Simultaneous Interpretation Utilizing Example-based Incremental Transfer

Hideki Mima[†], Hitoshi Iida and Osamu Furuse[††]
ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-0288, Japan
EMAIL: H.Mima@doc.mmu.ac.uk, iida@itl.atr.co.jp, furuse@cslab.kecl.ntt.co.jp

Abstract

This paper describes a practical method of automatic simultaneous interpretation utilizing an example-based incremental transfer mechanism. We primarily show how incremental translation is achieved in the context of an example-based framework. We then examine the type of translation examples required for a simultaneous interpretation to create naturally communicative dialogs. Finally, we propose a scheme for automatic simultaneous interpretation exploiting this example-based incremental translation mechanism. Preliminary experimentation analyzing the performance of our example-based incremental translation mechanism leads us to believe that the proposed scheme can be utilized to achieve a practical simultaneous interpretation system.

Introduction

Speech-to-speech translation necessitates quick and perspicuous responses to natural communication. Furthermore, since dialogues continuously expand, it is essential to incrementally translate inputs to avoid interrupting the coherency of communications. Therefore, a high degree of incrementality and acceptability in translation such as simultaneous interpretation is essential. To satisfy these requirements, an incremental translation system, which functions as a simultaneous interpreter, is seen as an efficient solution in this field.

The main characteristic of incremental translations is the translation process. This is activated synchronously with the input, in contrast with conventional sentence-by-sentence-based translation which cannot start processing until the end of an input (Kitano, 1994). However, in incremental translation, we believe that the following issues must be resolved to achieve actual simultaneous interpretation :

- **How to define Information Units (IUs)** (Halliday, 1994) **to determine appropriate components for translation** – Since differences exist among the word order of various languages, especially between linguistically distant languages such as English and Japanese, appropriate transfer units, equally effective for both the source and target languages, have to be defined.

- **How to determine plausible translation for each IU** – In terms of the information content, the greater the number of words contained in IUs, the less semantic ambiguity in translation, or the later the response is obtained. Because of time restrictions, deterministic processing by exploiting specious measures (e.g. linguistical or statistical plausibility) is required for each IU translation in order to shorten the length of IUs.

- **How to install simultaneous interpreters' know-how (i.e. empirical knowledge)** – In practical simultaneous interpretation, human translators generally use strong sentence planning using particular empirical know-how. The exploitation of this kind of knowledge is essential for achieving practical simultaneous interpretation (Kitano, 1994).

Transfer-Driven Machine Translation (TDMT) (Furuse, 1994a) (Mima, 1997) has been proposed, and an efficient method of spoken dialog translation. TDMT has the following key features:

- **Utilization of Constituent Boundary Patterns (CB-Patterns)** (Furuse, 1994b) (Furuse, 1996) – CB-Patterns based on meaningful information units are applied to parse an input incrementally and produce translations based on the synchronization of the source and target language structure pairs (Abeillé, 1990) (Shieber, 1990). This contrasts with the linguistic manner of applying grammar rules.

The result of this provides for incremental translations that can even handle lengthy input

---

efficiently by splitting the input into appropriate and meaningful chunks. In addition,

- **Existence of efficient disambiguation scheme**
  – by dealing with best-only substructures utilizing stored empirical translation examples compiled from a linguistic database, the explosion of structural ambiguities is significantly constrained (Furuse, 1996).

Accordingly, TDMT has the advantage of having both the capability to define effective IU and an efficient deterministic processing scheme in incremental spoken-language translation.

Additionally, in exploiting the empirical knowledge that is required in practical simultaneous interpretation, we can assume that the empirical knowledge is described within the linguistic resource of simultaneous interpretation corpora. (Harbusch, 1992) proposed a method of default handling in incremental generation based on this observation.

In this paper, we describe the achievement of practical simultaneous interpretation using a TDMT. Furthermore, we discuss what kind of empirical knowledge is required for realizing efficient simultaneous interpretation, in terms of a simultaneous translator's knowledge, as well as proposing a method to exploit this empirical knowledge in an example-based framework in order to produce consistent translations.

A preliminary experiment analyzing our proposed scheme indicates that it should be able to be used in achieving simultaneous interpretation systems.

The next section of the paper briefly explains incremental translation using TDMT. Section 2 discusses the type of empirical knowledge necessary in simultaneous interpretation using some examples. Section 3 describes our proposed scheme for exploiting simultaneous interpretation examples. Section 4 presents a preliminary experiment for analyzing our proposed scheme to confirm its feasibility. Section 5 examines some related research in the field of incremental translation. Finally, a summary of our approach concludes this paper.

1 Incremental Translation Using Transfer-Driven Machine Translation

1.1 Constituent Boundary Pattern

In TDMT, translation is performed by applying stored empirical transfer knowledge, which describes the correspondence between source language expressions and target language expressions at various linguistic levels. The source and target expressions from the transfer knowledge in TDMT are expressed as CB-Patterns, which represent meaningful units for linguistic structure and transfer. The efficient application of transfer knowledge source components to an input string plays a key role in our basic incremental translation scheme. A pattern is defined as a sequence that consists of variables and constituent boundaries such as surface functional words.

The transfer knowledge is compiled from actual translation examples in every source pattern.

1.2 Incremental Pattern Application

The incremental application of CB-Patterns is based on the idea of incremental chart parsing (Furuse, 1996) (Amtrup, 1995) with notions of linguistic levels.

The procedure for the application of CB-Patterns is as follows:
(a) Determination of possible pattern applications.
(b) Translation candidate determination and structural disambiguation of patterns by semantic distance calculation.

Our scheme determines the best translation and structure parallel with an input sequence and can restrain the number of competing structures (possible translation candidates) at the possible utterance point in the input by performing (a) in parallel with (b), thus reducing the translation costs in time. The structure selected in (b) has its result transferred with head word-information using semantic distance calculations when combined incrementally with other structures. The output sentence is generated as a translation result from the structure for the whole input, which is composed of best-first substructures.

In order to limit the combinations of patterns and control the appropriate timing of each partial utterance during pattern application, we distinguish pattern levels, and specify the linguistic sublevel permitted for use in the assigned variables for each linguistic level. This is because if any combinations of patterns are permitted, it is obvious that the possibility of combinations are easily exploded. Table 1 shows examples of the relationship between linguistic levels. Every CB-pattern is categorised as one of the linguistic levels, and a variable on a given level is instantiated by a string on the linguistic levels in the second column of Table 1.

For instance, in the noun phrase "X of Y", the variables X and Y cannot be instantiated by a

simple sentence pattern, but can be instatiated by NP such as a noun phrase pattern or a compound noun pattern.

Moreover, these levels give a guideline to the timing of utterance production (i.e. the timing of when an utterance is said). For example, each simple sentence level pattern has utterance markers (Table 2, where '/' indicates the utterance markers) for possible insertion of an utterance during left-to-right application of the pattern. Thus, redundant or incomplete partial matchings can be eliminated and an appropriate trigger of utterance can be obtained.

(Furuse, 1996) provides further details of the algorithm for incremental CB-Parsing.

Table 1 Possible linguistic sublevels in variables

Linguistic level	Sublevels of variables
Simple sentence	VP, NP, ...
Verb phrase (VP)	VP, NP, verb, ...
Noun phrase (NP)	NP, CN, proper-noun, ...
compound noun (CN)	CN, noun, ...

Table 2 Utterance markers

English pattern	Japanese Pattern
By the way / *X*	*tokorode* / *X'*
No / *X*	*iie* / *X'*
X but / *Y*	*X'* / *shikashi Y'*
X if / *Y*	*X'* / *moshi Y'*
X / *where Y*	*X'* / *Y'*

1.3 Disambiguation of Translation Candidate

The CB-pattern "*X no Y*" with the particle "*no*" is a frequently used expression in Japanese. We can observe the following Japanese-to-English transfer knowledge about "*X no Y*" from such translation examples as the source-target pairs of : "*hoteru no jûsho*" → "*the address of the hotel*", "*eigo no paNfuretto*" → "*the pamphlet written in English*", etc.

X no Y =>

   Y' of X' ((*hoteru , jûsho*), ...),

     'hotel' 'address'

   Y' written in X' ((*eigo, paNfuretto*), ...),

     'English' 'pamphlet'

   Y' for X' ((*asu, tenkô*), ...),

     'tomorrow' 'weather'

   :

Within this pattern, *X'* is the target word corresponding to *X*, and a corresponding English word is written below each Japanese word. For example, "*hoteru*" means '*hotel*', and "*jûsho*" means 'address'.

This transfer knowledge expression indicates that the Japanese pattern "*X no Y*" corresponds to many possible English expressions. (*hoteru, jûsho*) are sample bindings for "*X no Y*", where *X* = *hoteru*, and *Y = jûsho*.

TDMT makes the most of an example-based framework, which produces an output sentence by mimicking the closest translation example to an input sentence. The semantic distance from the input is calculated for all examples. Then the example closest to the input is chosen, and the target expression of that example is extracted.

Suppose that the input is "*nihoNgo no paNfuretto*", where *nihoNgo* means '*Japanese*', and the input is closest to (*eigo, paNfuretto*); "*the pamphlet written in Japanese*" can be gained by choosing *Y'* *written in X'* as the best target expression.

Furthermore, ambiguity in the combination of patterns, which have not been constrained by the linguistic levels, is also dissolved incrementally by using the total sum of the semantic distances of patterns contained (Furuse, 1996).

The distance between an input and a translation example is measured based on the semantic distance between the words contained, and the semantic distance between words is calculated in terms of a thesaurus hierarchy. (Sumita, 1991) provides further details of the semantic distance caluculation.

2 Exploitation of a Simultaneous Interpreter's Empirical Knowledge

In practical simultaneous interpretation, human translators generally use strong sentence planning such as transformation between the active and the passive voice, transformation from a lengthy interrogative sentence to a tag question, and topicalization transformation. Moreover, the input is produced and modified in a step-by-step manner, so that it can be temporarily incomplete – although as a whole sentence it may become sufficient. Thus, the consistency of translations has to be adjusted appropriately when a contradiction occurs between a previously uttered part of the translation and the part currently being translated.

As a consequence of under specification, simultaneous interpretation is essentially based on

working with empirical knowledge – e.g. simultaneous interpreters' translation examples.

In this section, we first describe the kinds of examples that are required to achieve simultaneous interpretation using some sample sentences.

2.1 Empirical Knowledge

- **Transformation to a tag question**

Let us consider the following Japanese utterance:

(E1) *Nani-mo moNdai-wa ari-maseN* -<pause>- *de-shô-ka.* (what problem exist -<pause>- is there)[1]

In Japanese, an interrogative is specified at the end of the sentence, while in English, it is generally specified in front of the sentence. Thus, although a translation of the whole sentence of (E1) is *'Is everything all right'*, in some cases, *'Everything is all right'* could be uttered after the pause in the incremental framework. In this case, the meaning of the previously uttered part is no longer consistent with the current translation. However, even in this case, translation can be continued transforming to a tag question as (E1)' by using a peculiar translation example [TE1] without interruption by semantic inconsistency and the insertion of a restatement.

[TE1] *(X de-shô-ka)* → *(X', isn't it)*

(E1)' *Everything is alright, isn't it.*({[TE1]: X' = 'Everything is alright'})

- **Negative sentence**

Let us consider the following utterance:

(E2) *TsuiNrûmu-wa gozai-masu* -<pause>- *ga, hoNjitsu-wa goriyô-ni-nare-maseN.* (twin room exist -<pause>- but today not-available)

In Japanese, negation is also specified at the end of the sentence while in English it has to be specified in front of the finite verb. In addition, an expression *"X wa gozai-masu"* in (E2) has possible translations as *"we have X'"* or *"X' is available"*. Thus, although the whole translation should ideally read as "We have twin rooms, but none are available today", "A twin room is available" might be selected as a part of the translation in some cases. Although one solution could be to restate previously uttered phrases such

as: "no, sorry, we do have twin rooms, but none .....", such restatements should not be used frequently. This is because the restatements tend to break in general, coherency of human interaction

However, in this case, translation can be continued as (E2)' by using a peculiar translation example [TE2], with no restatement.

[TE2] *(X ga, Y)* → *(X' usually, but Y')*

(E2)' A *twin room is available usually, but we do not have any vacancies today.* ({[TE2]: X'= 'A twin room is available', Y'='we do not have any vacancies today'})

- **Failure of prediction**

In simultaneous interpretation, elements are usually uttered before the input consumption has been finished. Thus, because of the uncertainty in assumptions, a system with this facility must be able to adjust the whole content of the translation when it is realized that the assumption is incorrect from information given later.

Consider the following English utterance:

(E3) *That restaurant is open* -<pause>- *as only as in the evening.*

In the case of the part of the translation already uttered, "sono-resutoraN-wa ôpuN-shite-I-masu", it should have been inserted "yoru nomi" in front of the phrase "ôpuN-shite-I-masu", when the whole sentence is translated.

However the translation can be continued as it is as in (E3)' by using a peculiar translation example [TE3].

[TE3] *(X as only as Y)* → *(X' I-masu, ga, Y' nomi-desu)*

(E3)' *Sono-resutoraN-wa ôpuN-shite I-masu, ga, yoru nomi-desu* ({[TE3]: X'= 'ôpuN-shite', Y'='yoru'})

As the above example shows, simultaneous interpretation as skilled as that performed by a human interpreter is achievable by exploiting peculiar translation examples - i.e. simultaneous interpretation examples (or SI-examples, in short).

In the next section, we propose an algorithm to handle these kinds of SI-example with the best-first example-based incremental MT mechanism.

3 Simultaneous Interpretation Algorithm

Although the main characteristic of example-based translation is the use of the most similar examples as the main knowledge source for translation, the exploitation of SI-examples is drawn from the following consideration :

---

[1] In this paper, sample Japanese is Romanized in italic based on the Hepburn system with the corresponding English words following in parentheses.

- **A translation should use an example consistent with previously uttered information**

Thus, the key translation process with exploiting SI-examples consists of the following stages:

(1) Checking the contextual consistency between previously uttered phrases[2] and the phrase to be uttered next.

(2) Retrieving the most plausible example according to both the contextual sequence and similarity.

(3) Re-translating the phrase to be uttered next by using the example retrieved in (2)

The algorithm is described as follows. In the algorithm, the input phrase to be considered as a combination of structures shown in Figure 1 to facilitate understanding of the algorithm. For example, in the case of (E3), ST_1 indicates "*The restaurant is open*", ST_2 indicates "*open as only as in the evening*", and $ST_{1,2}$ indicates the whole phrase. In addition, trans(ST_i) returns word sequence indicating translation of ST_i, trans(ST_i, E) also returns word sequence indicating the translation of ST_i using example E, and i indicates the current processing part. Since the algorithm for the exploitation of SI-examples is applied only if a previous translated phrase exists, the proposed algorithm is executed in the case of i>=2.

Algorithm:

Start.

1. Retrieve the similar examples of ST_i from the total example database (normal + SI-examples) and assign the list to the {SE} with the appropriate semantic distance.

2. Produce trans(ST_i, E), where E indicates the most similar example listed in {SE}.

3. Remove the example E from {SE}.

4. If trans($ST_{i-1,i}$, E) == trans(ST_{i-1}) +[3] trans(ST_i, E)[4],

---

[2] In this paper, we only state the context within a sentence and do not refer to contexts between dialogs.

[3] Indicating sequencial appending operation, which includes removal operation of the common sub-sequence among the last of the first item and the first of the second item. For example, word sequences "A B" + word sequences "B C" indicates "A B C" .

[4] i.e. trans(ST_{i-1}) and trans(ST_i) are contextually continuous. In this paper, we define contextually continuous from the view point of sequences of concrete words (phrases) contained, in terms of combination with an example-based framework.
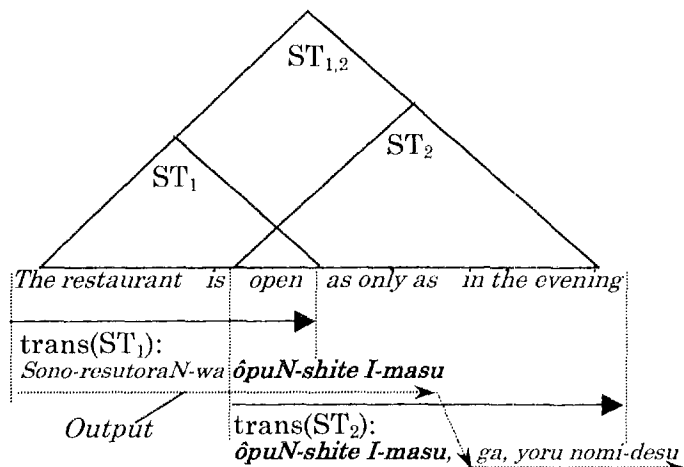
Figure 1 Notation of Substructures

then, output the difference between trans(ST_i, E) and trans(ST_{i-1}), then **goto End**.

5. **Goto 2.**

End.

In the majority of conventional example-based frameworks, only a semantic similarity is considered in retrieving the examples to be applied. In our scheme, on the other hand, not only semantic similarity but also contextual consistency with the previous translation is considered. In other words, the key notion of the scheme is its mechanism for selecting appropriate examples. Hence, as the above algorithm shows, exploitation of SI-examples can be combined smoothly with the conventional example-based framework.

Let us explain the algorithm in terms of sentence (E3) as an example. First, assuming that trans(ST_1) = "*Sono-resutoraN-wa ôpuN-shite I-masu*" (the-restaurant open), the most similar example of ST_2 is normally:

[TE4] *(X as only as Y)* → *(Y' nomi X' I-masu)*

Thus, trans(ST_2, TE4) can be "*yoru nomi ôpuN-shite I-masu*" (evening only open) and as the phrase "*yoru nomi ...*" is, in this case, not contextually continuous, and the next example should be extracted from the similar example list {SE}. Then, the example is [TE3], since trans(ST_2, TE3) = "*ôpuN-shite I-masu, ga, yoru nomi-desu*", in terms of the contextual order of the words, this translation can be continuous. Thus, the difference between trans(ST_1) and the trans(ST_2, TE3), "*ga, yoru nomi-desu*" can be obtained as the next utterance.

859

4 Preliminary Experiments

We conducted a preliminary experiment with respect to (a) the quality of example-based translation in relation to IUs (i.e., meaningful units), and (b) the quality and speed of incremental parsing (CB-Parsing), to confirm the feasibility of our proposed scheme.

In the evaluation of (a), we conducted a jack-knife experiment to measure the average success rate of translation for the most frequently used expressions (i.e. the most ambiguous) in Japanese, "X no Y" and "X wo Y". We prepared 774 and 689 examples for the expressions respectively, and conducted the experiment in increments of 100 examples (Furuse, 1994a). The examples were extracted by random sampling. We then evaluated the 10 translations of corresponding expressions in the dialog database for each case.

Figure 2 shows the average rate of the evaluation for 10 translations.

Although the translation quality of each unit depended on the type of expression, the graph shows that, in general, the more examples the system has, the better the quality[5].

Conditions of our experiment and evaluation for (b) are that the number of CB-patterns for Japanese-English translation and English-Japanese translation are 777 and 1241, respectively, and the number of total examples are 10000 and 8000, respectively. In the evaluation, we set the system to retain only one substructure in the semantic distance calculation in order to confirm the feasibility of deterministic processing at each incremental step.

CB-Parsing for 69-77 unseen dialogs (of 1,000 different unseen sentences) were manually evaluated by assigning a grade indicating success or failure. All of the parsing times include accessing time for an example database (i.e. corresponding to the whole transfer time) and were measured on a Sparc Station 10 workstation with 256 MB of memory.

Table 3 shows the experimental results. For CB-Parsing accuracy, a success rate of approximately 76 % was achieved for both translations, rates that are fairly high for spoken-language parsing.
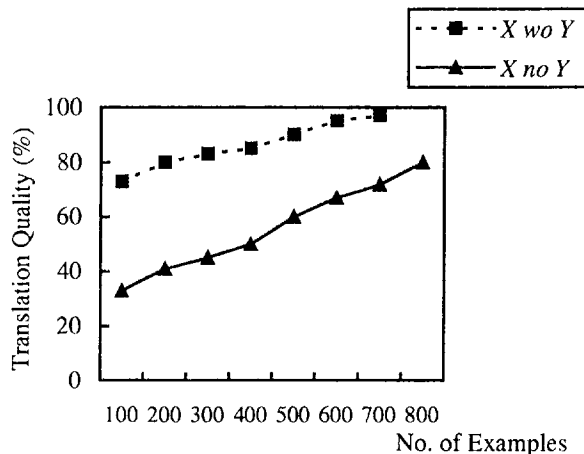

Figure 2 Quality of Example-based Transfer

Table 3 Evaluation Results

	J-E	E-J
No. of test dialogues (sent.)	69 (1225)	77 (1341)
Morphemes / sentence	9.7	7.1
CB-Parsing Accuracy	76.7 %	76.0 %
Parsing Time (average)	0.4 sec.	0.3 sec.

The main problem in the parsing procedure involved an insufficient number of examples for the CB-Pattern. However, as Figure 2 shows, an increase in the ratio with the number of examples could be observed with our framework. Thus, overall accuracy and acceptability should improve in proportion to an increase in transfer examples.

Although the speed depends on the amount of knowledge and sentence length, the average time was less than 0.4 seconds, which is fairly rapid. Thus, our translation scheme can be seen as an efficient translation mechanism in achieving a practical simultaneous interpretation system.

5 Related Research

Several schemes have been proposed with respect to incremental translation based on the synchronization of input and output fragments and the use of specialized information for simultaneous interpretation. (Kitano, 1994) proposes incremental translation that is based on marker-passing memory-based translation. Although the technique adopts a cost-oriented best-first strategy to avoid the explosion of structural ambiguity, the strategy does not pay attention to actual aspects of the overall meaning such as in the case when a previously made assumption turns out to be incorrect. (Matsubara, 1997) proposed a method to handle *extra-*

---

[5]However, we also have to ascertain the practical satiation limit, or how much the transfer knowledge can be expanded, as a future work.

860

grammatical phenomena with a chart-based incremental English-Japanese MT system based on observations of a translation corpus. However, this system was only capable of English to Japanese translation. In this paper, the aspects of flexible order, repetitions, and ellipses are only briefly considered and necessary extensions, such as the adjustment of consistency in related to the whole sentence by employing simultaneous interpreters' knowledge have not been previously investigated.

Conclusion

We have described a practical method of automatic simultaneous interpretation. In the exploitation of empirical knowledge, we examined the kind of empirical knowledge required to achieve efficient simultaneous interpretation. We then have proposed a method to exploit these empirical simultaneous translation examples in an example-based framework to produce a practical method of simultaneous interpretation.

Preliminary experimentation analyzing our proposed scheme showed that it can be utilized to achieve a simultaneous interpretation system.

The possibility of applying this sort of example-based framework into multilingual translation, such as a Japanese-German pair and a Japanese-Korean pair, has been shown in (Furuse, 1995) and (Mima, 1997). Therefore, the algorithm can also be expected to work for not only an English-Japanese pair but also other language pairs.

Important areas of future research will involve methods for:

- **Predicting the contents of the next utterance by using dialog-specific discourse analysis** (Levin, 1995)

- **Handling linguistic differences between the source and target languages such as subject ellipsis**

We believe that some situational information, such as the speakers-roles in the conversation (Mima, 1997) could be potentially helpful for both predicting the contents of the next utterance and resolving linguistic differences. The integration of statistical/stochastic approaches, such as Decision-Tree Learning (Yamamoto, 1997) for the above discourse-related issues is another area of interest for future work.

References

A. Abeillé, Y. Schabes and A. K. Joshi (1990) *Using Lexicalized Tags for Machine Translation*. In Proc. of Coling'90, pages 1—6.

J. W. Amtrup (1995) *Chart-based Incremental Transfer in Machine Translation*. In Proc. of 6th TMI. pages 188—195.

O. Furuse, E. Sumita, and H. Iida (1994a) *Transfer-Driven Machine Translation Utilizing Empirical Knowledge* (in Japanese). Trans. of Information Processing Society of Japan. Vol. 35, No. 3, pages 414–425.

O. Furuse, and H. Iida (1994b) *Constituent Boundary Parsing for Example-Based Machine Translation*. In Proc. of Coling '94, pages 105–111.

O. Furuse, J. Kawai, H. Iida, S. Akamine, and D. Kim (1995) *Multi-lingual Spoken-Language Translation Utilizing Translation Examples*. In Proc. of NLPRS'95, pages 544–549.

O. Furuse and H. Iida (1996) *Incremental Translation Utilizing Constituent Boundary patterns*. In Proc. of Coling '96, pages 412—417.

K. Harbusch, G. Kikui, and A. Kilger (1994) *Default Handling in Incremental Generation*. In Proc. of Coling'94. pages 356—362.

M. A. K. Halliday (1994) *An Introduction to Functional Grammar*. Edward Arnold.

H. Kitano (1994) *The Φ DM-DIALOG System*. In Speech-To-Speech Translation. H. Kitano. Kluwer Academic Publishers. Pages 47—113.

L. Levin, O. Glickman, Y. Qu, D. Gates, A. Lavie, C. P. Rosé, C. V. Ess-Dykema and A. Waibel (1995) *Using Context in Machine Translation of Spoken Language*. In Proc. of 6th TMI. pages 173—187.

S. Matsubara and Y. Inagaki (1997) *Utilizing Extra-grammatical Phenomena in Incremental English-Japanese Machine Translation*. In Proc. of 7th TMI. pages 31—38.

H. Mima, O. Furuse, and H. Iida (1997) *Improving Performance of Transfer-Driven Machine Translation with Extra-linguistic Information from Context, Situation, and Environment*. In Proc. of IJCAI'97. pages 983—988.

S. M. Shieber and Y. Schabes (1990) *Synchronous Tree-Adjoining Grammars*. In Proc. of Coling'90, pages 253—258.

E. Sumita and H. Iida (1991) *Experiments and Prospects of Example-based Machine Translation*. In Proc. of 29th ACL. pages 185—192.

K. Yamamoto, E. Sumita, O. Furuse, and H. Iida (1997) *Ellipsis Resolution in Dialogues via Decision-Tree Learning*. In Proc. of NLPRS'97. pages 423—428.