Translation using Information on Dialogue Participants

Setsuo Yamada, Eiichiro Sumita and Hideki Kashioka
ATR Interpreting Telecommunications Research Laboratories*
2-2, Hikaridai, Seika-cho, Soraku-gun,
Kyoto, 619-0288, JAPAN
{*syamada, sumita, kashioka*} *@itl.atr.co.jp*[†]

Abstract

This paper proposes a way to improve the translation quality by using information on dialogue participants that is easily obtained from outside the translation component. We incorporated information on participants' social roles and genders into transfer rules and dictionary entries. An experiment with 23 unseen dialogues demonstrated a recall of 65% and a precision of 86%. These results showed that our simple and easy-to-implement method is effective, and is a key technology enabling smooth conversation with a dialogue translation system.

1 Introduction

Recently, various dialogue translation systems have been proposed (Bub and others, 1997; Kurematsu and Morimoto, 1996; Rayner and Carter, 1997; Rosé and Levin, 1998; Sumita and others, 1999; Yang and Park, 1997; Vidal, 1997). If we want to make a conversation proceed smoothly using these translation systems, it is important to use not only linguistic information, which comes from the source language, but also extra-linguistic information, which does not come from the source language, but, is shared between the participants of the conversation.

Several dialogue translation methods that use extra-linguistic information have been proposed. Horiguchi outlined how "spoken language pragmatic information" can be translated (Horiguchi, 1997). However, she did not apply this idea to a dialogue translation system. LuperFoy et al. proposed a software architec-

---

*Current affiliation is ATR Spoken Language Translation Research Laboratories
[†]Current mail addresses are
{*setsuo.yamada, eiichiro.sumita, hideki.kashioka*} *@slt.atr.co.jp*

ture that uses "a pragmatic adaptation" (LuperFoy and others, 1998), and Mima et al. proposed a method that uses "situational information" (Mima and others, 1997). LuperFoy et al. simulated their method on man-machine interfaces and Mima et al. preliminarily evaluated their method. Neither study, however, applied its proposals to an actual dialogue translation system.

The above mentioned methods will need time to work in practice, since it is *hard to obtain the extra-linguistic information on which they depend.*

We have been paying special attention to "politeness," because a lack of politeness can interfere with a smooth conversation between two participants, such as a clerk and a customer. It is easy for a dialogue translation system to know which participant is the clerk and which is the customer from the interface (such as the wires to the microphones).

This paper describes a method of "politeness" selection according to a participant's social role (a clerk or a customer), which is easily obtained from the extra-linguistic environment. We incorporated each participant's social role into transfer rules and transfer dictionary entries. We then conducted an experiment with 23 unseen dialogues (344 utterances). Our method achieved a recall of 65% and a precision of 86%. These rates could be improved to 86% and 96%, respectively (see Section 4). It is therefore possible to use a "participant's social role" (a clerk or a customer in this case) to appropriately make the translation results "polite," and to make the conversation proceed smoothly with a dialogue translation system.

Section 2 analyzes the relationship between a particular participant's social role (a clerk) and politeness in Japanese. Section 3 describes our proposal in detail using an English-to-Japanese

37

translation system. Section 4 shows an experiment and results, followed by a discussion in Section 5. Finally, Section 6 concludes this paper.

2 A Participant's Social Role and Politeness

This section focuses on one participant's social role. We investigated Japanese outputs of a dialogue translation system to see how many utterances should be polite expressions in a current translation system for travel arrangement. We input 1,409 clerk utterances into a Transfer Driven Machine Translation system (Sumita and others, 1999) (TDMT for short). The inputs were closed utterances, meaning the system already knew the utterances, enabling the utterances to be transferred at a good quality. Therefore, we used closed utterances as the inputs to avoid translation errors.

As a result, it was shown that about 70% (952) of all utterances should be improved to use polite expressions. This result shows that a current translation system is not enough to make a conversation smoothly. Not surprisingly, if all expressions were polite, some Japanese speakers would feel insulted. Therefore, Japanese speakers do not have to use polite expression in all utterances.

We classified the investigated data into different types of English expressions for Japanese politeness, i.e., into honorific titles, parts of speech such as verbs, and canned phrases, as shown in Table 1; however, not all types appeared in the data. For example, when the clerk said "How will you be paying, Mr. Suzuki," the Japanese translation was made polite as "donoyouni oshiharaininarimasu-ka suzuki-sama" in place of the standard expression "donoyouni shiharaimasu-ka suzuki-san."

Table 1 shows that there is a difference in how expressions should be made more polite according to the type, and that many polite expressions can be translated by using only local information, i.e., transfer rules and dictionary entries. In the next section, we describe how to incorporate the information on dialogue participants, such as roles and genders, into transfer rules and dictionary entries in a dialogue translation system.

3 A Method of Using Information on Dialogue Participants

This section describes how to use information on dialogue participants, such as participants' social roles and genders. First, we describe TDMT, which we also used in our experiment. Second, we mention how to modify transfer rules and transfer dictionary entries according to information on dialogue participants.

3.1 Transfer Driven Machine Translation

TDMT uses bottom-up left-to-right chart parsing with transfer rules as shown in Figure 1. The parsing determines the best structure and best transferred result locally by performing structural disambiguation using semantic distance calculations, in parallel with the derivation of possible structures. The semantic distance is defined by a thesaurus.

```
(source pattern)
⇒
((target pattern 1)
 ((source example 1)
  (source example 2)
 ... )
 (target pattern 2)
 ... )
```
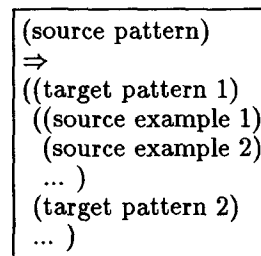
Figure 1: Transfer rule format

A transfer rule consists of a source pattern, a target pattern, and a source example. The source pattern consists of variables and constituent boundaries (Furuse and Iida, 1996). A constituent boundary is either a functional word or the part-of-speech of a left constituent's last word and the part-of-speech of a right constituent's first word. In Example (1), the constituent boundary $\langle V\text{--}CN \rangle$ is inserted between "accept" and "payment," because "accept" is a Verb and "payment" is a Common Noun. The target pattern consists of variables that correspond to variables in the source pattern and words of the target language. The source example consists of words that come from utterances referred to when a person creates transfer rules (we call such utterances closed utterances).

Figure 2 shows a transfer rule whose source pattern is (X $\langle V\text{--}CN \rangle$ Y). Variable X corresponds to x, which is used in the target pattern, and Y corresponds to y, which is also

Table 1: Examples of polite expressions

Type:	**verb, title**		
Eng:	How will you be <u>paying, Mr</u>. Suzuki		
Standard:	*donoyouni*	*shiharaimasu-ka*	*suzuki-san*
Polite:	*donoyouni*	*oshiharaininarimasu-ka*	*suzuki-sama*
Gloss:	How	pay-QUESTION	suzuki-Mr.

Type:	**verb, common noun**		
Eng:	We <u>have</u> two types of <u>rooms</u> available		
Standard:	*aiteiru*	*ni-shurui-no*	*heya-ga*	*arimasu*
Polite:	*aiteiru*	*ni-shurui-no*	*oheya-ga*	*gozaimasu*
Gloss:	available	two-types-of	room-TOP	have

Type:	**auxiliary verb**		
Eng:	You <u>can</u> shop for hours		
Standard:	*suujikan*	*kaimono-wo*	*surukotogadekimasu*
Polite:	*suujikan*	*kaimono-wo*	*shiteitadakemasu*
Gloss:	for hours	make-OBJ	can

Type:	**pronoun**		
Eng:	<u>Your</u> room number, please		
Standard:	*anatano*	*heya*	*bangou-wo*	*onegaishimasu*
Polite:	*okyakusamano*	*heya*	*bangou-wo*	*onegaishimasu*
Gloss:	Your	room	number-sc obj	please

Type:	**canned phrase**		
Eng:	How can I help you	
Standard:	*dou*	*shimashitaka*
Polite:	*douitta*	*goyoukendeshouka*
Gloss:	How	can I help you

Example (1)

Eng:	We accept payment by credit card			
Standard:	*watashitachi-wa*	*kurejitto-kaado-deno*	*shiharai-wo*	*uketsukemasu*
Polite:	*watashidomo-wa*	*kurejitto-kaado-deno*	*oshiharai-wo*	*oukeshimasu*
Gloss:	We-TOP	credit-card-by	payment-OBJ	accept

---

used in the target pattern. The source example (("accept") ("payment")) comes from Example (1), and the other source examples come from the other closed utterances. This transfer rule means that if the source pattern is (X $\langle V-CN \rangle$ Y) then (y "wo" x) or (y "ni" x) is selected as the target pattern, where an input word pair corresponding to X and Y is *semantically the most similar in a thesaurus* to, or *exactly the same* as, the source example. For example, if an input word pair corresponding to X and Y is *semantically the most similar in a thesaurus* to, or *exactly the same* as, (("accept") ("payment")), then the target pattern (y "wo" x) is selected in Figure 2. As a result, an appropriate target pattern is selected.

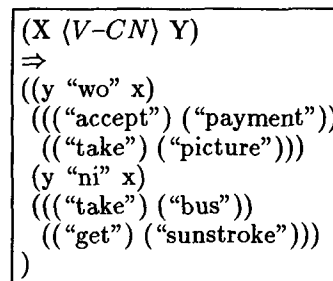After a target pattern is selected, TDMT creates a target structure according to the pattern

```
(X ⟨V-CN⟩ Y)
⇒
((y "wo" x)
 ((("accept") ("payment"))
  (("take") ("picture"))))
 (y "ni" x)
 ((("take") ("bus"))
  (("get") ("sunstroke"))))
)
```

Figure 2: Transfer rule example

by referring to a transfer dictionary, as shown in Figure 3. If the input is "accept $\langle V-CN \rangle$ payment," then this part is translated into "shiharai wo uketsukeru." "wo" is derived from the target pattern (y "wo" x), and "shiharai" and "uketsukeru" are derived from the transfer dictionary, as shown in Figure 4.

```
(source pattern)
⇒
(((target pattern 1₁) :pattern-cond 1₁
  (target pattern 1₂) :pattern-cond 1₂
  ...
  (target pattern 1ₙ) :default)
 ((source example 1)
  ... )
 (((source example 1) → (target word 1₁) :word-cond 1₁
   (source example 1) → (target word 1₂) :word-cond 1₂
   ...
   (source example 1) → (target word 1ₘ) :default)
  ... )
 (((target pattern 2₁) :pattern-cond 2₁
  ... )))
```

Figure 5: Transfer rule format with information on dialogue participants

```
(((source word 1) → (target word 1₁) :cond 1₁
  (source word 1) → (target word 1₂) :cond 1₂
  ...
  (source word 1) → (target word 1ₖ) :default)
 ... )
```

Figure 6: Dictionary format with information on dialogue participants

```
((source word) → (target word)
 ... )
```

Figure 3: Transfer dictionary format

```
(("accept") → ("uketsukeru")
 ("payment") → ("shiharai"))
```

Figure 4: Transfer dictionary example

```
(X "sama")
⇒
((("Mr." x) :h-gender male
  ("Ms." x) :h-gender female
  ("Mr-ms." x))
 ((("case"))
  (("room number")))
)
```

Figure 7: Transfer rule example with the participant's gender

3.2 Transfer Rules and Entries according to Information on Dialogue Participants

For this research, we modified the transfer rules and the transfer dictionary entries, as shown in Figures 5 and 6. In Figure 5, the target pattern

"target pattern 1₁" and the source word "source example 1" are used to change the translation according to information on dialogue participants. For example, if ":pattern-cond 1₁" is defined as ":h-gender male" as shown in Figure 7, then "target pattern 1₁" is selected when the hearer is a male, that is, "("Mr." x)" is selected. Moreover, if ":word-cond 1₁" is defined as ":s-role clerk" as shown in Figure 8, then "source example 1" is translated into "target word 1₁" when the speaker is a clerk, that is, "accept" is translated into "oukesuru." Translations such as "target word 1₁" are valid only in the source pattern; that is, a source example might not always be translated into one of these target words. If we always want to produce translations according to information on dialogue participants, then we need to modify the entries in the transfer dictionary like Figure 6 shows. Conversely, if we do not want to always change the translation, then we should not modify the entries but modify the transfer rules. Several conditions can also be given to ":word-cond" and ":pattern-cond." For example, ":s-role customer and :s-gender female," which means the speaker is a customer and a female, can be given. In Figure 5, ":default" means the de-

40

fault target pattern or word if no condition is matched. The condition is checked from up to down in order; that is, first, ":pattern-cond 1_1," second, ":pattern-cond 1_2," ... and so on.

```
(X ⟨V-CN⟩ Y)
⇒
((y "wo" x)
 ((("accept") ("payment"))
  (("take") ("picture")))
 ((("accept") → ("oukesuru") :s-role clerk
   ("accept") → ("uketsukeru")))
)
```

Figure 8: Transfer rule example with a participant's role

```
((("payment") → ("oshiharai") :s-role clerk
  ("payment") → ("shiharai"))
 (("we") → ("watashidomo") :s-role clerk
  ("we") → ("watashitachi")))
```

Figure 9: Transfer dictionary example with a speaker's role

Even though we do not have rules and entries for pattern conditions and word conditions according to another participant's information, such as ":s-role customer"(which means the speaker's role is a customer) and ":s-gender male" (which means the speaker's gender is male), TDMT can translate expressions corresponding to this information too. For example, "Very good, please let me confirm them" will be translated into "shouchiitashimasita kakunin sasete itadakimasu" when the speaker is a clerk or "soredekekkoudesu kakunin sasete kudasai" when the speaker is a customer, as shown in Example (2).

By making a rule and an entry like the examples shown in Figures 8 and 9, the utterance of Example (1) will be translated into "watashidomo wa kurejitto kaado deno oshiharai wo oukeshimasu" when the speaker is a clerk.

4 An Experiment

The TDMT system for English-to-Japanese at the time of the experiment had about 1,500 transfer rules and 8,000 transfer dictionary entries. In other words, this TDMT system was capable of translating 8,000 English words into Japanese words. About 300 transfer rules and 40 transfer dictionary entries were modified to improve the level of "politeness."

We conducted an experiment using the transfer rules and transfer dictionary for a clerk with 23 unseen dialogues (344 utterances). Our input was off-line, i.e., a transcription of dialogues, which was encoded with the participant's social role. In the on-line situation, our system can not infer whether the participant's social role is a clerk or a customer, but can instead determine the role without error from the interface (such as a microphone or a button).

In order to evaluate the experiment, we classified the Japanese translation results obtained for the 23 unseen dialogues (199 utterances from a clerk, and 145 utterances from a customer, making 344 utterances in total) into two types: expressions that had to be changed to more polite expressions, and expressions that did not. Table 2 shows the number of utterances that included an expression which had to be changed into a more polite one (indicated by "Yes") and those that did not (indicated by "No"). We neglected 74 utterances whose translations were too poor to judge whether to assign a "Yes" or "No."

Table 2: The number of utterances to be changed or not

Necessity of change	The number of utterances
Yes	104
No	166
Out of scope *	74
Total	344

* 74 translations were too poor to handle for the "politeness" problem, and so they are ignored in this paper.

The translation results were evaluated to see whether the impressions of the translated results were improved or not with/without modification for the clerk from the viewpoint of "politeness." Table 3 shows the impressions obtained according to the necessity of change shown in Table 2.

The evaluation criteria are recall and precision, which are defined as follows:

$$\text{Recall} = \frac{\text{number of utterances whose impression is better}}{\text{number of utterances which should be more polite}}$$

41

```
Example (2)

Eng:        Very good, please let me confirm them
Standard:   wakarimasita         kakunin   sasete   kudasai
Clerk:      shouchiitashimasita  kakunin   sasete   itadakimasu
Customer:   soredekekkoudesu     kakunin   sasete   kudasai
Gloss:      very good            confirm   let me   please
```

Table 3: Evaluation on using the speaker's role

Necessity of change	Impression	The number of utterances
Yes (104)	better	68
	same	5
	worse	3
	no-diff	28
No (166)	better	0
	same	3
	worse	0
	no-diff	163

better: Impression of a translation is better.

same: Impression of a translation has not changed.

worse: Impression of a translation is worse.

no-diff: There is no difference between the two translations.

Precision =

$$\frac{\text{number of utterances whose impression is better}}{\text{number of utterances whose expression has been changed by the modified rules and entries}}$$

The recall was 65% (= 68 ÷ (68 + 5 + 3 + 28)) and the precision was 86% (= 68 ÷ (68 + 5 + 3 + 0 + 3 + 0)).

There are two main reasons which bring down these rates. One reason is that TDMT does not know who or what the agent of the action in the utterance is; agents are also needed to select polite expressions. The other reason is that there are not enough rules and transfer dictionary entries for the clerk.

It is easier to take care of the latter problem than the former problem. If we resolve the latter problem, that is, if we expand the transfer rules and the transfer dictionary entries according to the "participant's social role" (a clerk and a customer), then the recall rate and the precision rate can be improved (to 86% and 96%, respectively, as we have found). As a result, we can say that our method is effective for smooth conversation with a dialogue translation system.

5 Discussion

In general, extra-linguistic information is hard to obtain. However, some extra-linguistic information can be easily obtained:

(1) One piece of information is the participant's social role, which can be obtained from the interface such as the microphone used. It was proven that a clerk and customer as the social roles of participants are useful for translation into Japanese. However, more research is required on another participant's social role.

(2) Another piece of information is the participant's gender, which can be obtained by a speech recognizer with high accuracy (Takezawa and others, 1998; Naito and others, 1998). We have considered how expressions can be useful by using the hearer's gender for Japanese-to-English translation.

Let us consider the Japanese honorific title "sama" or "san." If the hearer's gender is male, then it should be translated "Mr." and if the hearer's gender is female, then it should be translated "Ms." as shown in Figure 7. Additionally, the participant's gender is useful for translating typical expressions for males or females. For example, Japanese "wa" is often attached at the end of the utterance by females.

It is also important for a dialogue translation system to use extra-linguistic information which the system can obtain easily, in order to make a conversation proceed smoothly and comfortably for humans using the translation system. We expect that other pieces of usable information can be easily obtained in the future. For example, age might be obtained from a cellular telephone if it were always carried by the same person and provided with personal information. In this case, if the system knew the hearer was a child, it could change complex expressions into easier ones.

6 Conclusion

We have proposed a method of translation using information on dialogue participants, which

42

is easily obtained from outside the translation component, and applied it to a dialogue translation system for travel arrangement. This method can select a polite expression for an utterance according to the "participant's social role," which is easily determined by the interface (such as the wires to the microphones). For example, if the microphone is for the clerk (the speaker is a clerk), then the dialogue translation system can select a more polite expression.

In an English-to-Japanese translation system, we added additional transfer rules and transfer dictionary entries for the clerk to be more polite than the customer. Then, we conducted an experiment with 23 unseen dialogues (344 utterances). We evaluated the translation results to see whether the impressions of the results improved or not. Our method achieved a recall of 65% and a precision of 86%. These rates could easily be improved to 86% and 96%, respectively. Therefore, we can say that our method is effective for smooth conversation with a dialogue translation system.

Our proposal has a limitation in that if the system does not know who or what the agent of an action in an utterance is, it cannot appropriately select a polite expression. We are considering ways to enable identification of the agent of an action in an utterance and to expand the current framework to improve the level of politeness even more. In addition, we intend to apply other extra-linguistic information to a dialogue translation system.

References

Thomas Bub et al. 1997. Verbmobil: The combination of deep and shallow processing for spontaneous speech translation. In *the 1997 International Conference on Acoustics, Speech, and Signal Processing: ICASSP 97*, pages 71–74, Munich.

Osamu Furuse and Hitoshi Iida. 1996. Incremental translation utilizing constituent boundary patterns. In *Proceedings of COLING-96*, pages 412–417, Copenhagen.

Keiko Horiguchi. 1997. Towards translating spoken language pragmatics in an analogical framework. In *Proceedings of ACL/EACL-97 workshop on Spoken Language Translation*, pages 16–23, Madrid.

Akira Kurematsu and Tsuyoshi Morimoto. 1996. *Automatic Speech Translation*. Gordon and Breach Publishers.

Susann LuperFoy et al. 1998. An architecture for dialogue management, context tracking, and pragmatic adaptation in spoken dialogue system. In *Proceedings of COLING-ACL'98*, pages 794–801, Montreal.

Hideki Mima et al. 1997. A situation-based approach to spoken dialogue translation between different social roles. In *Proceedings of TMI-97*, pages 176–183, Santa Fe.

Masaki Naito et al. 1998. Acoustic and language model for speech translation system ATR-MATRIX. In *the Proceedings of the 1998 Spring Meeting of the Acoustical Society of Japan*, pages 159–160 (in Japanese).

Manny Rayner and David Carter. 1997. Hybrid language processing in the spoken language translator. In *the 1997 International Conference on Acoustics, Speech, and Signal Processing: ICASSP 97*, pages 107–110, Munich.

Carolyn Penstein Rosé and Lori S. Levin. 1998. An interactive domain independent approach to robust dialogue interpretation. In *Proceedings of COLING-ACL'98*, pages 1129–1135, Montreal.

Eiichiro Sumita et al. 1999. Solutions to problems inherent in spoken-language translation: The ATR-MATRIX approach. In *the Machine Translation Summit VII*, pages 229–235, Singapore.

Toshiyuki Takezawa et al. 1998. A Japanese-to-English speech translation system: ATR-MATRIX. In *the 5th International Conference On Spoken Language Processing: ICSLP-98*, pages 2779–2782, Sydney.

Enrique Vidal. 1997. Finite-state speech-to-speech translation. In *the 1997 International Conference on Acoustics, Speech, and Signal Processing: ICASSP 97*, pages 111–114, Munich.

Jae-Woo Yang and Jun Park. 1997. An experiment on Korean-to-English and Korean-to-Japanese spoken language translation. In *the 1997 International Conference on Acoustics, Speech, and Signal Processing: ICASSP 97*, pages 87–90, Munich.