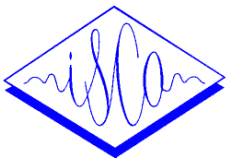

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

DETECTION OF UNKNOWN WORDS IN SPONTANEOUS SPEECH *

Pablo Fetter, Fritz Class, Udo Haiber, Alfred Kaltenmeier,
Ute Kilian, and Peter Regel-Brietzmann
Daimler-Benz AG, Research and Technology, Wilhelm-Runge-Str. 11,
D-89081 Ulm, Germany
e-mail: fetter@dbag.ulm.daimlerbenz.com

ABSTRACT

This paper presents an analysis of the unknown-word problem and results of experiments in acoustic and language modeling of unknown words. In particular, we introduce the method of *iterative substitution* for correcting distortions caused by unknown words in the language model.

1 INTRODUCTION

Most speech recognition systems are designed to find the maximum *a posteriori* probability of a string sequence given an acoustic utterance and a lexicon. As a consequence, the words to be recognized must be predefined (as *known words*), and the lexicon becomes closed. Furthermore, the system can only output words that are contained in its lexicon—if a user utters an *unknown word*, the system will misrecognize it, outputting at best the most similar entry in its lexicon [1].

Recently it has been reported that, on the average, every unknown word causes between 1.2 [2] and 1.6 [8] recognition errors, which could be avoided if unknown words could be detected as such. Furthermore, in an application such as Verbmobil [7] it is absolutely necessary that the system be able to expand its lexicon on-line, for example with proper names.

In Section 2 we make an analysis of the unknown-word problem. Results of experiments in acoustic and language modeling of unknown words are presented in Sections 3 and 4. In Section 4 we also introduce the method of *iterative substitution* for correcting distortions caused by unknown words in the language model.

*This work was partly supported by the German Federal Ministry of Education, Science, Research and Technology.

2 ANALYSIS

Our work is based on the Verbmobil database (also called the German Spontaneous Scheduling Task [6]). To date, about 200 human-to-human dialogs have been collected and transcribed at different German universities. The whole corpus contains over 100,000 entries, including phenomena such as pronunciation variations, word fragments[1], etc.

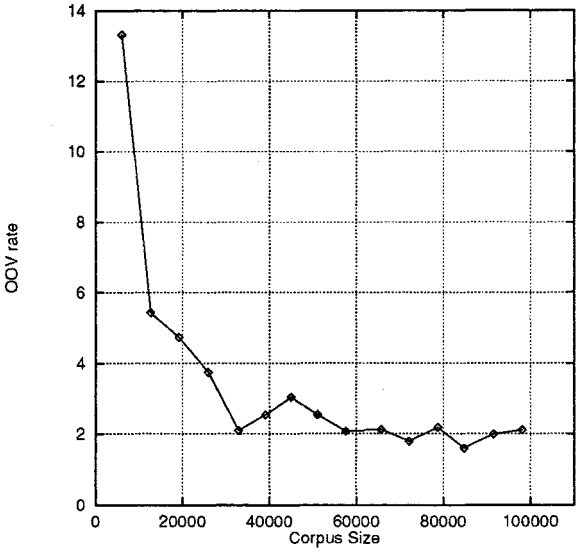

Figure 1: *OOV rate vs. corpus size*

In order to simulate the occurrence of unknown words, we analyzed, in a random order,

[1]Word fragments are always considered unknown words, since they are not included in recognition lexica. They make up about 1% of the corpus, independently of the corpus size.

each of the dialogs from the Verbmobil corpus. For each dialogue we registered all words which had not occurred in a previous dialogue as *unknown* and then added them to the lexicon, thus making them *known* for subsequent dialogs. Figure 1 shows the Out-Of-Vocabulary rate (OOV rate) as a function of the corpus size at different stages of this process. It can be seen that the OOV rate decays exponentially down to a base of aprox. 2%[2]. It can also be seen that the OOV rate remains almost constant after (only) 60,000 words have been seen.

Using this technique we predicted an OOV rate of 2% for the 1995 evaluation within the Verbmobil project. The actual rate was 2.5%.

3 ACOUSTIC MODELING

We trained 15 whole-word models for unknown words using the complete training material. The topology of these 15 models was similar to that of models for frequent (known) words [3]: The number of states depends on the word length; there are loops for every state and skip arcs for adjacent states. The first entry of Table 1 shows the results on the Verbmobil '95 evaluation set with a 3.3KW lexicon and *without* a language model. All results reported here are for word lattices of approximately 5 *hyp/word.*

mod.	*LM*	*WA*	*RCL*	*PRC*
0	no	62.9%	0.0%	0.0%
15	no	62.9%	5.0%	8.84%
15	yes	79.1%	24.9%	11.05%

Table 1: *Word Accuracy (WA), Recall (RCL) and Precision (PRC) without and then with explicit unknown models (mod.); without and then with Language Model (LM).*

Recall (RCL) is the ratio of the number of correctly detected unknown words and the total nuber of unknown words. *Precision (PRC)* is the ratio of correctly detected unknown words and the nuber of hypothetisized unknown words (including false alarms). The *PRC* may seem very low, but not if we consider that the *PRC* of the known vocabulary can at most be 20% in a word lattice with 5 *hyp/word.*

As expected, these very general whole-word models were seldom hypothesized, since they can rarely compete with the known words modeled with sub-word units.

As an alternative to this, we also experimented with augmenting the recognition lexicon. We tried incorporating 35% and then 75% more words into the official 3.3KW lexicon; but in both cases we obtained similar *RCL* for a given *PRC*, compared to the smaller lexicon. Therefore we continued our experiments with the previous acoustic modeling scheme.

4 LANGUAGE MODELING

We then ran the same test with a bigram language model trained with all Verbmobil dialogs available before the evaluation. Even though the overall performance increases significantly, we still did not achieve good performance on unknown words (Table 1, 3rd entry).

This is due to the fact that the probability given by the language model to <UNK>[3] is very low, and also context independent. Both facts are not in accordance with reality, since, as noted above, unknown words are very frequent; on the other hand, the probability distribution of <UNK> should not be constant over all known words. Intuitively, one would say that it is much more probable for <UNK> to follow, say, the sequence *"my name is ..."* than, say, *"let's meet on ..."*.

In [5] it was proposed that all unknown words in the training corpus be mapped into one symbol, and thus to model the context of unknown words explicitly. But how can this particular context be modeled when all words in the corpus are *in* the vocabulary, and thus *known*?

We developed a method which we dubbed *iterative substitution*, which consists of simulating the process of collecting new data and incorporating all new words into the lexicon in a later stage. The basic algorithm is as follows:

- **INIT**

  Define a certain section of the corpus as *known*. The remaining part of the corpus is called *unknown*.

[2]This number represents the minimum word error rate in a system that cannot hypothesize unknown words.

[3]We will use the symbol <UNK> to denote the class *unknown words.* in the language model

- **BODY**
  - make a vocabulary of all words in the *known* section of the corpus.
  - replace all words in the *unknown* section which are not in this vocabulary with the symbol <UNK>. Output the modified corpus.
  - enlarge the *known* section of the corpus by a certain amount.
  - repeat **BODY** until the whole corpus is *known*.
- **END**

  Construct an "augmented corpus" containing all the previously outputted corpora.

At the end of the process we obtain an augmented corpus that is as many times larger than the original one, as iterations made in the process. The probability distributions of the words never followed by <UNK> in the augmented corpus do not change with respect to the original corpus. Moreover, if a word was followed by <UNK> at some stage of the iterative process, a portion of its probability mass will be taken away from the known vocabulary and assigned to <UNK>.

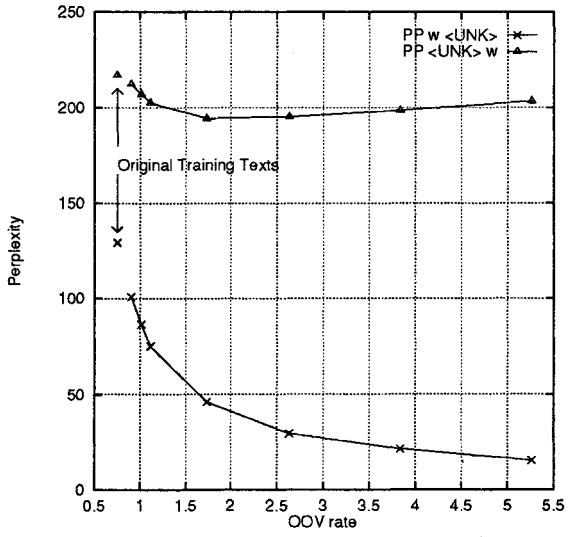

Figure 2: *Perplexity in the vicinity of* <UNK>

In order to test the quality of the language model in the vicinity of <UNK>, we constrained the definition of perplexity PP as follows:

$$PP_{w\text{<UNK>}} = 2^{-\frac{1}{n}\sum_{i=1}^{n} LP(w_i|w_{i-1})}, \quad \text{if } w_i = \text{<UNK>}$$

$$PP_{\text{<UNK>}w} = 2^{-\frac{1}{n}\sum_{i=1}^{n} LP(w_i|w_{i-1})}, \quad \text{if } w_{i-1} = \text{<UNK>}$$

where $w_1 \ldots w_n$ represents, as usual, a word sequence, and LP its log probability.

We then trained several language models with different starting points in the corpus and varying number of iterations. Afterwards, we tested them on unseen texts[4].

Fig. 2 shows the perplexities for different relative frequencies of <UNK> in the training text. As can be seen, the more <UNK>s we have, the "easier" (more probable) it becomes for <UNK> to be predicted by its predecessors ($PP_{w\text{<UNK>}}$ decreases); on the other side, it becomes "harder" for <UNK> to predict its successor ($PP_{\text{<UNK>}w}$ increases) simply because its probability distribution becomes "flatter." This tradeoff results in a U-shaped curve of the overall perplexity. As expected, the minimum is located where the OOV rate of the training text is the same as the test text (Fig. 3).

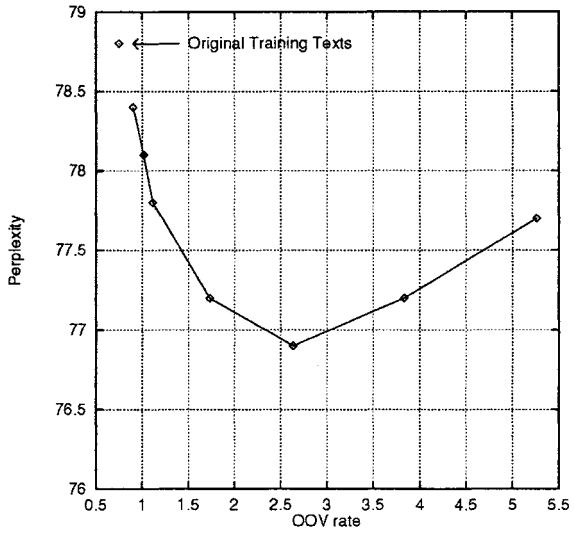

Figure 3: *Perplexity*

We participated in the category "Detection of unknown words" of the Verbmobil Evaluation '95. There we delivered two result sets: a baseline system to compare our progress (see also 1st entry in Table 1), and a system with both acoustic and language modeling of unknown words.

We optimized the language model on the ex-

[4]The material of the Verbmobil Evaluation '94.

pected OOV rate of the test set, and obtained, on the average, a 30% performance gain in detecting unknown words (having a PRC decay of 3%, from 11% to 8%). The results are shown in Figure 4.

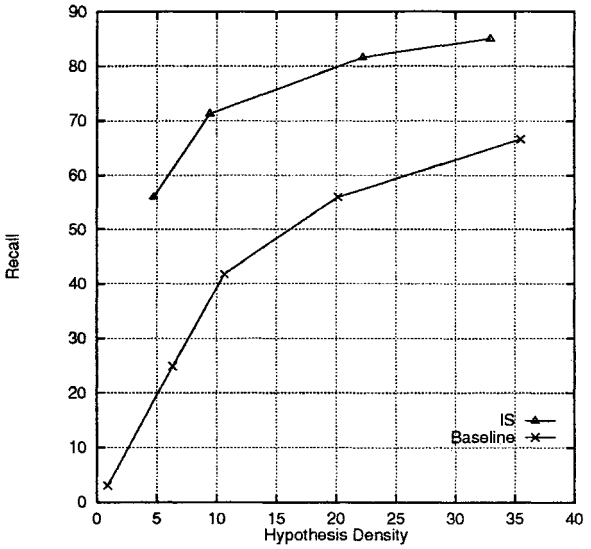

Figure 4: *Recall vs. Hypothesis Density*

5 CONCLUSIONS

Coping with unknown words is an indispensable requirement for real-world applications of speech recognition. We have presented different schemes for explicitly modeling unknown words. On the acoustic side, we tested coarse whole-word models as well as lexicon expansions to capture unknown words. Our conclusion is that coarse models do as well as a lexicon expansion in detecting unknown words (for a given *PRC*). Furthermore, we have modeled the context of unknown words with the method of *iterative substitution*. Using this method we have achieved very positive results.

Acknowledgments

The authors gratefully thank Roni Rosenfeld for providing the excellent SLM Toolkit, as well as his on-line expertise.

REFERENCES

[1] Ayman Asadi et al., *Automatic Detection of New Words in a Large Vocabulary Continuous Speech Recognition System*, ICASSP 1990, pp. 125-8.

[2] L. Chase et al., *Improvements in Language, Lexical, and Phonetic Modeling in Sphinx-II*, Spoken Language Technology Workshop, Austin, Jan. 22-23, 1995.

[3] F. Class et al., *Optimization of an HMM-Based Continuous Speech Recognizer*, Eurospeech '93, pp. 803-6.

[4] E. Paulus and M. Lehning, *Die Evaluierung von Spracherkennungssystemen in Deutschland*, Fortschritte der Akustik, DAGA '49, (in German).

[5] B. Suhm et al., *Detection and Transcription of New Words*, Eurospeech '93, pp. 2179-82.

[6] B. Suhm et al., *JANUS: Towards Multilingual Spoken Language Translation*, Spoken Language Technology Workshop, Austin, Jan. 22-23, 1995.

[7] W. Wahlster, *Verbmobil, Translation of Face-To-Face Dialogs*, Eurospeech '93, Opening and Plenary Sessions, pp. 29-38.

[8] P. Woodland et al., *The Development of the 1994 HTK Large Vocabulary Speech Recognition System*, Spoken Language Technology Workshop, Austin, Jan. 22-23, 1995.