JANUS 93: TOWARDS SPONTANEOUS SPEECH TRANSLATION

M. Woszczyna   N. Aoki-Waibel   F. D. Buø   N. Coccaro   K. Horiguchi   T. Kemp   A. Lavie
A. McNair   T. Polzin   I. Rogina   C. P. Rose   T. Schultz   B. Suhm   M. Tomita   A. Waibel
Carnegie Mellon University — USA
University of Karlsruhe — Germany

ABSTRACT

We present first results from our efforts toward translation of spontaneously spoken speech. Improvements in-
...ude increasing coverage, robustness, generality and speed
JANUS, the speech-to-speech translation system of
...egie Mellon and Karlsruhe University. Recognition
...chine Translation Engine have been upgraded to
...requirements introduced by spontaneous human
...s. To allow for development and evalua-
...on adequate data, a large database with
...g dialogs is being gathered for En-
...sh.

public demonstration in spring 1993. Independently, other speech translation systems [4, 5] have been presented, showing the growing interest in the field.

To begin extending our system to spontaneous human-to-human dialogs, however, improvements and changes along several dimensions of our earlier system are necessary to increase speed, robustness and coverage of the system in the face of ill-formed and ungrammatical spontaneous input. In the following, we will report on the state of these most recent efforts.

2. THE SCHEDULING TASK DATABASE

We have started collecting a large database of human to human dialogs centered around the scenario of appointment scheduling. In each recording session, two subjects are each given a calendar (one of 13 scenarios) and asked to schedule a meeting with the dialog partner. The recording setup allows one person to speak at a time by way of a push-button. Data is being collected in similar fashion for English and Spanish using the same setup at ...sity and Carnegie Mellon.

1. OVERVIEW

...g early systems to attempt the
... It had initially been built
...12 read dialogs of the confer-
...sing a vocabulary of around
...speaker-independent sys-
...nces from English and
...ish or Japanese.

	recorded	transcribed
	450	150
	200	140
...ity)	70	70
	75	30

...ous Scheduling Task Database

...Semantic Parser

...dialogs cover about 10-12 utter-
...long. The test set perplexities
...n initial set of dialogs were
...sh and 88 for German; the
...is about 40% larger
...larger number of inflec-
...size L2
...ribed following the
...anscription format
...in L2
...mispronunciations,
...as pauses.
...ish and Ger-
...[3] dialogs and
...speech
...ated by a

This was achieved by using the word dependent N best
algorithm[8] as a backward pass in the forward backward
pruning algorithm.
Recent experiments with the Wall Street Journal Task
ow that the recognizer can handle vocabularies of up to
0 words at a perplexity of 110.

. Language Models

The most successful language model so far used for the
scheduling task is a model that interpolates a cluster based
model with smoothed bigrams. The cluster based model
s using automatically build classes. The optimization cri-
erion is to minimize the leaving-one-out perplexity on the
aining set [9]. The results shown in table 4 were obtained
ng 150 word classes.

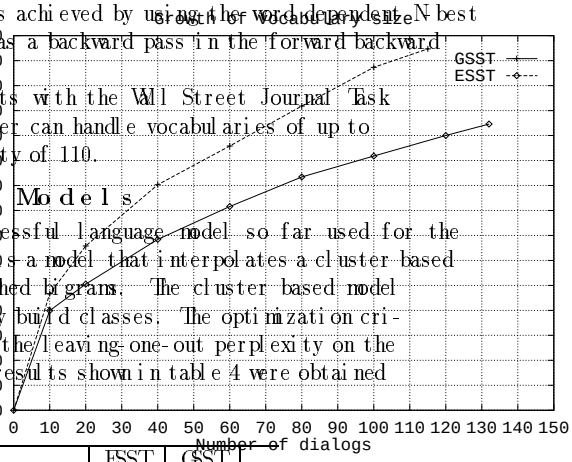

Figure 2. Development of Vocabulary Size

	ESST	GSST
Smoothed Bigrams	46.8	88.2
Cluster	42.2	71.8
Interpolated	39.2	62.1

Perplexity Reduction due to automatic clustering

3. RECOGNITION ENGINE

3.1. Acoustic Modeling

For acoustic modeling, several alternative algorithms are
being explored including TDNN, MS-TDNN, MLP and
LVQ [6, 7]. In the main JANUS system an LVQ algo-
rithm with context dependent phonems is currently used
for speaker independent recognition. for each phoneme,
there is a context independent set of prototypical vectors.
The output scores for each phoneme segment are computed
the euclidean distance using context dependent seg-
ights.
t changes include the introduction of noise mod-
l as the improvement of the training algorithms;
esults in table 2 were obtained using triphone
corrective training and feature weights. Further
using cross word triphones are possible but
ere.

Between Word Noise Models

Noises, filled pauses, and restarts, are much more frequent
in spontaneous speech compared to read speech, particu-
larly as our domain involves human-to-human negotiation.

Noise	English	German
human		
breathing	2166	1017
smack	992	144
	36	4
	21	9
	14	
	75	
	136	

	1991	1993
Registration	9.1 %	3.7 %
gement	24 %	7.5 %

mparison of error-rates

of the recognizer builds a sorted list of
s. Speed and memory requirements have
bly (table 3).

	1991	1993
	3-5 min	3-10 sec
	50Mbytes	7Mbytes
500	10000	
	120	

Search Module.
Task.

GEST-TIME. Other typical speech acts are *STATE-CONSTRAINT, *AFFIRM and *REQUEST-RESPONSE. ... the interlingua ... applied in a modular ... languages can be added ... also allows the separate evaluation of parser and ... by matching against and generating from a set ... interlingua representations.

... The total relative ... 6 generalized noise models ... reduction of between word ... 17% ... English Spontaneous Scheduling task.

```
...ACTUALLY JULY TWENTY SIXTH AND
   TWENTY SEVENTH LOOKS GOOD)
...CT *SUGGEST-TIME)
...ENCE-TYPE *STATE)
...E *FREE)
...RAME *TIME-LIST) (CONNECTIVE AND)
   (ITEMS
    (*MULTIPLE*
     ((FRAME *SIMPLE-TIME)
      (DAY 26)
      (MONTH 7))
     ((FRAME *SIMPLE-TIME)
      (DAY 27)))))
...RB ACTUALLY))
```

...re 3. Example for Interlingua Output

4. THE MACHINE TRANSLATION (MT) ENGINE

The MT-component that we have previously used has now ... that can run several alternate processing strategies in parallel. In translating spoken language from one language to another, the analysis of spoken sentences which suffer from ill-formed input and recognition errors is most certainly the hardest part. Based on ... list of N-best hypotheses delivered by the recognition ..., we can now attempt to select and analyze the most ... sentence hypothesis in view of producing an accu... meaningful translation. ... central in this attempt: *high fidelity* and *accu*... wherever possible, and *robustness* or *grace*... face of ill-formed or misrecognized input. ... parallel modules attempt to address these ... parser 2) a semantic pattern based ap... connectionist approach. The most useful ... is mapped onto a common In... but domain-specific rep...

...arser

... translation process ... parsing ... Compiler. It can use syntactic or ... For application to the sponta... scheduling task, we found ... parser ...

Semantic Pattern Based Parsing

Our robust semantic parser combines frame based semantics with semantic phrase grammars [12]. We use a frame based parser similar to the DYPAR parser used by *Carbonell, et al.* to process ill-formed text [11], and the MINDS system ... developed at CMU. Semantic information is represented in a set of frames. Each frame contains a set of slots ... pieces of information. In order to fill the slots ..., we use semantic fragment grammars. The ... parser can be viewed as "phrase spotting". ... interpretations are pursued simultane... ... a frame with some of its slots ... is represented by a separate Recursive ... which specifies all ways of saying the ... slot. The grammar is a seman... are semantic concepts instead ...

...ist Parsing

... connectionist parsing system PARSEC [13] is used as ... if the LR parser fails to analyze the in... important aspect of the PARSEC system is that ... sentences from a corpus of training ex... the very difficult work of writing ... aspect is that it has proven ro... utterances which frequently are ... restarts, repairs or ungram... integration with other infor... is easier. ... recent developments in PAR...

4.5. The Generator

The generation of target language from an Interlingua representation involves two steps. First, with the same Transformation Kit used in the analysis phase, Interlingua representation is mapped into syntactic f-structure of the target language. The f-structure is then fed into sentence generation software called "GENKIT" to produce a sentence in the target language. A grammar for GENKIT is written in the same formalism as the Generalized LR Parser: phrase structure rules augmented with pseudo unification equations.

As a first experiment for generation on the spontaneous scheduling task, we tried Japanese generation on 264 new hand-coded ILTs. More than 75% of the generated sentences were found to be good or acceptable (see table 6).

Output quality	%
good	65.2
acceptable	11.0
bad	4.5
no output	19.3

…le 6. Quality of Japanese Generation

5. CONCLUSION

…paper we have described a number of system im…
…nd extensions that have recently been intro…
…, to accomodate extension of the speech …
…aneously spoken (not read) human-to-…

…neous negotiation dialogs is being …
…ish and Spanish, and first results …
…s data have been reported.

…DGEMENTS

…edge the collaboration with …
…ons Laboratory, NEC, …
…ssible.

REFERENCES

[1] 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, and W. Ward, *Recent Advances in Janus, a Speech to Speech Translation System*, EUROSPEECH 1993.

[2] L. Osterholtz, A. McNair, I. Rogina, H. Saito, T. Sloboda, J. Tebelskis, A. Waibel, and M. Woszczyna, *Testing Generality in JANUS: A Multi-Lingual Speech Translation System*, ICASSP 1992, volume

… T. Takezawa, F. Yato, S. Sagayama, …a, and A. Kurematsu, *ATR's …SLR…*, EUROSPEECH …