A Situation-based Approach to Spoken Dialog Translation Between Different Social Roles

Hideki Mima, Osamu Furuse and Hitoshi Iida

ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02, Japan
{mima, furuse, iida}@itl.atr.co.jp

Abstract. This paper presents a situation-based approach to spoken dialog translation between speakers with different social roles. In evaluating what constitutes natural communications by languages, particular attention was paid to the social standings of speakers in situations. By taking their social roles into account, we are able to discuss what kinds of situational information are useful for creating naturally communicative dialogs and show one type of information that can resolve contextual ambiguities with respect to zero-subject resolution, word usage selection, and so on. We then propose a method of improving the precision of translation by utilizing this information. Through preliminary experimentation showing performance improvements in English to Japanese translation, the proposed scheme is expected to improve the performance of spoken dialog MT.

1 Motivation

In an attempt to develop a speech-to-speech dialog system, several groups have recently tried to understand spoken dialogs [Goddeau et al., 1994][Levin et al., 1995][Wahlster et al., 1993] [Rayner et al., 1994]. Systems that deal with spoken dialogs generally require different techniques than systems that deal with written languages. The main requirements for the former are techniques to handle 1) spoken languages containing ungrammatical expressions, 2) real-time translation to avoid interrupting smooth communications [Sumita et al., 1995], and 3) appropriate expressions under environmentally influenced situations[1].

Transfer-Driven Machine Translation (TDMT) [Furuse et al., 1995] has been proposed and is one of the efficient methods for spoken dialog translation. In TDMT, constituent boundary patterns [Furuse and Iida, 1996] are applied to an input incrementally; this contrasts with the linguistic manner of applying grammar rules. The result provides for robust parsing that can even handle ungrammatical phenomena such as derivation in metonymical relationships. Additionally, by dealing with best-only substructures utilizing translation examples, the explosion of structural ambiguities is significantly constrained. Accordingly, robust and efficient translation of a spoken-language input can be achieved.

However, developing a general solution for dealing with situations such as context understanding still remains one of the most difficult problems. Most conventional approaches to contextual understanding have tried to determine the goal or the reference within a sentence [Levin et al., 1995], i.e, the relationship between what the speaker is referring to and the entity existing in the real world. Although this requires various kinds of possibilities to be inferred from the context, most of the items ultimately are eliminated despite the cost of computing. Additionally, to handle the euphemistic expressions[2] required by spoken-language translation, the selection of such euphemistic expressions depends solely on the situational influences. A conventional MT system provides a large-scale dictionary, but this dictionary gives little information on word usage under appropriate situations or context. The analysis process works

[1] In this paper, we generally assume that a situation includes the environment and context.

[2] We define euphemistic expressions as expressions said when paying one's respects to the hearer indirectly, such as by offering superficial choices to the hearer even if the real intention is a command. For example, it is more polite to say “I would appreciate it if you could help me” rather than saying “Will you help me”, when you ask something of your superior, in general.

only for individual sentences, separating them from the context without considering information about their individual situations.

Nevertheless, there are several cases which can make this problem simple to solve when some situational information is accessed. Thus, in this paper, we will primarily discuss situational information (including the social roles, that is social location, social rank, gender, and so on, of the speakers) that the system can use to naturalize communications by contextual disambiguation with respect to zero-subject resolution, word usage selection, and so on. We will also propose a method to improve the translation accuracy by utilizing such information.

From preliminary experimentation of a new TDMT system, the proposed scheme is also expected to improve the performance of spoken dialog MT.

The next section of this paper explains situation-based translation together with some examples. Section 3 describes our proposed method for improving the translation accuracy. Section 4 presents a performance improvement of the TDMT prototype system[3] through experimental results using about 1000 unseen input sentences. In section 5, we state our conclusions.

2 Situation-based Approach for Dialog Translation

2.1 Situational Information

As we have already mentioned, the treatment of polite remarks such as in the selection of euphemistic expressions generally depends mostly on environmental influences. For example, Japanese donatory auxiliary verbs such as *'ageru'* have multiple possible expressions depending on the speaker's honorific attitude to the hearer[4]. Maeda et al. presented a unification based approach to Japanese honorifics [Maeda et al., 1988]. Similarly, the translation of set phrases, such as *'shitsurei-itashi-masu',* which in Japanese has multiple possible translations like 'hello', 'good bye' and 'excuse me'[5], is determined through the context or situation.

Any contextual constraint should be described in a certain domain, that is, in a sublanguage, and the real utterance should be expressed based on the appropriate usage under every situation (Fig. 1). Consequently, a spoken dialog translation system must work in situations where the speaker properties (role, gender, and rank), local topics, focus in the utterances, dialog domains, objects mentioned, actions mentioned, and derivational forms of the words are all understood.

In this section, a few examples show the typical information a dialog translation system can utilize to resolve contextual ambiguity with respect to zero-subject resolution, word usage, and so on.

2.2 Social Role and Dialog Domain

In the dialog domain of travel, a clerk in a hotel, as opposed to a traveler, is usually not dining, boarding or using a tourist-oriented transportation system. Such information can constrain the possibilities of word selection with respect to polite remarks. For example, the word 'eat' can be translated into Japanese as *'taberu' 'meshiagaru*-honorific' or, *'itadaku*-humble'. However, if the dialog domain is travel, and the speaker is a clerk and not a traveler, 'eat' is never translated into *'itadaku*-humble'. In the same way, we can consider the translation of 'your' in the following sentence:

(1) "Could you please tell me your telephone number?"

If the speaker's social role is a clerk and the listener's is a guest, "your telephone number" should be translated as *"o-kyaku-sama-no o-denwa-bango"*(guest-polite-possessive telephone-number-polite)[6].

[3] Since TDMT and its prototype system have already been presented such as in [Furuse et al., 1995] [Furuse and Iida, 1996], we do not describe the mechanism here.

[4] The complexity of word usage selection for verbs of Giving and Receiving is presented on the web; *http://central.itp.berkeley.edu/~eal/Jpnotes/donatory_verbs.html.*

[5] For example, when entering someone else's house and when leaving the house, *'shitsurei-itashi-masu'* is used in Japanese for both cases.

[6] In this paper, sample Japanese sentences are Romanized in italic based on the Hepburn system, and the corresponding English words with usage modifiers follow in parentheses.

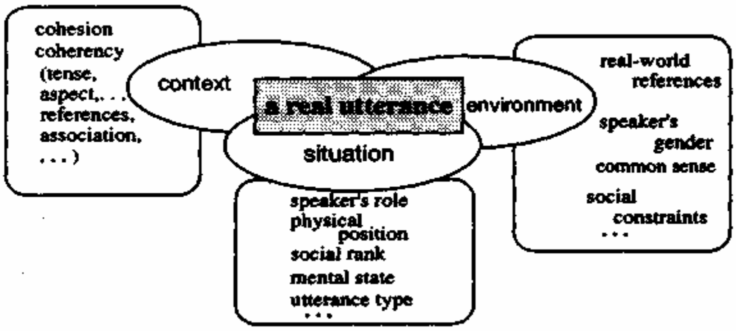

Figure 1. Determination of an utterance in a situation

In addition, it is well-known that the subject tends to be omitted in spoken Japanese. Thus, having this knowledge would clearly be helpful in predicting the English for missing Japanese subjects. For example, if a cook says,

(2) *"Moshi o-meshiagaru no deshitara, sugu tsukuri-masu"*
(if eat-honorific, immediately cook-polite)

it is appropriate to translate this by supplying "you" as the subject of "eat" and "I" as the subject of "cook".

2.3 Gender

There are several differences between male expressions and female expressions in Japanese. For example, men can use *'boku'* or *'ore'* to mean "I". However, women never use these words in standard Japanese. It has also been reported that women usually use euphemistic expressions more frequently than do men, and likewise women tend to use softer expressions than do men. Therefore, information on whether the speaker/listener is male or female allows the translation system not only to assign the correct honorific title, such as Mr. or Ms., but also enables the translation to preserve expressions appropriate to the speaker's gender.

2.4 Social Rank

There are various types of expressions: polite, honorific, humble, and euphemistic. These expressions are used depending on the social interactions between speakers[7]. For example, in making a hotel reservation,

(3) "How many people will there be?"

'will there be' can be translated into Japanese in at least two forms, *'i-masu-ka'* or *'irasshai-masu-ka'*. If the listener in this case is at a socially higher rank than the speaker (e.g., the relationship between a guest and a clerk), the latter expression would be preferred as it bears more esteem. In the same way,

(4) "Then, we will confirm your reservation."

should be translated into:

(4') *"Yoyaku-o kakunin-sasete-itadaki-masu"*
('reservation-objective, 'confirm-euphemism')

by using the euphemistic alternative.

[7] Although this type of word usage selection seems a little bit Japanese (or Korean) oriented, the authors believe that handling this kind of difference between languages is very important for smooth dialog translation such as in the case of diplomatic meetings. In fact, some sentences were judged as incorrect translations only because of their lack of adequate politeness, in our experimental evaluation with an EJ translation system.

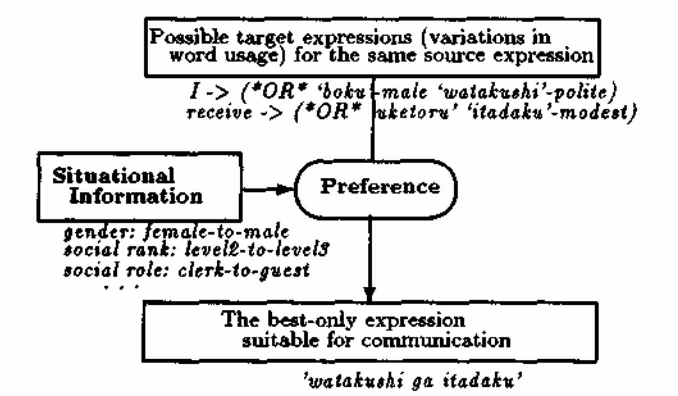
Possible target expressions (variations in word usage) for the same source expression
I -> (*OR* 'boku-male 'watakushi'-polite)
receive -> (*OR* uketoru' 'itadaku'-modest)
Situational Information
Preference
gender: female-to-male
social rank: level2-to-level3
social role: clerk-to-guest
. . .
The best-only expression suitable for communication
'watakushi ga itadaku'

Figure 2. Preference of the best-only expression.

EJ: *(X <noun-verb> Y)* =>
X' ga Y' ((I , receive) →
(*OR* ((*OR* '*boku*-male '*watashi'*-female ...*), 'uketoru')*
('watakushi-polite', '*itadaku*-polite-humble') ...), ...)

JE: *(moshi Y no deshitara)* =>
if *X' Y'* (('*taberu'*-speaker:person-in-service) →
((default you), eat), ...)

Figure 3. Examples of Transfer Knowledge with Situational Information.

3 Utilization of Situational Information in TDMT

In this section, we propose a scheme for utilizing situational information in spoken dialog MT systems to enable preferences appropriate for varying types of communications to be established while preserving the efficiency.

At first, to illustrate the utterance selection mechanism in relation to the situational information, we assume that the model is as simple as that shown in Fig. 2. For example, the word 'receive' is translated into the Japanese word *'itadaku'* if a more polite expression is required than the possible translation *'uketoru'*. However, the model should independently work for each reasonable unit within a sentence. This is because each object (e.g., a person who should be respected) of a predicate of a unit might differ depending on the subject. For instance, let us consider the sentence:

(5) "Could you tell me the telephone number where I can contact you please?"

Since the object in the unit phrase "I can contact you" must be the listener, an honorific expression is required, in contrast to the object in the unit phrase "Could you tell me ...", which must be the speaker.

Additionally, as we have already mentioned, in some cases, omitted subjects in Japanese sentences can be determined depending on the relation of the situation and the action by default. Similarly, there is a preference for linguistic co-occurrence for word usages within individual units in general. For example, if *'itadaku'* is used as the translation of 'receive', *'watakushi'* is supposed to be more suitable as the subject of the unit phrase in many cases. In order to achieve these word usage preferences and default handlings for each unit phrase in sentences, we propose a scheme that incrementally applies the model to each phrase transfer of a translation. In TDMT, constituent boundary patterns are applied to an input sentence, and by dealing with best-only substructures utilizing phrase level translation examples, a deterministic translation can be achieved for each pattern incrementally [Furuse and Iida, 1996]. Consequently, the design principles of our scheme are:

- Word usage examples or default subjects with a relation to situational parameters are attached to each transfer pattern (Fig. 3)

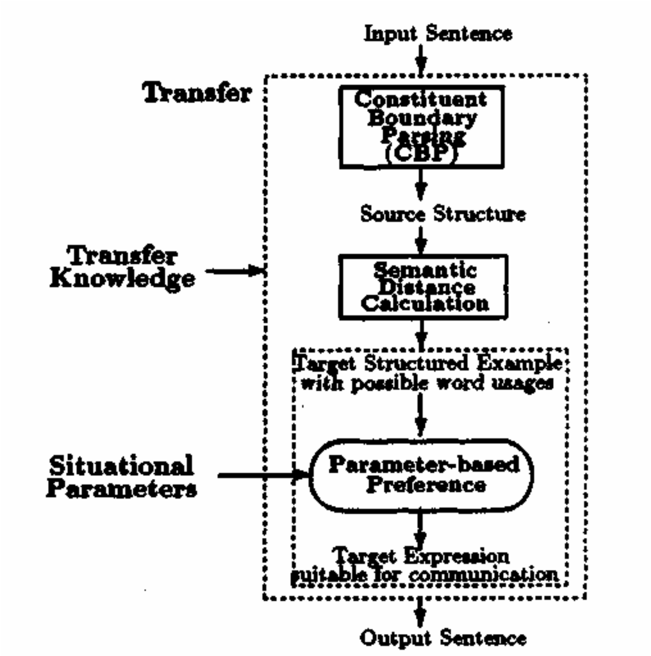
Input Sentence
Transfer
Constituent Boundary Parsing (CBP)
Source Structure
Transfer Knowledge
Semantic Distance Calculation
Target Structured Example with possible word usages
Situational Parameters
Parameter-based Preference
Target Expression suitable for communication
Output Sentence

Figure 4. Utilizing situational information in TDMT.

- The best-only phrase structure selection mechanism using examples and the word usage selection model with situational parameters are combined (Fig. 4)
- Word candidates for particular usage are selected simultaneously after each phrase transfer has settled

Furthermore, since this extension preserves the graduality for phrase chunk transfer in TDMT, it is reasonable to extend our method to simultaneous interpretation even for lengthy inputs in order to achieve real-time translation. By utilizing our scheme in TDMT, along with a situational parameter such as *:clerk-to-guest,* sample sentence (5) can be translated into more plausible Japanese:

(5') *"o-kyaku-sama-ni go-renraku-dekiru denwa-bangō-o uketamaware-masu-ka."*
('you-guest-role-objective' 'contact-polite-humble-possible' 'telephone-number-objective' 'tell-me-humble-polite')

4 Experimental Results

In this section, we describe the performance improvement achieved by our method from a comparison between a situation-based TDMT prototype system and the conventional TDMT system. Each TDMT system, whose domain is travel conversations, is implemented in LISP and runs on UNIX-based machines. Systems dealing with spoken-languages generally require quick and accurate responses, rather than grammatical responses, in order to provide smooth communications. Moreover, since every process, including speech recognition, translation and speech synthesis, runs automatically from start to finish, there is no room for manual pre/post-editing of the input/output sentences in order to make the sentences easier for either the translation process or the user to read. In other words, assuring acceptability in spoken-language translation is the most crucial task in devising such a system. Therefore, we evaluated both TDMTs for acceptability of the translation and analyzed the evaluation results from this point of view.

4.1 The Evaluation Procedure

We evaluated the TDMTs' translation quality by separately using morphological analysis and a translation module (including a generation module). Thus, manually analyzed morpheme

Table 1. Experimental conditions.

	JE	JK	JG	EJ	KJ
Vocabulary size	10000			6000	3000
No. of training sentences (diff.)	2602	1195	1553	2431	493
average morphemes/sentence	10.1	9.0	9.3	8.4	7.5
No. of patterns for transfer	887	624	787	1194	320
No. of examples for transfer	10227	3605	2941	7008	1701

Table 2. Evaluation results.

	JE	JK	JG	EJ	KJ
No. of test dialogs	69 (1247 sentences)			73(1323)	87(1169)
average of morphemes/sentence		9.4		7.1	8.0
(A) (%)	30.1	46.6	27.6	23.7	34.4
(B) (%)	18.1	29.1	11.4	17.2	17.5
(C) (%)	20.6	14.4	10.6	18.5	21.2
(D) (%)	31.2	9.9	50.4	40.5	26.9

sequences were used to avoid errors and unknown words in testing the translation module itself. This allowed us to assess how well the TDMTs would function individually. The details of the evaluations with a morphological analyzer are not described here[8].

Table 1 shows the conditions of our experiments and evaluations. The reader should note that as the JG and KJ translation projects have just started, the transfer knowledge should not be expected to be of the same quality for all language pairs. However, Sumita et al. [Sumita et al., 1991] did show that the translation correctness improves in proportion to the amount of translation examples[9] and also showed that the amount of transfer knowledge required for correct translation depends on the linguistic distance between the source and target languages [Furuse et al., 1995]. In addition, though the quality of the thesaurus for each language is an important topic for example-based frameworks, according to our experimental results obtained by applying some kinds of thesauruses into TDMT, no remarkable differences in the translation quality were observed except for a difference in the number of translation outputs.

The translation of each sentence for 69-87 dialogs (about 1,000 unseen different sentences) was manually evaluated by assigning a grade. Two or three native speakers of the target languages performed the assessments, where all of the examiners were also familiar with the source languages in order to judge the correctness of the information. We used the same dialogs for all of the translations whose source language was Japanese: i.e., JE, JK, and JG translations. This allowed us to compare the differences in the quality of the transfer knowledge (patterns, examples, etc), the linguistic distance between languages, and so on. Each sentence was assigned one of four grades for translation quality; (A) No problem - a fluent translation with all information conveyed correctly; (B) Fair - a translation that made it easy to understand the expressions but with some unimportant expressions missing grammatical elements; (C) Acceptable - an acceptable translation; (D) Nonsense, incorrect - an unacceptable translation or the incorrect translation of important information.

[8] The success rate of perceiving morphemes was more than 99%, and that of assigning linguistic categories was more than 98%.

[9] This is accepted in general for the example-based framework, since the exact match ratio is certain to increase in proportion to the increase in translation examples. In fact, a total accuracy of more than 93% was achieved in our closed test evaluation for over 1,000 sentences for EJ TDMT translation. However, we have to ascertain the satiation limit, i.e., how much the transfer knowledge can be expanded depending on the language pairs, in terms of the practicality.

Table 3. Breakdown of Changed Sentences.

Previous rank\New rank	(A)	(B)	(C)	(D)
(B) (No. of sentences)	9	4	0	0
(C) (No. of sentences)	15	11	4	0
(D) (No. of sentences))	3	3	43	23

4.2 Evaluation Results of Conventional TDMT

Table 2 shows evaluation results for the TDMT. All ratios were taken from the average of two or three examiners. As the table shows, almost 70% acceptability[10] was achieved in the JE result and almost 60% acceptability was achieved in the EJ result; remarkably, more than 90% acceptability was achieved in the JK translation, although the JK translation needed less transfer knowledge (Table 1) than the others. As we mentioned above, this observation can be explained from the viewpoint of linguistic similarity; while the Japanese-English (German) language pair is linguistically distant, the Japanese-Korean pair is rather close.

4.3 Preliminary Experiment with Situation-based Approach

We conducted a preliminary experiment to examine the performance of this scheme while using the current EJ TDMT system. We first tested principally to improve the politeness, with regard to the social role and social rank. Transfer patterns with a word usage dictionary for about 50 words were prepared for the experiment. In this preliminary experiment, we assigned the situational parameters manually to independently evaluate the scheme. In order to evaluate the effectiveness of TDMT while utilizing word usage preferences, we used 316 of the "clerk's" sentences in regard to hotel conversation dialogs previously graded (B)(C)(D)[11] in Table 2.

Translation examples of utterances from a clerk to a guest (modified by ") compared to identical utterances from a guest to a clerk (modified by '), are shown in the following.

(6) "We have a question for you."
(6') Guest-to-Clerk: *"shitsumon-ga ari-masu."*
('question' 'we-have-polite')
(6") Clerk-to-Guest: *"o-kyaku-sama-ni go-shitsumon-ga go-zai-masu."*
('you'-guest-role-polite-objectivel' 'question-polite-objective2' 'we-have-humble-polite')

The following results were obtained in the evaluation.

— 115 (36.4%) sentences were changed into other expressions.
— 84 (26.5%) sentences were improved by at least one grade.
— The rest of the changed sentences were still assigned the same grades because of different problems.

Table 3 shows a breakdown of the changed sentences. The table indicates that there were nine sentences obtaining an (A) grade from a (B) grade, and so on. By improving all results having a (D) grade in the hotel conversation dialogs (complete in 637 sentences), the total acceptability can be improved by about 8% using this scheme, indicating that these results are fairly good. However, one problem is the possibility of constructing overly polite expressions in Japanese. This basically derives from the fact that the current generation of software does not have sufficient knowledge yet to treat collocational constraints of politeness between patterns. Nevertheless, the problem can be easily overcome by using statistical/stochastic methods such as n-grams or the co-occurrence of phrases. Consequently, it is clear that the proposed scheme should be able to improve the performance of a spoken dialog MT system.

[10] In this evaluation, we assumed the acceptability of translation as the sum of the (A), (B) and (C) grade sentences.

[11] As we have already mentioned, some sentences were judged as of the (D) grade only because of their lack of adequate politeness.

5 Conclusions

Developing a general solution to deal with the context still remains one of the most difficult problems. We have discussed several types of situational information which can make this problem simple to solve.

For handling the situational information required for spoken dialog translation, we have discussed situational constraints on zero-subject resolution, word usage selection, and so on, that naturalize communicative dialogs. Furthermore, we have proposed a method to improve the accuracy of translations by utilizing this information. Preliminary experimentation of our new TDMT has shown that the proposed scheme can be expected to improve the performance of spoken dialog MT.

One important area of future research will be to develop a method capable of extracting the information from a situation efficiently. At least two schemes will be considered:

1) Information manually assigned in the system (as one of the system's properties)
There are some information types, such as a speaker's social role, which are considered to be more efficient when assigned manually in the system, if the system is settled in a static environment.
2) Information automatically assigned dynamically depending on the situation
Other information types, such as the various preferences for the social rank, gender and conversational domain, can potentially be found in the large corpora of linguistic expressions with some kind of stochastic information, such as the usage frequency or word usage n-gram in relation to these situational parameters.

Currently, though corpora including the situational parameters have not been provided completely yet in general, the automatic training of translation knowledge can also be expected by using corpora. Therefore, our proposed scheme using an example-based framework is considered to have the advantage of handling situations even when scaling up the system for practical use. One area of interest for future work is the automatic extraction of situational information from dialogs by utilizing statistical/stochastic approaches.

References

[Furuse et al., 1995] Osamu Furuse, Jun Kawai, Hitoshi Iida, Susumu Akamine and Deok-Bong Kim. 1995. Multi-lingual Spoken-Language Translation Utilizing Translation Examples. In *Proceedings of Natural Language Pacific Rim Symposium '95,* pages 544-549.

[Furuse and Iida, 1996] Osamu Furuse and Hitoshi Iida. 1996. Incremental Translation Utilizing Constituent Boundary Patterns. In *Proceedings of the 16th International Conference on Computational Linguistics,* Copenhagen, pages 412-417.

[Goddeau et al., 1994] David Goddeau, Eric Brill, James Glass, Christine Pao, Michael Phillips, Joseph Polifroni, Stephanie Seneff, and Victor Zue. 1994. Galaxy: A Human-Language Interface to On-Line Travel Information. In *Proceedings of International Conference on Spoken Language Processing '94,* Yokohama, Japan, pages 707-710.

[Levin et al., 1995] Lori Levin, Oren Glickman, Yan Qu, Donna Gates, Alon Lavie, Carolyn P. Rose, Carol Van Ess-Dykema and Alex Waibel. 1995. Using Context in Machine Translation of Spoken Language. In *Proceedings of Sixth International Conference on Theoretical and Methodological Issues in Machine Translation TMI 95,* Leuven, Belgium, pages 173-187.

[Maeda et al., 1988] Hiroyuki Maeda, Susumu Kato, Kiyoshi Kogure and Hitoshi Iida. 1988. Parsing Japanese Honorifics in Unification-based Grammar. In *Proceedings of S6th Annual Conference of the Association for Computational Linguistics,* Buffalo, New York, pages 139-146.

[Rayner et al., 1994] M. Rayner, H. Alshawi, I. Bretan, D. M. Carter, V. Digalakis, B. Gamback, J. Kaja, J. Karlgren, B. Lyberg, S. G. Pulman, P. Price, and C. Samuelsson. A Speech to Speech Translation System Built from Standard Components. Abstract of SRI Cambridge Technical Report CRC-031; *http: //www.cam.sri.com/tr/crc031/abstract.html.*

[Sumita et al., 1991] Eiichiro Sumita and Hitoshi Iida. 1991. Experiments and Prospects of Example-based Machine Translation. In *Proceedings of 29th Annual Conference of the Association for Computational Linguistics,* Berkeley, Calif., pages 185-192.

[Sumita et al., 1995] Eiichiro Sumita and Hitoshi Iida. 1995. Heterogeneous Computing for Example-Based Translation of Spoken Language. In *Proceedings of Sixth International Conference on Theoretical and Methodological Issues in Machine Translation TMI 95,* Leuven, Belgium, pages 273-285.

[Wahlster et al., 1993] Wolfgang Wahlster. 1993. Verbmobil: Translation of Face-To-Face Dialogs. In *Proceedings of the Fourth Machine Translation Summit: MT Summit IV,* Kobe, Japan, pages 127-135.