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

6th International Conference on Spoken Language Processing (ICSLP 2000)
Beijing, China
October 16-20, 2000

Fine keyword clustering using a thesaurus and example sentences for speech translation.

Yumi WAKITA† § *Kenji MATSUI*† *Yoshinori SAGISAKA*‡ §
†Advanced Technology Research Laboratories, Matsushita Electric Industrial Co., Ltd., Kyoto, Japan
‡ATR Spoken Language Translation Research Laboratories, Kyoto, Japan
§Graduate School of Science and Technology, Kobe University, Kobe, Japan

Abstract

For robust speech translation, we propose a new language translation method in which speech recognition results are mapped to example sentences using keywords. In this method, the keyword clustering is used to cope with recognition errors and the wide variety of words that do not appear in the training corpus. Initial classes defined using only thesaurus are re-defined by using the dependency between the keywords in limited number of example sentences. The effectiveness of our keyword clustering method is confirmed through example sentence search experiments. These experiments were done using keyword sets of (a) different sentences including keywords not in the example sentences and (b) recognition results those sentences in which recognition errors were obtained. Compared with the search method which uses keyword sets defined by using only a thesaurus, our proposed method offered improved search error rates.

1. Introduction

For practical speech translation, it is necessary to develop a robust speech translation technique capable to dealing with sentences involving recognition errors. To develop such a technique, we apply an example-sentence-driven method to the language translation portion in domain that use simple and limited expressions, such as travel conversation domain.

Instead of directly translating recognition results using conventional language translation methods, our method selects an example sentence from among a fixed number of prestored sentences by considering only the dependency between keywords in the recognition result. Translation is then carried out by substituting constituent words into the selected sentence. Once the keywords are correctly recognized, this selection method is free from the misrecognize other words and the inserted words.

For domains using only simple and limited expressions, it is possible to collect translation pair examples covering almost all expressions, it is difficult, however, to collect examples covering all keywords used in the domain. As a result a suitable keyword class must be defined for all keywords in order to search for a suitable example from a recognition result that includes keywords not in the example sentences.

The effectiveness of a thesaurus class has been confirmed to structural disambiguation for example-based speech translation[1], but many words belong to several semantic classes. To select a suitable example certainly, defining only one suitable keyword class from several semantic classes is necessary.

Several methods have been proposed for word sense disambiguation using statistical or knowledge based methods, and the effectiveness have been already confirmed[2][3][4]. In this paper, we propose a fine keyword clustering method to search for suitable sentences certainly and also we report the effectiveness of proposed keyword clustering to select a suitable example sentences selection for robust speech translation.

In Section 2 we summarize our speech translation method using selected example sentences In Section 3, we describe our keyword clustering method. In Section 4, we report evaluation results confirming the effectiveness of our clustering method in a speech translation system.

2. Speech translation using example sentences driven method.

Figure 1 shows the structure of our speech translation method. Instead of directly translating recognition results as conventional language translation methods do, an example sentence is selected from among a fixed number of prestored sentences by using the dependency between only keywords in the recognition result. Afterwords, translation is done by substituting constituent words into the selected sentence. Once keywords are correctly recognized, this selection method is free from the misrecognition of other words.

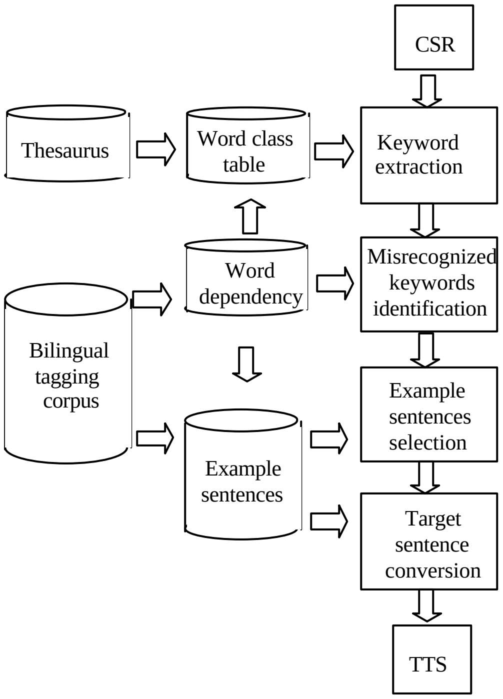

Fig. 1. Structure of speech translation method

If keywords are misrecognized, erroneous translation must be avoided. Our method is capable of identifying misrecognized keywords by considering the dependency between keywords in example sentences and deleting these misrecognized keywords from the group of keywords in recognition results. The strength of the dependency between recognized keywords is evaluated by comparing keywords

$$\mathrm{Sim}(W_j^A \bullet W_i^A) = \sum_{k=1}^{K} R(W_j^A \bullet Z_k) \times \sum_{k=1}^{K} R(W_i^A \bullet Z$$

whose dependency have already analyzed. If a recognized keyword is uncorrelated with the others, the keyword is identified as an error. Thus a revised set of new keywords is used for selecting example sentences.

In addition to these strengths, for a compact speech translation system, our translation method has a mechanism that interprets speech-act as opposed to directly translating sentences. This method maps the simplified expressions instead of doing deep parsing. All training sentences in the bilingual corpus are rewritten into the simplified expressions and grouped by rules. By using the simplified expressions, the system reduces a number of translation rules and processing time. [5][6]

3. Fine keyword clustering

3.1 Why is keyword clustering necessary?

In the domain of simple and limited type of expressions such as travel guide tasks, it is possible to collect the bilingual translation corpus which covers almost all of the necessary expression patterns. But, collecting all of the vocabulary related to the domain is very difficult. It is therefore necessary to handle the keywords as the suitable keyword classes, so that the example sentence search module can look at not the actual keyword but the classes. For example, the sentence "I would like to have a coffee", can be "I would like to have [drink]." A thesaurus can be used to define keyword classes, but following problems occur:

(1) Difficulty in defining a single suitable keyword class because many of the words can belong to multiple semantic classes,

(2) Difficulty in distinguishing words whose meaning are different from other words in the same cluster.

To solve these problems, we propose a fine clustering method that considers the similarity between keywords. This similarity is defined by using the dependency analysis results of example sentences. The dependency is an asymmetric binary relationship between a word called head and another word called modifier[7]. The clustering is done to only head side of words. When two different heads frequently depend on the same modifires, these heads are clustered into one class. As a result of clustering, (1) suitable classes for limited domain can be selected from all thesaurus classes and (2) the words whose meaning are different from other words can be split from the original thesaurus class. In the following sub-sections, we describe how the similarity value between keywords is calculated and how fine clustering using the similarity values is defined.

3.2 Keyword clustering using similarity of dependency between keywords

This similarity is calculated by using the following formula. If both heads, Wj and Wk, belong to class A, the similarity between Wj and Wk is calculated as follows.

Where

where

if $R(Wj^{A}\ \ Zk) = 0$, then $R(Wi^{A}\ \ Zk) = 0$

if $R(Wi^{A}\ \ Zk) = 0$, then $R(Wj^{A}\ \ Zk) = 0$

$$R(W_j^A \bullet Z_k) = \frac{FreqPair\ (W_j^A \bullet Z_k)}{Freq\ (W_j^A)}$$

$Sim(Wj^{A}, \bullet Wi^{A})$: similarity value between Wj and Wi

Zk : k-th words of modifires

FreqPair(Wj^{A},Zk) :

Number of case Wj^{A} that depends on Zk

Freq(Wj) : Frequency of Wj

K : Number of all modifires

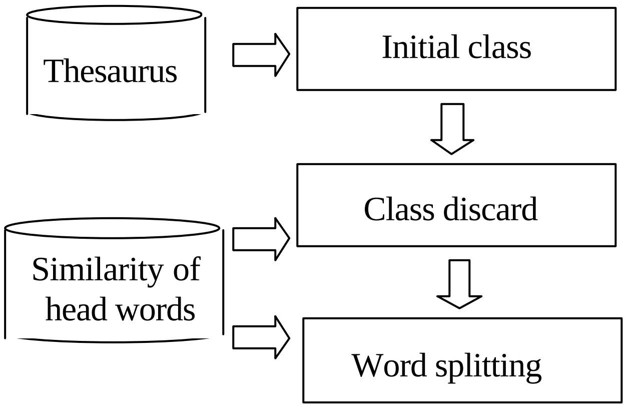

Fig.2 Fine clustering method

First, the initial keyword classes are defined by using the thesaurus information. Next, the initial classes are re-defined according to the following conditions.

Condition 1 The average of the similarity value between heads are calculated by each class. The classes with average less than threshold are discarded.

Condition 2 If all of the similarity values between a head word and other heads are less than the threshold value, the head word is split from the class.

3.3 Evaluation of the Fine Clustering

We performed an evaluation to make sure the effectiveness of using the similarity between keywords based on dependency relationship for clustering and the clustering conditions described in section 3.2 is reasonable. The evaluated sentences are generated based on the example sentences by substituting the keywords in example sentences to other words in same class. By evaluating the understandability of the generated sentences by two people, we tested our clustering method. First, each person gave one of the following five score (i.e. 1 through 5) to each sentence and next, the average of understandability scores, called AUS, were calculated for each cluster. The classes with the higher AUS were regarded as more suitable classes.

1. Incomprehensible
2. Mostly incomprehensible, but a portion could be understood.
3. Comprehensible, but expression was awkward
4. Comprehensible, but expression was slightly awkward
5. Comprehensible and natural.

In our evaluation, we used 770 sentences from customers in a restaurant retrieved from the ATR language database.

3.3.1 Effectiveness of class dissolution using similarity between keywords

We performed an evaluation test to make sure the “clustering condition 1” described in Section 3.2 is reasonable. That is the process of discarded clusters whose average similarity values are less than the threshold, we compared the AUS of the classes with higher similarity values with those with lower ones. Table 1. shows the similarity values and AUS for each class. The classes whose similarity values were high had also high AUS values. This suggests that using similarity values is useful, and the classes with low similarity values should be dissolved in fine clustering decisions.

Table 1. Effectiveness of dependency similarity between keywords in fine clustering

High similarity value			Low similarity value		
Class	Ave. of Similarity	AUS Value	Class	Ave. of Similarity	AUS Value
drink	0.662	[illegible]	pronoun	0.033	[illegible]
food	0.513	[illegible]	color	0.014	[illegible]
seasoning	0.457	[illegible]	place	0.008	[illegible]

3.3.2 Effectiveness of keywords splitting

We also performed a evaluation test to see if the “clustering condition 2” described in Section 3.2 is effective. That is words with low similarity values are split from the class we compared AUS before splitting with those after splitting. Table 2 shows that the AUS for classes whose similarity values are high. The AUS value after splitting were higher. This suggests that splitting is useful in fine clustering.

Table2. Effectiveness of word splitting in fine clustering

	AUS	
Class	before splitting	after splitting
drink	0.662	0.678
food	0.513	0.602
seasoning	0.457	0.510

Example Selection Experiments

4.1 Experiment conditions

To demonstrate the effectiveness of proposed clustering for example sentences selection, we compared correct sentences selection rate by proposed clustering with its by the clustering using only thesaurus. For the evaluation, we used sentences of two domains, the restaurant domain used the experiments in section 3.3 and the shopping domain. The database conditions are shown in Table.3.

Table 3. Experimental conditions

Domain	restaurant	shopping
Training of clustering and examples	770 sentences	583 sentences
Test - open sentences	160 sentences	170 sentences
- recognition error sentences	88 sentences	97 sentences
Keywords of head side	675 words	532 words
Number of classes using only thesaurus	144 classes (675 words)	158 classes (532 words)
Number of classes using proposed clustering	61 classes (367 words)	50 classes (296 words)

4.2 Example sentence selection experiment for correct recognition results

To confirm the effectiveness of our clustering toward example sentences selection for correct recognition results, The meaning of the selection results were evaluated by comparing with those of the open correct sentences. The selection results were evaluated by two people. Each gave a score (i.e. 1 through 5) to each sentence.

1.Incomprehensible the meaning of input sentences
2.Incomprehensible the meaning, but a portion could be understandable.
3. Comprehensible the meaning, of the input, but the expression was quite awkward.
4. Comprehensible the meaning of the input, but the expression was a little awkward.
5. Comprehensible the meaning of input perfectly.

Each of average rate of the two domains and two evaluators are shown in Table 4. The average scores for the proposed clustering is high than the case of using only thesaurus (proposed: 3.66, only thesaurus: 3.22). The number of comprehensible results (over score 3) using proposed method are more than the ones using thesaurus only (proposed: 64.5% , only thesaurus: 58.5%). The results suggest that the proposed clustering is effective for improvement of sentences selection rate. Also the rejection rate by using the proposed method is higher. When the inputs are open sentences which include many unknown keywords, they were rejected by the proposed method. But by thesaurus only clustering, the meaning of selected sentences are sometimes quite different from the ones of input sentences. It also shows that the proposed clustering is effective to avoid selecting incomprehensible examples.

4.3 Example sentence selection rates in case the recognition errors occur

To confirm the effectiveness of our clustering toward example sentences selection for misrecognized recognition results, the same experiments as the previous section 4.2 were done using misrecognized results. The result is shown in Table 5. It suggests that the example selection errors by using only thesaurus could be changed to the rejection case by using the proposed method. The rejection case increased by using proposed clustering (from 52.5% ro 60.5%), on the other hand, The error case (score 1) decreased by using proposed clustering (from 14,5% to 6.5%). The proposed method is effective to decrease the selection error rate. However, the rate of comprehensible results by the proposed method is almost same as the one by using the proposed method (proposed: 26% only thesaurus: 24.4%). It suggests the proposed method is not effective to change the incomprehensible results to the comprehensible results.

Table 4. The effect of proposed clustering toward example sentences selection for correct recognition results. (%)

score						Rej .
using only thesaurus	14.5	15.6	18.6	15.1	24.8	11.4
Using similarity of dependency	4.1	13.1	19.7	15.4	29.4	18.8

Table 5. The effect of proposed clustering toward example sentences selection for misrecognized results. (%)

score						Rej .
using only thesaurus	14.5	8.6	4	13.4	7	52.5
Using similarity of dependency	6.5	7.0	10.9	4.3	10.8	60.5

Conclusion

We proposed a keyword clustering method using similarity of dependency between keywords in limited number of example sentences. First, the initial keyword clusters are defined by using the thesaurus information only. Next the initial clusters are re-defined by using similarity of dependency between keywords. To confirm the suitability of the proposed clustering method, we evaluated the understandability of the sentences generated by substituting the keywords to other keywords in same class. We confirmed the average of understandability score (AUS) of each class increased by using proposed clustering method rather than initial clustering by using thesaurus only. Also the effectiveness of our clustering method for examples selection was confirmed by comparing with the clustering using the clustering only thesaurus dictionary. Our method is useful to increase the correct example selection rate and also effective to avoid the in comprehensible results both for open sentences and misrecognized sentences. In future, we plan to apply our method into a translation system and evaluate the performance.

6. References

[1] O.Furuse,H.Iida: Constituent Boundary Parsing for Example-based Machine Trasnlation: Proc. Coring94, pp. 105-111 (1994)

[2] D.Yorowsky: Word-Sense Disambiguation using Statistical Models of Rogets Categories Trained LargeCorpola. Proc. Coring92, pp.454-460 (1992)

[3] E.Agirre, G.Rigau : Word Sense Disambiguation using Conceptual Density. Proc. of Coling96, pp.16-22

[4] Philip Resnik: Disambiguating Noun Groupings with Respect to WordNet Senses. Proc. of the 3rd Workshop on Very Large Cprpola, MIT (1995)

[5] Chengqing Zong, Yumi Wakita Bo Xu, Kenji Matsui and Zhenbiao Chen : Japanese-to-Chinese Spoken Language Translation Based on the Simplified Expression, ICSLP2000(2000)

[6] Yumi Wakita, Kenji Matsui, Yoshinori Sagisaka: Robust Speech Tanslation using Fine Keyword Clustering, Workshop on Multi-Lingual Speech (2000.10)

[7] Dekang Lin: Using Syntactic Dependency as Local Context to Resolve Word Sense Ambiguity, Proc. of ACL-EACL' 97, pp.64-61.(1997)