AN EXPERIMENT ON KOREAN-TO-ENGLISH AND KOREAN-TO-JAPANESE SPOKEN LANGUAGE TRANSLATION

Jae-Woo Yang and Jun Park

Human Interface Department, ETRI
161 Kajong-dong, Yusung, Taejon, 305-350, Korea
jwyang@media.etri.re.kr

ABSTRACT

We have implemented a Korean-to-English and Korean-to-Japanese spoken language translation system prototype. The system can translate speech in travel planning domain with 5,000 word vocabulary. In our prototype, we concentrate on how to transfer the intention of a user to the partner in spite of current limitation of spoken language processing technology. We measured the end-to-end performance of the prototype to test whether the output of the system is understandable using a subjective measure. We also used an objective measure to evaluate the system performance and found that it generates coherent result with the subjective test. The test result shows that the user can understand the output even in the case that the system cannot translate speech correctly. Thus it is important to provide even partially correct translation output to the user, in order not to neglect the possibility that the user can infer the intended message using the context and his/her intelligence.

1. INTRODUCTION

Spoken Language Translation (SLT hereafter) system is an important tool for global communication in order to break the language barrier. International collaboration is a key to success because of its dependence on languages. Five countries agreed to perform international joint experiment on multimedia, multilingual SLT system in C-STAR consortium. We, ETRI have a role to provide Korean-to-English and Korean-to-Japanese SLT system. English-to-Korean and Japanese-to-Korean translation will be provided by CMU and ATR, respectively. We built an SLT system prototype and evaluated the performance.

The current SLT system suffers from the performance limitation. Especially, the speech recognition performance of spontaneous is far from satisfactory. The goal of our approach is to achieve a high end-to-end, i.e., human-to-human performance in contrast to those of most conventional speech translation systems pursuing only high input-to-output performances. The system can provide even partially correct translation output to the user, in order not to neglect the possibility that the user can infer the intended message using the context and his/her intelligence. Our system also provides multimedia channels which can be utilized to produce more successful results from the error-prone translated speech channel.

2. THE PROTOTYPE

Figure 1 shows the block diagram of ETRI's SLT system prototype. We use microphone array for remote speech input. To eliminate button control, we detect speech period of the signal from microphone array. The speech recognizer processes detected speech signal into Korean text[1]. To translate this text into English, we use the interlingua approach. To translate the Korean text into Japanese, we use Token-based transfer-driven machine translation (TB-TDMT)[2]. Then the English and Japanese speech synthesizer produces English and Japanese speech, respectively.

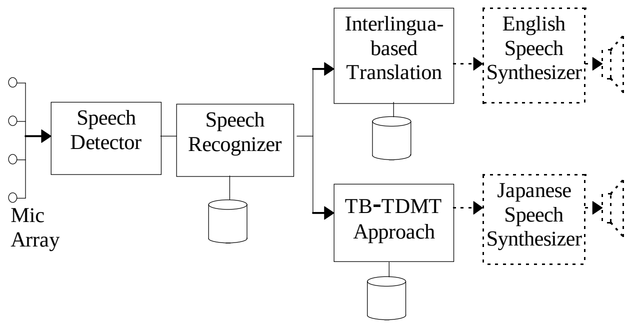

Figure 1. System configuration of the SLT system.

3. TEST DATA

Total 590 bilingual dialogues are collected for travel planning domain. The travel planning includes hotel reservation, air ticket reservation, scheduling, and so on. A half of them is Korean-to-English and the rest is Korean-to-Japanese. In the bilingual data collection situation, human interpreter acts as an SLT machine, and controls each speaker's turns to avoid the overlap of utterances.

Table 1. Characteristics of data sets

set	description	dialogs	utterances	words	voc.
A	Korean Dialogue	590	3090	43244	5499
B	Korean Test Dialogue	55	283	3925	1316
C	Recognized Dialogue	55	283	2972	1216
D1	Translated in Japanese	55	283	2634	1239
D2	Translated in English	55	283	2775	398

Unlike the monolingual data collection, dialogues in bilingual database are more straight forward. Speakers do not talk much about out-of-domain issues. They do not feel comfortable as in monolingual dialogue, so they tend to finish the utterance as soon as possible. The number of vocabulary used is much smaller than the monolingual case. However, vocabulary size is still quite large.

Among the 590 dialogues, we select 480 dialogues as training data, and 55 dialogues as a test set. The remains are reserved for future evaluation. The specification of the speech database used in experiments is shown in Figure 2 as well as in Table 1.

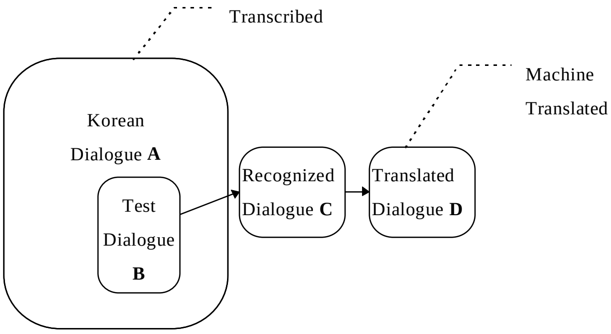

Figure 2. Data sets in bilingual dialogue experiment.

4. SUBJECTIVE TEST

Performance of Korean-to-English Translation

To measure the understandability of speaker's intention, we graded the output quality in three levels; A: user understand the speaker's intention perfectly, B: user understand the speaker's intention despite of minor errors, and C: user cannot understand. Then, we assigned 1 to grade A, 0.7 to B, and 0 to C, respectively. Table 2 shows the understandabilty of the translated data. Here, each point (dialogue set) is 5 dialogues.

The resultant understandability is low. This is partly due to the strict criterion for the case C. For example, in travel planning task, there are many digit information such as phone number, credit card number, and date. While the digit in Korean is rather hard to recognize, but it contains the critical information in travel planning task. Thus, we classified the output utterance as C whenever any of these information was not correctly translated.

This also implies that multimedia could be quite useful for SLT. If you use keyboard for input of those digit information and/or proper names, the understandability of the system will be increased considerably.

Table 2. Understandability of dialogue set

dialogue set	understandability
1	0.48
2	0.30
3	0.48
4	0.30
5	0.49
6	0.49
7	0.55
8	0.62
9	0.56
10	0.37
11	0.30

Performance of Korean-to-Japanese Translation

We used the different evaluation method for Korean-to-Japanese translation. We measured the understandability with 5 grades, i.e. perfectly understandable, almost understandable, moderately understandable, mostly incomprehensible, and nothing understandable. For each grade, we assign number 1 to 5 (the less, the better).

Four language experts participated in the evaluation. Table 3 and Figure 3 show the correlation among the experts. Here, we multiplied the number of utterances to the understandability of each dialogue because experts evaluated the performance utterance by utterance. This results show that there exists rather a large variation between evaluating experts.

Table 3. Correlation coefficients of evaluations made by experts

	Expert B	Expert C	Expert D
Expert A	0.88	0.72	0.60
Expert B	-	0.67	0.80
Expert C	-	-	0.79

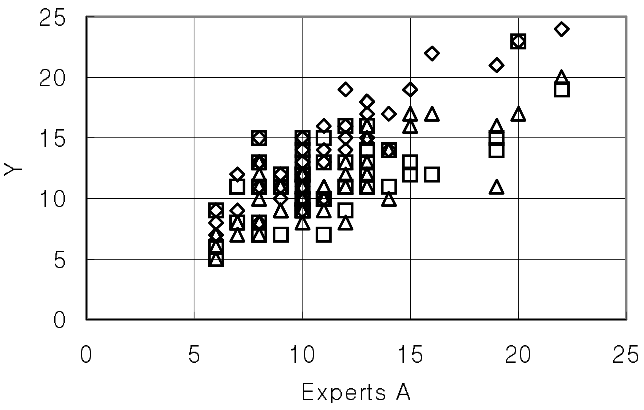

Figure 3. Correlation between evaluations made by experts weighted by no. of utterances (evaluation results from three experts are drawn with respect to the expert A's evaluation).

5. OBJECTIVE TEST

An objective test is more robust and less expensive than a subjective test. In this section, we define an objective measure for SLT systems.

First, we define the information quantity of a word using Shannon's information theory [3, 4]. Let *p(s)* denotes the occurrence probability of the word *s* in language A. We define the information quantity *Q(s)* as

$$Q(s) = -\log p(s). \qquad (1)$$

Then, the information quantity *Q(s)* of a sentence **s** = {s_1, s_2, ... , s_N } is

$$Q(\mathbf{s}) = -\sum_{i=1}^{N} \log p(s_i) \qquad (2)$$

assuming that there is no correlation among the words in **s**.

Suppose a sentence **s** in language A is transformed into the sequence of translated words $\mathbf{t} = \{t_1, t_2, , t_M\}$ in language B. We define the information quantity T(**t**) as

$$T(\mathbf{t}) = -\sum_{i \in C} \log r(t_i) + \sum_{j \in I} \log r(t_j), \qquad (3)$$

where $r(t_i)$ is the occurrence probability of word t_i in language *B*, and *C* and *I* are the set of correctly and incorrectly translated words, respectively.

Since the incorrectly translated words can prevent a listener from understanding the speaker, they destroy the information conveyed in the correctly translated words. This effect is incorporated as the second term of (3).

We define the performance *e* of SLT as

$$e = \frac{T(\mathbf{t})}{Q(\mathbf{s})}. \qquad (4)$$

Observe that *e* becomes less than one even if the translation is 'correct' for all words if a sequence of words in one language is translated into a sequence of words in another language with less discriminative power.

Performance of Korean-to-English Translation

We calculated the proposed objective measure for Korean-to-English translation output and indicated its correlation with understandability in Figure 4. As in the understandability evaluation, each point (dialogue set) is 5 dialogues. It shows the objective measure is coherent with the understandability. The correlation coefficient between *e* and understandability is 0.87.

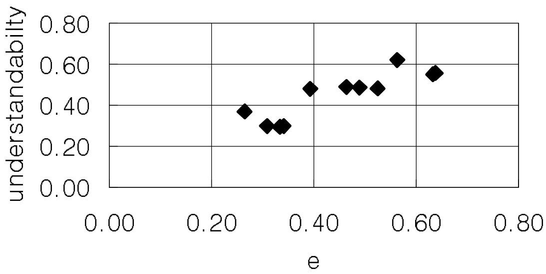

Figure 4. Correlation between *e* and understandability in Korean-to-English translation.

Performance of Korean-to-Japanese Translation

Figure 5 also shows the correlation between *e* and understandability for Korean-to-Japanese translation. Remember that we assigned the number one to five to each understandability grade. We found stronger correlation than that of Korean-to-English translation (correlation coefficient = 0.90). This is because Korean and Japanese have more similar language structure and word-to-word correspondence.

6. SYSTEM PERFORMANCE

The performances of each module and the whole system are shown in Table 4. Here, we assume there is no error in speech synthesis. The recognition performance is measured in terms of word accuracy. For the evaluation of the translation and system performance, we simply classified both Case A and Case B as correct translation, rather than assigning the weight for each case. That is why the performance is better than that in Table 2. Note that the system performance is larger than the product of speech recognition performance and machine translation performance. This shows that the user can understand the output even in the case that the system cannot translate speech correctly.

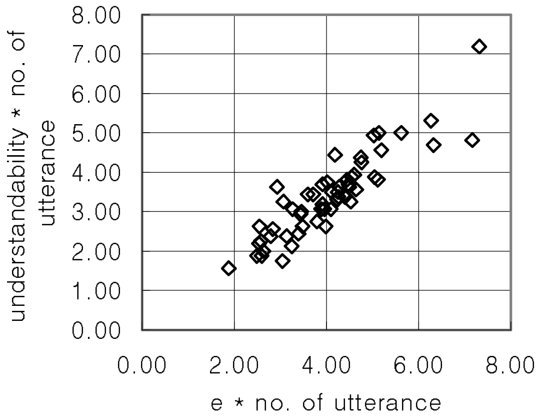

Figure 5. Correlation between e and understandability weighted by no. of utterances in Korean-to-Japanese translation

Table 4. Performances of modules and the whole system

	Speech Recognition performance	Machine Translation Performance	System Performance
Korean-to-English	0.72	0.76	0.56
Korean-to-Japanese	0.72	0.94	0.76

7. CONCLUDING REMARKS

We implemented Korean-to-English and Korean-to-Japanese SLT based on the strong belief that an SLT system should be designed to utilize human information processing capability that catches the meaning even from a partially translated utterance. We have also proposed an objective, end-to-end performance measure for SLT systems. This measure shows a strong support for the idea that the translation of spoken language should be synchronized with multimedia channels.

We have tested the performance of SLT system prototype. We have collected 590 Korean-to-English and Korean-to-Japanese bilingual dialogues. Among them, 55 dialogues (283 utterance) are used for the test. From the test we have found the followings.

1. Partial translation output can be very useful for understanding of speaker's utterance.
2. Subjective test for understanding reveals large variance at utterance level. But there exists correlation among them in a large scale.
3. Proposed objective measure is coherent with human understandability.

To make an SLT system useful in the near future, a lot of works should be done. One of the highest priority work is to adopt multimedia and to design multimodal interaction carefully to reduce the errors from the out-of-vocabulary or out-of-domain words. The interaction should guide users to restrict their vocabulary, and the constraint should be designed to be natural enough not to limit the content of their communication.

REFERENCES

[1] H.S. Lee, J. Park, and H.-R. Kim, "An Implementation of Korean Spontaneous Speech Recognition System," 1996 International Conference on Signal Processing Applications & Technology, pp. 1801-1805, Boston, USA, Oct., 1996.

[2] J.-W. Yang and Y. Lee, "Toward translating Korean Speech into Other Languages," Proc. of International Conference on Spoken Language Processing, Philadelphia, USA, 1996.

[3] C.E. Shannon, "A Mathematical Theory of Communication," Bell syst. Tech. Journal, 27, pp.379-423m 1948.

[4] J.-W. Yang et al., "Multimedia Spoken Language Translation," IEICE Trans. on Information and Systems, Vol. E79-D, No. 6, 1996.