

Automatic Interpretation System Integrating Free-style Sentence Translation and Parallel Text Based Translation

Takahiro Ikeda Shinichi Ando Kenji Satoh Akitoshi Okumura Takao Watanabe
Multimedia Res. Labs. NEC Labs.
4-1-1 Miyazaki, Miyamae-ku, Kawasaki, Kanagawa 216
t-ikeda@di.jp.nec.com, s-ando@cw.jp.nec.com, k-satoh@da.jp.nec.com,
a-okumura@bx.jp.nec.com, t-watanabe@ay.jp.nec.com

Abstract

This paper proposes an automatic interpretation system that integrates free-style sentence translation and parallel text based translation. Free-style sentence translation accepts natural language sentences and translates them by machine translation. Parallel text based translation provides a proper translation for a sentence in the parallel text by referring to a corresponding translation of the sentence and supplements free-style sentence translation. We developed a prototype of an automatic interpretation system for Japanese overseas travelers with parallel text based translation using 9206 parallel bilingual sentences prepared in task-oriented manner. Evaluation results show that the parallel text based translation covers 72% of typical utterances for overseas travel and the user can easily find an appropriate sentence from a natural utterance for 64% of typical traveler's tasks. This indicates that the user can benefit from reliable translation based on parallel text for fundamental utterances necessary for overseas travel.

1 Introduction

A speech-to-speech translation system must integrate at least three components — speech recognition, machine translation, and speech synthesis. In practice, each component does not always output the correct result for various inputs, and an error in one component often leads to an incorrect result

being produced by the total system even for a limited domain. Clearly, we need ways to complement speech-to-speech translation systems that cannot reliably produce a correct result.

Although some robust methods that make the erroneous results of other components acceptable have been proposed (Yumi et al., 1997; Furuse et al., 1998), there is no guarantee that the final output from a system will be appropriate even with these methods. To deal with this problem, we have taken a more practical approach to developing an automatic interpretation system where the user can obtain a correct result instead of having to apply additional operations and judgment.

In actual use of a speech-to-speech translation system, an error in the speech-recognition or speech-synthesis components is not a large problem if the system has a screen that displays each result. The user of the system can correct errors in the recognition result on the screen, and can communicate by showing the other person the translated sentence on the screen.

On the other hand, an error in the machine-translation component is critical because a user who is not familiar with the target language is unlikely to notice the error in some cases. When a nonsensical sentence is generated by machine translation, the user may realize that the listener does not understand the translated sentence. However, when a plausible sentence that means something different from the intended meaning is generated by the machine translation, the user may incorrectly assume that the utterance was properly communicated. Consequently, the user can seldom be sure that the listener correctly understood the intended meaning when using a speech-to-speech translation system. A conversa-

tion could continue for some time before it became apparent that the two sides misunderstood what the other was saying.

Moreover, if the user realizes that there is an error in the machine translation, correcting it will be difficult. Without knowing the source of the error, the user cannot modify the input to obtain a correct result.

These error problems severely limit the usability of speech-to-speech translation.

In this paper, we propose an automatic interpretation system that integrates free-style sentence translation and parallel text based translation. In this system, free-style sentence translation accepts natural language sentences and translates them by machine translation without guaranteeing the quality of the translation. On the other hand, parallel text based translation uses parallel bilingual sentences registered in the system and translates a registered sentence by referring to the corresponding translation. Although this translation process limits the input to registered sentences, it is a robust means of handling input with recognition errors and consistently provides a correct translation. We integrated these two types of translation to realize a robust translation system where the two types of translation compensate for the shortcomings of each other.

For appropriate integration of free-style sentence translation and parallel text based translation, we had to consider three main points.

1. User interface: how best to present the two functions to the user?
2. Content of registered sentences: How many utterances should be covered by registered sentences?
3. Retrieval system: What methods of searching among the registered sentences should be provided to the user?

In this paper, we discuss these three points with respect to a translation system for Japanese travelers in the overseas travel domain. We construct a model of the integration of free-style sentence translation and parallel text based translation in Section 2. We describe a prototype system based on the model in Section 3 and evaluate it in Section 4. Related work on translation systems utilizing parallel text are discussed in Section 5, and we conclude in Section 6.

2 The Integration Model

2.1 User Interface

Although parallel text based translation provides a correct result, the registered parallel bilingual sentences cannot cover all possible utterances by the user in the target domain. Free-style sentence translation, on the contrary, accepts free-style input sentences but provides no guarantee as to the quality of results.

For many routine situations, users will clearly benefit from using parallel text based translation. In such cases, the system will probably include a sentence that totally or partially fits what they want to say. To ensure high translation reliability, users should use free-style sentence translation only for utterances not covered by the registered sentences.

However, users usually will not know what sentences are registered in the system and will have to search for an appropriate sentence before they can use parallel text based translation. In some cases, the user will be forced to use free-style sentence translation if unable to find an appropriate sentence.

A seamless user interface that allows the user to easily switch between free-style sentence translation and parallel text based translation is therefore needed in a system integrating these two forms of translation. Two conditions in particular had to be met to make the system easy to use.

1. The user should be able to use an input sentence seamlessly as both a source sentence for free-style sentence translation and a key sentence for registered sentence retrieval.
2. The user should be able to use each sentence included in the results of the registered sentence retrieval and the input sentence as a source sentence for translation. (The former would be used for parallel text based translation, and the latter would be used for free-style sentence translation.)

2.2 Content of Registered Sentences

Registered sentences must cover the utterances necessary for accomplishing typical tasks in the target domain to provide correct translation for minimal communication. In a translation system for overseas travelers, some typical tasks are changing money, checking in at a hotel, and ordering at a restaurant.

We adopted a three-tier model that consists of scenes, tasks, and subtasks to prepare a sufficient set

Table 1: Examples of scenes, tasks, subtasks, and templates of sentences

Scene	Task	Subtask	Template of sentence
Hotel	Check-in	Checking in	I'd like to check in, please.
Hotel	Check-in	Requesting a type of room	I'd like a room with <u>the ocean view</u> .
Restaurant	Order	Requesting cooking time for your steak	<u>Medium</u> , please.
Restaurant	Order	Asking what they recommend	What do you recommend for <u>appetizers</u> ?

of necessary sentences to be registered in the system. A scene comprised a place or situation that corresponds to where a traveler is likely to be (e.g., a hotel) and a problem that could arise. We made a list of typical travelers’ tasks that would be necessary in various travel scenes, divided each task into smaller primitive tasks (subtasks), and assigned a sentence template to each subtask based on the model.

In general, more than one round of conversation is necessary to accomplish each task. We assumed that a task would consist of smaller subtasks, each of which would correspond to one round of conversation that consisted simply of an utterance from a traveler to a respondent and a response from the respondent to the traveler. For example, the task of checking in to a hotel consists of subtasks such as giving your name, confirming your departure date, and so on. Each subtask should be the smallest unit of a task because users cannot use a registered sentence effectively if it includes more than what they want to say.

In this way, only one sentence template is needed for each subtask with regard to an utterance from a traveler to a respondent. For example, we can assign a sentence template of “I’d like to have ...” to the subtask of ordering a dish in a restaurant. We can provide a sufficient number of sentences by enabling the user to fill in the part denoted as “...” (referred to as a slot) with words applicable to the situation.

Table 1 shows examples of scenes, tasks, subtasks, and sentence templates. An underlined part represents a slot. We define a list of words individually for each slot.

For each task, both the utterances from a traveler to a respondent and the responses from a respondent to a traveler are significant. Responses should also be supported by parallel text based translation to ensure reliable communication. However, inputting the response and retrieving a registered sentence that matches it will be difficult and time consuming for the respondent who is unlikely to be familiar with the translation system.

We use a system that presents a menu of responses

for the respondent to choose from. The system keeps typical responses in parallel bilingual form for each registered sentence that the traveler can use and displays these as candidate responses when the traveler uses the sentence. The system then shows the traveler the translation of the response selected by the respondent.

This approach enables travelers to obtain a reliable response and also enables respondents to easily select an appropriate response.

2.3 Retrieval System

The retrieval system to search for a registered sentence that we use is based on a combination of three conditions — the natural language sentence, scene, and action.

Registered sentence retrieval based on a natural language sentence is essential for seamless integration of free-style sentence translation and parallel text based translation. We used a simple keyword-based retrieval system for registered sentence retrieval. This system extracts keywords from an inputted natural language sentence, searches for sentences including the keywords, and presents the results ranked mainly by the number of keywords included in each sentence.

The system retrieves all sentences including more than one keyword to reduce the chance of an appropriate sentence not being retrieved. We overcame the increased retrieval noise in the result by applying an additional retrieval system to search for registered sentences in terms of the scene and action.

Each registered sentence to be retrieved for translation corresponds to a set of a scene, a task, and a subtask as described in the previous section. A scene represents a place or a situation where the user wishes to accomplish the task and the subtask. A task and a subtask represent a user’s actions. This means that the user’s utterance is related to the user’s intention regarding where (scene) the user wants to do something (action).

We use the additional retrieval system in situations where the user has to search for sentences from

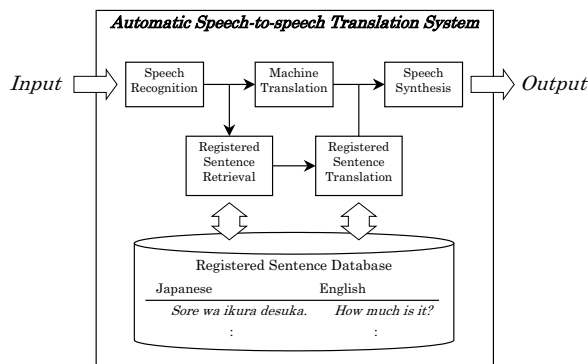


Figure 1: The configuration of our prototype system

the points of view of scene and action.

1) Search by scene

The number of scenes where travelers are likely to have a conversation is limited and can be systematically classified regarding places such as an airport, a hotel, or a restaurant.

We provide a directory-type search system that can be used to search for sentences by scene. We built up the travel-scene directory tree and assigned sentences to the leaf nodes of the tree. When the user selects a scene in the tree, sentences belonging to that scene are presented to the user. The selected scene does not change until the user selects another scene in this search system since the user generally will not move to a different scene while talking.

2) Search by action

Since it is difficult to represent actions with keywords and a traveler's range of probable actions in overseas travel is limited, we also provided a directory-type search system to search for sentences by action. We constructed a directory tree of traveler actions, and the user can obtain the sentences used for an action by selecting the action from the tree.

By inputting a natural language sentence and selecting a scene and an action, the user can obtain sentences that include the keywords extracted from the input sentence and that match the selected scene and action. When the user selects a different scene or action, the system again searches through the registered sentences using the new condition regarding the scene or action along with the original condition that was not changed. This enables the user to dynamically adjust the search conditions.

Table 2: Top layer nodes of the scene directory

Scenes
Using Interpretation Machine
Basic Expressions
On the airplane
Airport
Hotel
Restaurant
Shopping
Transportation
Rent a car / Driving
Sightseeing / Entertainment
Telephone / Mail / Bank
Property loss / Incident / Accident
Sickness / Injury

Table 3: Action directory

Actions
Making a request
Asking for permission
Asking a question
Complaining
Explaining
Greeting

3 Prototype System

We have integrated free-style sentence translation (Watanabe et al., 2000) and parallel text based translation based on the model described in the previous section and built a new prototype system. Here, we describe the system configuration, the contents of the registered sentences in the system, and the scene and action directories. We also explain how the user operates the system interface.

3.1 System Configuration

The prototype system consists of six components — speech recognition, machine translation, registered sentence retrieval, parallel text based translation, registered sentence database, speech synthesis (Figure 1). We have utilized the speech recognition, machine translation, and speech synthesis described in (Watanabe et al., 2000). The registered sentence retrieval component searches the registered sentence database using the system input. The parallel text based translation component produces a translation of the registered sentence selected by the user from the search results provided by the registered sentence retrieval component. The system input can be used as the machine-translation target and as a search key for registered sentence retrieval according to the user's instruction.

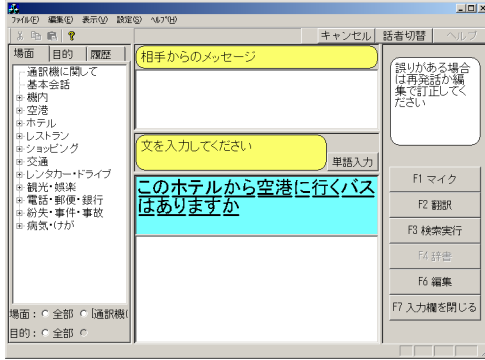


Figure 2: A sentence inputted to the automatic interpretation system

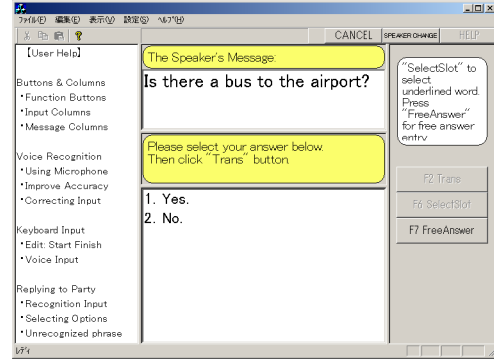


Figure 4: English translation of the first registered sentence in Figure 3

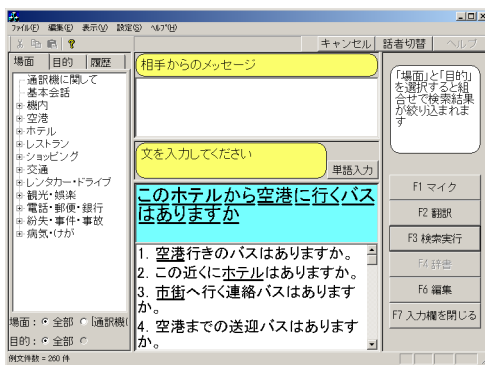


Figure 3: The result of registered sentence retrieval for the sentence in Figure 2

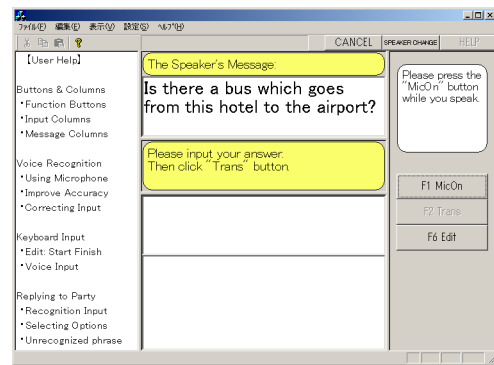


Figure 5: English translation of the input sentence in Figure 2

3.2 Registered Sentences

We first listed a traveler's typical tasks in eleven scenes where travelers often have to speak to people and then made a list of typical subtasks by analyzing the process necessary to accomplish each task. Next, we composed a sentence template for each subtask and a list of typical words that could be inserted into each slot of the templates. We have composed 2590 templates, which can be used to generate 7410 sentences with the slot word-lists, and have installed these in the system.

We have also composed 1185 templates, which can be used to generate 1796 sentences through slot word expansion, as response candidates for the respondent. Sharing a set of response candidates among several sentences for the traveler decreases the total number of response templates needed. A set of response candidates is linked to every sentence for the traveler to which the respondent can respond.

3.3 Scene and Action Directories

For each of the eleven scenes, we listed the relevant tasks in a two-layered tree with 70 leaf nodes to create a scene directory. Table 2 shows the top layer nodes of the scene directory.

We used only the six actions listed in Table 3 for the action directory and constructed a one-layered tree since it is difficult for the user to select the action if actions are classified in detail.

3.4 User Interface

Figure 2 shows the display screen of the prototype system. In this example, the user inputs the Japanese sentence “*Kono hoteru kara kūkō ni iku basu wa arimasuka.* (Is there a bus going to the airport from this hotel?)” by speaking. The result of the speech recognition is displayed in the input window at the center of the screen.

When the user clicks the “*kensaku jikkō* (search)” button in the screen, the system searches among

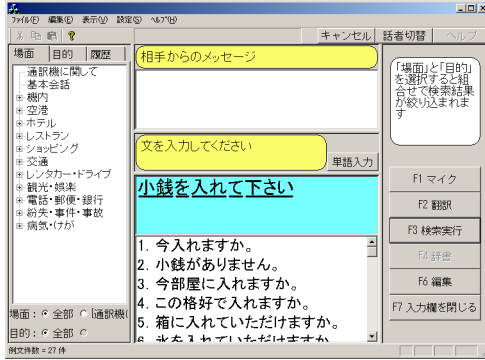


Figure 6: The result of registered sentence retrieval for the sentence “*Kozeni o irete kudasai*. (I’d like some small change.)”

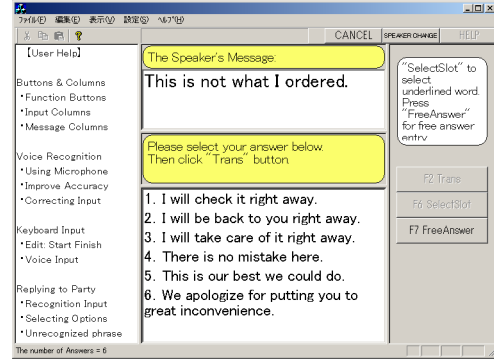


Figure 8: English translation of the registered sentence “*Kore wa chūmon to chigaimasu*. (This is not what I ordered.)”

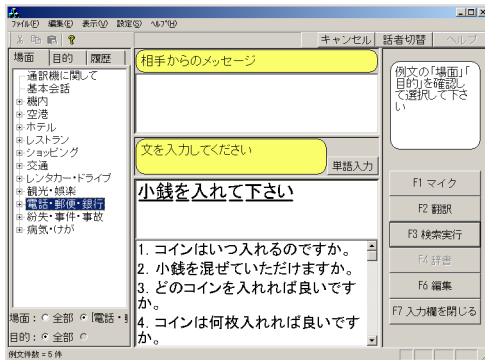


Figure 7: The screened result for the scene “*Denwa · Yūbin · Ginkō* (Telephone / Mail / Bank)”

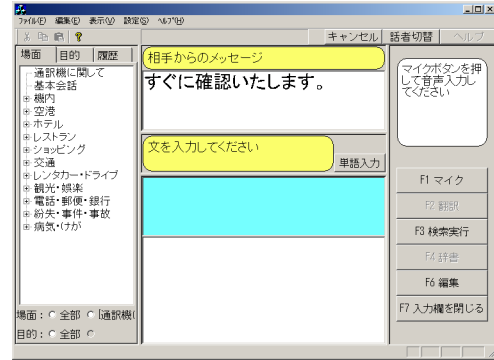


Figure 9: Japanese translation of the first response in Figure 8

the registered sentences using the input sentence as a key and displays the search result under the input window (Figure 3). The sky-blue color of the background in the window indicates the sentence selected as the target for translation. The user can select another sentence including the input sentence by clicking it.

When the user clicks the “*honyaku* (translate)” button after selecting the first registered sentence in Figure 3, the system retrieves an English translation of the sentence registered with the Japanese sentence, displays it (Figure 4), and reads it through the speech synthesis.

If the user cannot find an appropriate sentence in the search results, the user can resort to free-style sentence translation. When the user clicks the “*honyaku* (translate)” button after selecting the input sentence, the system translates it into English through the machine translation, displays it (Figure

5), and reads it through the speech synthesis.

In this way, the user can use free-style sentence translation and parallel text based translation seamlessly for the same input sentence.

Next, we explain how a user can narrow down the search result by using the directory.

Figure 6 shows the system display when the user inputs the Japanese sentence “*Kozeni o irete kudasai*. (I’d like some small change.)” by speaking and searches for a matching registered sentence. The search result is displayed in the lower central window. In this case, no appropriate sentence appears among the higher ranking sentences.

In Figure 6, the scene directory is displayed in the left part of the window. When the user selects the scene “*Denwa · Yūbin · Ginkō* (Telephone / Mail / Bank)”, the search result is narrowed down to the sentences associated with the scene in “Telephone / Mail / Bank” (Figure 7). The registered sentence

“Kozeni o mazete itadakemasuka. (I’d like some small change.)” is then displayed at the second of the results, and the user can use this sentence for translation.

The user can similarly narrow down the result with the action directory. If necessary, the user can also use a combination of the scene and the action directories.

We next explain how a respondent can respond by selecting from among the response candidates registered in the system.

Figure 8 shows the screen of the system when the user selects the registered Japanese sentence “Kore wa chūmon to chigaimasu. (This is not what I ordered.)” and translates it. The English translation is displayed in the upper central window of the screen, and the response candidates are listed in the lower central window.

When the respondent selects the first response from these and clicks the “Trans” button, the system displays Japanese translation of the response (Figure 9) and reads it through the speech synthesis.

In this way, the respondent can easily respond to the traveler by selecting a response from among the provided candidates when the traveler uses a registered sentence. The traveler can thus fully understand the response.

4 Evaluation

In this section, we evaluate the prototype system with respect to the extent that typical traveler utterances are covered by the registered sentences and whether the user can easily find a registered sentence that matches a natural utterance.

4.1 Coverage provided by the Registered Sentences

The system can provide correct translation if the input sentence can be matched to a registered sentence; otherwise, it can only provide a translation through machine translation whose quality is uncertain. To determine the proportion of commonly used sentences for which the system would provide a correct translation, we evaluated the coverage provided by the registered sentences for sentences randomly extracted from travel conversation corpora.

Table 4 displays the evaluation results. In the table, “closed set” denotes the result for sentences extracted from the corpora we referred to when developing the registered sentences and “open set” de-

Table 4: Coverage provided by the registered sentences for travel conversation corpora

	Total	Covered
Closed set	358	260 (72.6%)
Open set	308	163 (52.9%)

Table 5: The number of the subtask for which proper registered sentence retrieved

Total Subtask	Sentence Retrieved
116	74 (63.8%)

notes the result for sentences extracted from corpora not referred to when developing the registered sentences. The registered sentences covered 72.6% of the sentences in the closed set and 52.9% of the sentences in the open set.

The coverage of almost 73% for the closed set suggests that roughly 27% of the sentences in the corpora are not used for typical travel conversation. If the open set includes an equal proportion of atypical sentences, the registered sentences cover about 72% of the typical sentences used for travel in the open set.

We therefore believe that our prototype system can provide reliable translation for a minimum set of utterances necessary for overseas travel.

4.2 Basic Performance of the Registered Sentence Retrieval

Since registered sentences are retrieved mainly by using sentences inputted with natural language, a poorly performing retrieval system may prevent the user finding an appropriate sentence that is registered in the system. To determine whether the system can reliably retrieve an appropriate sentence from the utterance the user first thinks of, we randomly picked 116 subtasks for which registered sentences had been developed, had experimental subjects compose a natural sentence that could be used to accomplish each subtask, and evaluated whether the retrieval result when using a composed sentence as a key included the registered sentence.

Table 5 shows the evaluation results. The user could find a registered sentence corresponding to the natural utterance with our retrieval system for 63.8% of the traveler’s subtasks when a sentence for the subtask was registered in the system.

Although this retrieval system performance is not sufficient when we take into consideration the decreased performance caused by recognition errors in the input sentence, we expect to improve the retrieval performance by adding expressions synonymous with those in each sentence template to the index used for registered sentence retrieval.

5 Related Work

Example-based machine translation has been proposed as a method for translating with parallel bilingual sentences (Nagao, 1984; Shirai et al., 1997; Furuse et al., 1994). An example-based machine translation system retrieves example sentences similar to the input sentence, and translates by appropriately assembling and adapting the retrieved sentence to the input.

On the other hand, the parallel text based translation that we apply uses a translation that corresponds to each sentence as it is. Our system thus enables robust overall translation by allowing users to select the most appropriate sentence for the situation.

The technique of translation memory allows users to apply parallel bilingual sentences and widely used in commercial systems (Falcone, 2000). Previous translation results are stored and reused to provide similar translations for new input. This technique is mainly used for document translation to improve the efficiency of translation and to ensure the consistency in translation results.

On the other hand, the parallel text based translation that is used in our system was designed to allow seamless cooperation with real-time speech-to-speech translation. Also, it is equipped in advance with sentences specially composed for any likely conversation.

6 Conclusion

We have developed an automatic interpretation system by integrating free-style sentence translation and parallel text based translation. The system provides a predefined always-correct translation when the user can use a registered sentence. Users can easily switch between the two forms of translation if necessary.

Our prototype of the automatic interpretation system for Japanese overseas travelers includes 9206 task-oriented parallel bilingual sentences for travelers. The composed sentences cover 72% of the sentences typically used during overseas travel. The

system also provides predefined response candidates that a respondent can use when answering the traveler.

Registered sentences are searched for in response to natural language sentences input by the user. The user can narrow down the search results by specifying a scene and an action. We found that users could find a registered sentence that corresponded to a natural utterance for 64% of traveler's subtasks if a sentence for the subtask was registered in the system.

Our next step is to improve the performance of the registered sentence retrieval. We plan to add expressions synonymous with those in each sentence template to the index used for registered sentence retrieval. We also plan to restrict the search space by estimating the scene from the dialogue context.

References

- Suzanne Falcone. 2000. More translation memory tools (not many more, but good ones). *Translation Journal*, 4(2).
- Osamu Furuse, Eiichiro Sumita, and Hitoshi Iida. 1994. Transfer-driven machine translation utilizing empirical knowledge. *Transactions of IPSJ*, 35(3):414–425. in Japanese.
- Osamu Furuse, Setsuo Yamada, and Kazuhide Yamamoto. 1998. Splitting long or ill-formed input for robust spoken-language translation. In *Proceedings of COLING-ACL'98*, pages 421–427.
- Makoto Nagao. 1984. A framework of a mechanical translation between Japanese and English by analogy principle. In Alick Elithorn and Ranan Banerji, editors, *Artificial and Human Intelligence*, pages 173–180. North-Holland.
- Satoshi Shirai, Francis Bond, and Yamato Takahashi. 1997. A hybrid rule and example based method for machine translation. In *Proceedings of NLPRS-97*, pages 49–54.
- Takao Watanabe, Akitoshi Okumura, Shinsuke Sakai, Kiyoshi Yamabana, Shinichi Doi, and Ken Hanazawa. 2000. An automatic interpretation system for travel conversation. In *Proceedings of ICSLP 2000*.
- Wakita Yumi, Kawai Jun, and Iida Hitoshi. 1997. Correct parts extraction from speech recognition results using semantic distance calculation, and its application to speech translation. In *Proceedings of ACL/EACL-97 Workshop on Spoken Language Translation*, pages 24–31.