ISCA Archive
http://www.isca-speech.org/archive

6th European Conference on
Speech Communication and Technology
(EUROSPEECH'99)
Budapest, Hungary, September 5-9, 1999

# ERROR CORRECTION TRANSLATION USING TEXT CORPORA

*Kai Ishikawa and Eiichiro Sumita*
ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai, Seika-cho Soraku-gun, Kyoto 619-0288, Japan
ishikawa, sumita@itl.atr.co.jp

10.21437/Eurospeech.1999-440

## ABSTRACT

In this paper, we propose an error correction method using text corpora. In this method, recognition errors are corrected using phonetically similar examples in the text corpora. The reliability of the correction hypotheses are judged according to their semantic consistency and their phonetic similarity to the original input. We previously proposed an error correction method that uses a treebank [1]. However, the previous method was not flexible in its use of examples, because structural mismatches occurred between the input and examples due to recognition errors. In our new proposal, examples are treated as morpheme sequences. This enables us to use examples partially when there are no useful full-sentence-examples. We built our proposed method into a speech translation system and compared the translation quality for simple translation and translation with error correction. The rate of acceptable translation increased about 10% with our proposed method compared to simple translation.

## 1. INTRODUCTION

Extensive studies have been carried out recently on speech translation systems, because of the achievement of high accuracy in speech recognition and (spoken language) machine translation in the past few years. For such a system to be used in practical communication, robustness is quite important and this cannot be achieved by a simple connection between a speech recognition system and a translation system. In order to achieve robustness, it is essential to develop a framework for processing erroneous sentences.

Y. Wakita et al. proposed a robust translation method [2] which translates only reliable parts in recognized sentences, and showed that the misunderstanding rate for translations is reduced by applying the method. In this method, however, the original information of the input sentence is lost for erroneous parts.

In our previous paper announced at ICSLP98, we proposed an error correction method using example sentences. In this method, the hypotheses of the error corrections are produced from the input using example sentences in the treebank, and the reliability of each hypothesis is judged according to its semantic consistency and its phonetic similarity to the input. The method was proven to be feasible by preliminary experiments, however, it has a problem in the flexible utilization of example sentences. Structural mismatching occurs while matching the input structure with the example structure due to the existence of recognition errors.

In this paper, we propose an error correction method to overcome the structural mismatching problem. In our new method, the example sentences are applied by morpheme sequences. Because of this, the flexibility of applying example sentences is increased compared to the previous method using a treebank.

## 2. THE PROPOSED METHOD

### 2.1. Overall process flow

We explain here the overall flow of the processes of our method. The processes are shown below (see Fig. 1) categorized in three steps, i.e.: (1) correction necessity judgement of the input according to semantic consistency, (2) error correction of the input morpheme sequence using phonetically similar example sentences, and (3) reliability judgement of the correction hypotheses according to semantic consistency.

**(1) Semantic correction necessity judgement**

*(1-1) Parsing and semantic distance calculation*

Parse the recognition result using the Constituent Boundary parser (CB-parser) [3] and obtain the dependency structure and semantic distances.

*(1-2) Correction necessity judgement*

The input is judged needing recovery when the total semantic distance is larger than a threshold Γ. Otherwise, the following recovery processes are not executed.

*(1-3) Erroneous part extraction*

Extract sub-structures from the dependency whose semantic distances are larger than the threshold Γ.

**(2) Morpheme sequence correction**

*(2-1) Example retrieval from text corpus*

Retrieve example sentences from the corpus whose morpheme sequences are phonetically similar to the morpheme sequences of the sub-structures obtained in (1-3).

*(2-2) Erroneous part correspondence*

Determine the boundaries of the example phoneme sequences corresponding to the erroneous parts.

*(2-3) Morpheme sequence replacement*

Obtain a correction hypothesis for each example by replacing an erroneous part of the input morpheme sequence with an example morpheme sequence.

*(2-4) Phonetic similarity selection of hypotheses*

The hypotheses obtained in (2-3) are rejected if their phonetic distances are larger than a threshold Δ [1].

**(3) Semantic reliability judgement**

*(3-1) Parsing and semantic distance calculation*

Parse the correction hypotheses that remain in (2-4) in the same way as in (1-1).

*(3-2) Correction reliability judgement*

Output a hypothesis as the recovery result with a total semantic distance that is lower than the threshold Γ. When plural hypotheses are obtained, the most phonetically similar correction is output. When no hypothesis remains, return to (1-3) and continue the correction processes for the hypotheses obtained in (3-1).

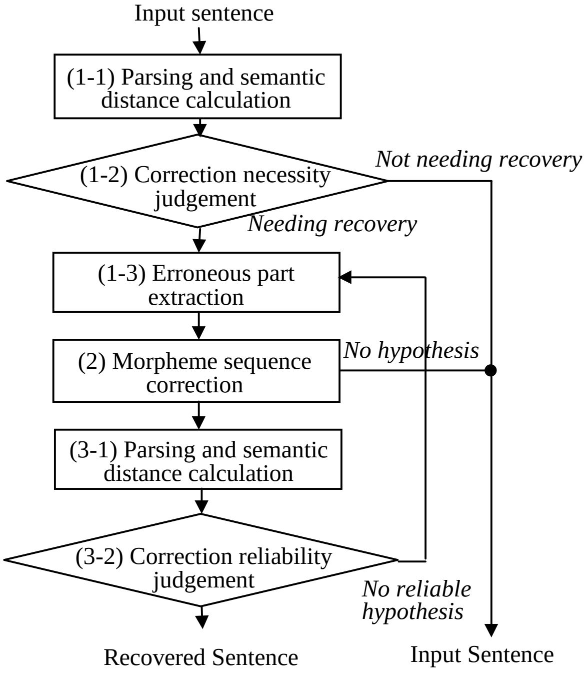
Input sentence
(1-1) Parsing and semantic distance calculation
(1-2) Correction necessity judgement
Not needing recovery
Needing recovery
(1-3) Erroneous part extraction
(2) Morpheme sequence correction
No hypothesis
(3-1) Parsing and semantic distance calculation
(3-2) Correction reliability judgement
No reliable hypothesis
Recovered Sentence
Input Sentence

**Figure 1:** Overall process flow

## 2.2. Details of the processes

### *2.2.1. Parsing and semantic distance calculation (1-1, 3-1)*

In steps (1-1) and (3-1), the recognition results and the error correction hypothesis are parsed. We utilize the Constituent Boundary parser (CB-parser) [3] to obtain the dependency structure and semantic distances. The CB-parser constructs the dependency structure by applying rules in a bottom-up manner. The rules are applied by matching their constituent boundary patterns as expressed by either a functional word or a part-of speech bigram marker. The local semantic distance for the applied rule is calculated according to the semantic distance between the linguistic constituent of the input and the examples defined in the rule for the corresponding constituent, where the distance values are defined according to the semantic hierarchy in the thesaurus [4]. The total semantic distance for a dependency structure is a summation of the local semantic distances for all dependencies. The smaller the semantic distance of the dependency structure, the more reliable the structure is assumed to be. This nature of semantic distance is utilized as the reliability of semantic consistency in our proposed method.

### *2.2.2. Erroneous part extraction (1-3)*

In step (1-3), the semantically unreliable part of the input is extracted according to the dependency structure and its local semantic distances. Each dependency is considered to be reliable semantically if its local semantic distance is small. We extract parts that contain unreliable dependencies as semantically unreliable parts. Firstly, we extract all unreliable dependencies that have semantic distances larger than the threshold value Γ from all dependencies. Secondly, we extract all of the partial dependency structures that contain those unreliable dependencies. Finally, we obtain morpheme sequences for the extracted partial structures as erroneous parts.

### *2.2.3. Example retrieval from the text corpus*

The example sentences in the text corpus are stored in the example database as morpheme sequences with part-of-speech tags. In (2-1) through (2-2), morpheme sequences which are phonetically similar to the erroneous part are extracted. In (2-1), example sentences containing morpheme sequences which are phonetically close to the erroneous part are extracted. Here, phonetic similarity is approximately evaluated by the edit distance of the character sequences, and the problem of example sequences is resolved into the ambiguous matching of the character sequences. In our method, retrieval of the example sentences is carried out utilizing the String Approximate Pattern-Matching method of Y. Lepage [5].

### *2.2.4. Morpheme sequence replacement*

In (2-2), we extract the partial morpheme sequences that contain morpheme sequences which are phonetically similar to the correction parts from the example sentences retrieved in (2-1). The boundaries of the partial morpheme sequences are obtained to minimize the values of the edit distances in the character sequences with the correction parts using DP matching. Consequently, the error correction hypotheses in the morpheme sequence are obtained by repairing the erroneous part of the input with the extracted partial morpheme sequence of the example sentences.

### *2.2.5. Recursive error correction*

In (3-2), the correction hypotheses are not accepted as the final correction result without being judged as recovered according to the semantic distance, even though they are close to the original recognition results. According to this semantic reliability judgement, some correction hypotheses that were partially recovered for errors can also be rejected, because of the existence of some other erroneous parts remaining after the correction. Those partially recovered hypotheses can be completely recovered by applying error correction recursively. In (1-3), we carry out error correction recursively from each correction hypothesis, when there is no correction hypothesis that satisfies the condition of semantic distance.

## 3. EVALUATION

To show the efficiency of the proposed method, we built it into a speech translation system and evaluated the translation results. 337 Japanese sentences from the ATR travel conversation database [6] were used as the input. These sentences are included in the training sentences of the translation system. And, 15,264 sentences from the same conversation database were used as the text corpus. Here, we used the threshold $\Gamma = 1.0$ for the semantic distance, and the threshold $\Delta = 0.3$ for the phonetic distance. Table 1 shows the correlation between the correction necessity judgement and recognition errors in the input sentences.

**Table 1:** Recovery judgement and recognition errors (total of 337 input sentences)

| Recovery Judgement | Erroneous Inputs<br>224 | Correct Inputs<br>93 |
|---|---|---|
| Needing Recovery 153 | **145 (59%)** | 8 (9%) |
| Not Needing Recovery 184 | 99 (41%) | **85 (91%)** |

According to this, 91% of all correct input sentences (no recognition error) are judged "not needing recovery". Table 2 itemizes 153 sentences needing recovery. 59% of all erroneous inputs (including recognition errors) are judged "needing recovery".

**Table 2:** Correction result and recognition errors for 153 sentences needing recovery

| Needing Recovery 153 | Erroneous Inputs 145 | Correct Inputs<br>8 |
|---|---|---|
| Correction obtained 66 | **66** | 0 |
| Correction not obtained 87 | 79 | **8** |

Recovery results were obtained for 66 sentences (27% of the 337 inputs). The recall of the error detection is not so high, but the precision is perfect. Note that, for the 8 correct inputs (9% of the 337 inputs) falsely judged as needing recovery, no recovery results were consequently obtained.

Next, we will explain the evaluation of the translation results. We introduce the evaluation measure shown in Fig.2 to evaluate the results. The following 4 levels correspond to the levels of success in communication. Here, the answer refers to the original correct sentence of the recognized input.

| | |
|---|---|
| **Level A** | Exactly the same meaning as the answer |
| **Level B** | Almost the same meaning as the answer |
| **Level C** | Partially expresses the meaning of the answer |
| **Level D** | Fails to express the meaning of the answer |
| **NIL** | No translation results output |

**Figure 2:** Evaluation measure of translation

In table 3, the evaluation results of the simple translation and the translation with our proposed method are compared for the 337 sentences.

**Table 3:** Improvement by recovery (total of 337 sentences)

| Rank | Simple Translation | Correction Translation |
|---|---|---|
| Level A+B+C | 64% (217) | 74% (249) |
| Level A | 41% (139) | 54% (185) |
| Level B | 11% (38) | 10% (34) |
| Level C | 12% (40) | 9% (30) |
| Level D | 25% (85) | 18% (61) |
| NIL | 10% (35) | 8% (27) |

The results show that the number of input sentences for level A increased after recovery. Table 4 itemizes 66 recovered sentences.

**Table 4:** Improvement by recovery for 66 inputs in which correction was obtained

| Rank | Simple Translation | Correction Translation |
|---|---|---|
| Level A+B+C | 45% (30) | 94% (62) |
| Level A | 5% (3) | 74% (49) |
| Level B | 20% (13) | 14% (9) |
| Level C | 21% (14) | 6% (4) |
| Level D | 42% (28) | 6% (4) |
| NIL | 12% (8) | 0% (0) |

The translation rate of the 66 sentences is 45%, which is much lower than the 64% of all inputs. On the contrary, the translation rate of 94% after correction is high. These results show that our proposed method is also effective for rather badly erroneous inputs, and can obtain appropriate error recovery results.

# 4. DISCUSSION

## 4.1. Comparison with the previously proposed method

In this paper, we proposed an error recovery method different from our previously proposed method in its utilization of example sentences. The example sentences are adopted as morpheme sequences to overcome the problem of structural mismatch in the previous method. Other characteristics in our method include its manner of erroneous part extraction and the recursive error correction in its error recovery. By considering the erroneous parts of the inputs according to the semantic distance, the costs for the retrieval of example sentences and the costs for generating correction hypotheses are reduced, compared to the previous method, in which all possible parts from the input dependency structure were considered. The validity of the recursive error correction is also confirmed for some input sentences that include several erroneous parts and cannot be recovered completely in a single recovery process. However, with regard to the performance of our method, 50% of the “needing recovery” input sentences are recovered, which is almost the same performance as the previous method. This result can be understood by taking into account the condition of the experiment, i.e., a closed test for the database and translation system. In a closed test for the database, the correct sentences of the input are also included in the example sentences of the database. Under this condition, the difference in matching between our new and previous methods does not cause a considerable difference in performance.

## 4.2. Open test for the text corpus

Under the condition of an open test for the text corpus, the existence of the answer for the input in the example sentences is not guaranteed. So, it is important to partially utilize useful examples for the correction, even though they are not the perfect answer for the input. In our proposed method, example sentences are stored as morpheme sequences and flexibly matched to the correction parts of the input. Compared to the previous method using a treebank, our new method is more flexible in utilizing example sentences by extracting useful parts for error correction. Under the condition of an open test for the text corpus, it is also important to store a sufficient number of useful example sentences in the text corpus. The size of the text corpus and the number of useful example sentences should be estimated by comparing performances in using different databases.

## 4.3. Open test for the translation system

The experiments presented in this paper use a closed test for the translation system, i.e., the input sentences are included in the training sentences for the translation system. The semantic distances of the training sentences are assumed to converge in the process of training the CB-parser in the translation system. So, in the closed test, the semantic distance of the input sentences also converges if they are free from recognition errors. This fact guarantees the reliability of utilizing semantic distance in the correction necessity judgement. Under the condition of an open test for the translation system, a decline in the preciseness of the semantic consistency judgement is expected. This decrease is easily predictable, because the convergence of the semantic distance is not assumed for the open input sentences. Moreover, the decline in the preciseness of the erroneous part extraction can also be predicted according to the same reason. We will estimate these declines and confirm the validity of our method for the open inputs.

# 5. CONCLUSION

In this paper, we presented an error recovery method which enables a more flexible use of example sentences compared to our previous method. The validity of our proposal is shown by evaluating the results of speech translation with the error correction method. We are now in the process of adapting our method to an open test for the database and an open test for the translation system. The validity of our method under those conditions will soon be confirmed.

# 6. REFERENCES

1 K. Ishikawa, E. Sumita, H. Iida. 1998. Example-Based Error Recovery Method for Speech Translation: Repairing Sub-Trees According to the Semantic Distance. ICSLP98, vol. 4, pp. 1147-1150.

2 Y. Wakita, J. Kawai, H. Iida. 1997. Correct parts extraction from speech recognition results using semantic distance calculation, and its application to speech translation. In Proc. of Spoken Language Translation (ACL '97 Workshop).

3 O. Furuse, H. Iida. 1996. Incremental Translation Utilizing Constituent Boundary Patterns. In Proc. of COLING '96, pp. 412-417.

4 S. Ohno and M. Hamanishi. 1981. Ruigo-Shin-Jiten. Kadokawa-Shoten.

5 Y. Lepage. 1997. String Approximate Pattern-Matching. 55th Meeting of the Information Processing Society of Japan, Fukuoka, August 1997. vol. 3, pp. 139-140.

6 Tsuyoshi Morimoto, Noriyoshi Uratani, Toshiyuki Takezawa, Osamu Furuse, Yasuhiro Sobashima, Hitoshi Iida, Atsushi Nakamura, Yoshinori Sagisaka, Norio Higuchi, and Yasuhiro Yamazaki, “A Speech and Language Database for Speech Translation Research ,” Proc. of ICSLP `94, pp. 1791-1794, 1994.