 rules is writt Conference Registration domain of En

```
((PREV-UTTERANCES ((SPEECH-ACT *ACKNOWL) (VALUE
(TIME *PRESENT)
(PARTY
((DEFINITE +) (NUMBER *SG)
(ANIM -)
(TYPE *CONFERENCE)
(CONCEPT *OFFICE)))
(SPEECH-ACT *IDENTIFY-OTHER))
```

Figure 2: Example: Interlingua Output

Figure 2 is an example of Interlingua representation produced from the sentence "Hello is this the conference of fice". In the example, "Hello" is represented act *ACKNOWLEDGEMENT, and the rest a act *IDENTIFY-OTHER.

3.3. The Generator

The generation of target la representation invo Transform g side there is a "built-in" robustness against these phenomena in a connectionist system

The connectionist parsing process is able to combine symbolic information (e.g. syntactic features of words) with non-symbolic information (e.g. statistical likelihood of sentence types). Moreover, the system can easily integrate different knowledge sources. For example stead of just training on the symbolic in trained PARSEC on both the symboli the pitch contour. After trai tem was able to use t mine the se were