Cooperation between Transfer and Analysis in Example-Based Framework

Osamu FURUSE and Hitoshi IIDA

ATR Interpreting Telephony Research Laboratories
2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan
e-mail: furuse{iida}%atr-la.atr.co.jp@uunet.uu.net

Abstract

Transfer-Driven Machine Translation (TDMT) is presented as a method which drives the translation processes according to the nature of the input. In TDMT, transfer knowledge is the central knowledge of translation, and various kinds and levels of knowledge are cooperatively applied to input sentences. TDMT effectively utilizes an example-based framework for transfer and analysis knowledge. A consistent framework of examples makes the cooperation between transfer and analysis effective, and efficient translation is achieved. The TDMT prototype system, which translates Japanese spoken dialogs into English, has shown great promise.

1 Introduction

Many applications dealing with spoken-language, such as automatic telephone interpreting system, need efficient and robust processing. The system must be capable of handling many idiomatic expressions and spoken-language-specific expressions which deviate from conventional grammar.

Also, spoken-language has both easy and difficult expressions to translate. In human translation when translating an easy sentence, the translated result is produced quickly using only surface-level knowledge. When translating a complex sentence, a more elaborate process is performed, using syntactic, semantic, and contextual knowledge. Thus, many strategies and various levels of knowledge are used to effectively translate spoken-language.

This paper proposes a method called Transfer-Driven Machine Translation (TDMT), which carries out efficient translation processing by driving the necessary translation processes according to the nature of the input sentence. An example-based framework can achieve quick processing and consistently describe knowledge. The integration of transfer and analysis in an example-based framework is proposed as a method for achieving TDMT. In this method, transfer and analysis proceed autonomously and cooperatively. Also, a well-balanced load on each process can be achieved by employing this integrated processing mechanism.

Section 2 explains the idea of TDMT. Section 3 explains distance calculation and transfer in an example-based framework. Section 4 explains analysis in an example-based framework. Section 5 reports on the TDMT prototype system, and Section 6 reports on the experimental results.

The explanations in the following sections use Japanese-to-English translation.

2 Transfer-Driven Machine Translation

TDMT performs efficient and robust spoken-language translation using various kinds of strategies to be able to treat diverse input. Its characteristics are explained in the following sub-sections.

2.1 Transfer-centered cooperation mechanism

Translation is essentially converting a source language expression into a target language expression. In TDMT, transfer knowledge consists of various levels of bilingual information. It is the primary knowledge used to solve translation problems. The transfer module retrieves the necessary transfer knowledge ranging from global unit like sentence structures to local unit like words. The retrieval and application of transfer knowledge are flexibly controlled depending on the knowledge necessary to translate the input. Basically, translation is performed by using transfer knowledge. A transfer module utilizes analysis knowledge (syntactic/semantic information) which helps to apply transfer knowledge to some part of the input. And generation and context knowledge are utilized for producing correct translation result. In other words, TDMT produces translation results by utilizing these different kinds of knowledge cooperatively and by centering on transfer, and achieves efficient translation according to the nature of the input.

2.2 Utilization of example-based framework

Transfer knowledge is the basic data which is used for totally controlling the translation process.

Most of the transfer knowledge in TDMT is described by the example-based framework. An example-based framework is useful for consistently describing transfer knowledge. The essence of the example-based framework is the distance calculation. This framework achieves the best-match based on the distance between the input and provided examples, and selects the most plausible target expression from many candidates. The distance is calculated quickly because of its simple mechanism. Through providing examples, various kinds and levels of knowledge can be described in the example-based framework.

2.3 Multi-level knowledge

TDMT provides multi-level transfer knowledge, which corresponds to each translation strategy. In the transfer knowledge of the TDMT prototype system, there is string-, pattern- and grammar-level knowledge. TDMT achieves efficient translation by utilizing multi-level knowledge effectively according to the nature of input.

Some conventional machine translation systems also provide multiple levels of transfer knowledge for idioms, syntax, semantics, and so on, and try to apply these levels of that knowledge in a fixed order to cover diverse input [Ikehara et al. 87]. However, this method proceeds with the analysis for deciding which level of knowledge should be applied for any given input sentence in a fixed order, placing heavy load on the analysis module. Also, the knowledge description is rather more complicated than that of the example-based framework. Therefore, the translation of a simple sentence is not always quick because the system tries to cover all translation strategies.

3 Example-based Transfer

TDMT utilizes distance calculation to determine the most plausible target expression and structure in transfer.

3.1 Word distance

We adopt the distance calculation method of Example-Based Machine Translation (EBMT) [Sumita and Iida 91]. The distance between words is defined as the closeness of semantic attributes in a thesaurus. Words have certain thesaurus codes, which correspond to particular semantic attributes. The distance between the semantic attributes is determined according to the relationship of their positions in the hierarchy of the thesaurus, and varies between 0 and 1 (Fig. 1). The distance between semantic attributes A and B is expressed as d(A, B). Provided that the words X and Y have the semantic attribute A and B, respectively, the distance between X and Y, d(X, Y), is equal to d(A, B).

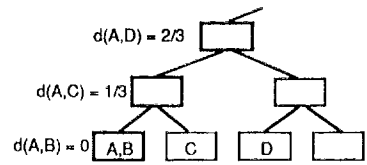

Figure 1 Distance between thesaurus codes

The hierarchy of the thesaurus that we use is in accordance with the thesaurus of everyday Japanese [Ohno and Hamanishi 84], and consists of four layers. when two values can be abstracted in the k-th layer from the bottom, the distance k/3 ($0 \leq k \leq 3$) is assigned. The value 0 means that two codes belong to exactly the same category, and 1 means that they are unrelated. The attributes "*writing*" and "*book*" are abstracted by the immediate upper attribute "*document*" and the distance is given as 1/3. Thus, the word "ronbun{technical paper}" which has thesaurus code "*writing*", and "yokoushuu{proceedings}" which has the thesaurus code "*book*", are assigned a distance of 1/3.

3.2 Description of Transfer Knowledge

Transfer knowledge describes the correspondence between source language expressions (SE) and target language expressions (TE) in certain meaningful units, preserving the translational equivalence [Tsujii and Fujita 91]. The condition under which a TE is chosen as a translation result of an SE is associated with the TE. Transfer knowledge in an example-based framework is described as follows:

SE => TE1 (E11, E12,...),
: :
TEn (En1, En2,...)

Each TE has several examples as conditions. Eij means the j-th example of TEi. The input is the SE's environment, and the most appropriate TE is selected according to the calculated distance between the input and the examples. The input and examples comprise a set of words.

Let us suppose that an input I and each example Eij consist of t elements as follows:

$I = (I1,...,It)$
$Eij = (Eij1,...,Eijt)$

Then the distance between I and Eij is calculated as follows:

$$d(I, Eij) = d((I1,...,It), (Eij1,...,Eijt)) = \sum_{k=1}^{t} d(Ik, Eijk) * Wk$$

The attribute weight Wk expresses the importance of the k-th element in the translation[1]. The distance from the input is calculated for all examples. Then the example whose distance to the input is least, is detected and the TE which has the example is selected. When Eij is closest to I, TEi is selected as the most plausible TE.

The enrichment of examples increases the accuracy of determining the TE because conditions become more detailed. Further, even if there is only one TE, but there is no example close to the input, the application of the transfer knowledge is rejected.

3.3 Wide application of distance calculation

Distance calculation is used to determine which TE has the example that is closest to the input, and can be used in various abstract level expressions depending on how the input words are provided.

Various levels of knowledge can be provided by the wide application of distance calculation. TDMT achieves efficient translation by utilizing multi-level knowledge effectively.

In the transfer knowledge of the TDMT prototype system, the string-, pattern- and grammar-level knowledge, the latter two of which can be described easily in an example-based framework, are now adopted. String-level knowledge is the most concrete, while grammar-level knowledge is the most abstract.

3.3.1 String-level transfer knowledge

Since this kind of knowledge has a condition outside the SE, the cooperation with such as context module is sometimes necessary.

In some cases the conditions can be described by the examples of the most closely related word in which the SE is used, as follows:

```
sochira   => this ((desu {be}[2])...),
             you  ((okuru {send})..),
             it   ((miru {see})...)
```

Applying this knowledge,"you" is selected as the word corresponding to the "sochira" in
"sochira ni{particle} tsutaeru" because of the small distance between "tsutaeru{convey}" and "okuru{send}".

3.3.2 Pattern-level transfer knowledge

Pattern-level transfer knowledge has variables. The binding words of the variables are regarded as input.

For example, "X o o-negaishimasu" {X particle will-ask-for} has a variable. Suppose that it is translated into two kinds of English expressions in the example below:

```
X o o-negaishimasu  =>
    may I speak to X'[3]   ((jimukyoku{office} ), ...),
    please give me X'      ((bangou{number}), ...)
```

In the translation of "X o o-negaishimasu", the TE is determined by the calculation below:

```
if   Min ( d((X), (jimukyoku)),....)
     <   Min ( d((X), (bangou)),....)

then  the TE is "may I speak to X'
else  the TE is "please give me X' "
```

The following two sentences have the pattern "X o o-negaishimasu":

(1) "jinjika{personnel section} o o-negaishimasu."
(2) "daimei {title} o o-negaishimasu."

The first sentence selects "may I speak to X' " because (jinjika) is close to (jimukyoku). The second sentence selects "please give me X' " ,because (daimei) is close to (bangou). Thus, we get the following translations:

(1') "may I speak to the personnel section."
(2') "please give me the title."

3.3.3 Grammar-level transfer knowledge

Grammar-level transfer knowledge is expressed in terms of grammatical categories. The examples consist of sets of words which are concrete instances of each category. The following transfer knowledge involves sets of three common nouns (CNs):

[1] Wk is given for each Ik by TE's distribution that semantic attribute of Ik brings {Sumita and Iida 91}.

[2] $\{w_1,...,w_n\}$ is the list of corresponding English words.

[3] A' is the transferred expression of A

```
CN1 CN2 CN3 =>
    CN3' of CN1'
        (("kaigi, kaisai, kikan
          {conference, opening, time}"),...),
    CN2' CN3' for CN1'
        (("sanka, moushikomi, youshi
          {participation, application, form}"),...),
      :
```

This transfer knowledge allows the following translations.

```
kenkyukai kaisai kikan {workshop, opening, time}
-> the time of the workshop
happyou moshikomi youshi
    {presentation, application, form}
-> the application form for presentation
```

The above translations select "CN3' of CN1' " and "CN2' CN3' for CN1' " as the most plausible TEs, as the result of distance calculations.

3.4 Disambiguation by total distance

When there are several ways to apply transfer knowledge to the input sentence, structural ambiguity may occur. In such cases, the most appropriate structure is selected on the basis of total distance. The least total distance implies that the chosen structure is the most suitable input structure. For example, when the pattern "X no Y" is applied to the sentence "kaigi no touroku hi no waribiki {conference, particle, registration, fee, particle, discount}", there are two possible structures:

(1) kaigi no (touroku hi no waribiki)
(2) (kaigi no touroku hi) no waribiki

The pattern "X no Y" has various TEs, such as in the following

```
X no Y  =>  Y' of X'   (E11, E12, ... ),
            Y' for X'  (E21, E22, ... ),
            Y' at X'   (E31, E32, ... ),
            X' Y'      (E41, E42, ... ),
              :              :
```

The respective TE tree representations constructed from structures (1) and (2) are shown in Figs. 2 and 3.

The structure of (1) transfers to "Y' of X' " with the distance value of 0.50 and "Y' of X' " with the distance value of 0.17, and generates (1') with a total distance value of 0.67. In structure (2), "Y' of X' " with the distance value of 0.17 and "Y' for X'" with the distance value of 0.17, generates (2') with a total distance value of 0.34. The latter result is selected because it has the least total distance value.

(1') "discount of the registration fee of the conference"
(2') "discount of registration fee for the conference"

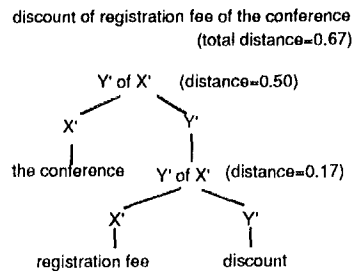

Figure 2 Translation of " kaigi no (touroku hi no waribiki) "

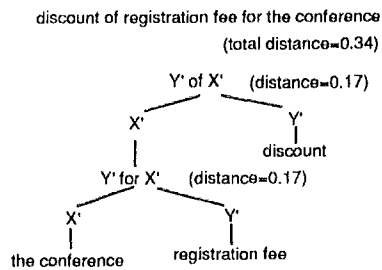

Figure 3 Translation of " (kaigi no touroku hi) no waribiki "

4 Example-based Analysis

For some structurally complex sentences, translations cannot be performed by applying only transfer knowledge. In such cases, analysis knowledge is also required. The analysis module applies analysis knowledge and supplies the resulting information to the transfer module, which then applies transfer knowledge on the basis of that information. When no analysis knowledge is necessary for translation, the application of only transfer knowledge produces the translation result. The analysis described in this paper is not the understanding of structure and meaning on the basis of a parsing of the input sentence according to grammar rules, but rather the extraction of the information

required to apply transfer knowledge and to produce the correct translation from the input sentence.

4.1 Description of analysis knowledge

Analysis knowledge is described by examples in the same way as transfer knowledge, as follows:

```
SE => Revised SE1  (E11, E12, ... ),
          :              :
      Revised SEn  (En1, En2, ... )
```

Although the form of knowledge description is virtually the same, transfer knowledge descriptions map onto TEs, whereas analysis knowledge descriptions map onto revised SEs.

4.2 Cooperation mechanism

The transfer and analysis processes operate autonomously but cooperatively to produce the translation result shown in Figure 4.

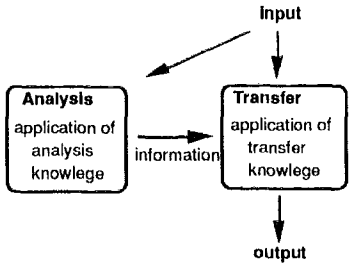

Figure 4 Relation between transfer and analysis

At present, we are providing analysis knowledge for normalization [Nagao 84] and for structuring with TDMT. In the following sections we will explain the cooperation mechanism between transfer and analysis based on these two kinds of analysis knowledge.

4.2.1 Analysis knowledge for normalization

Normalization is putting together minor colloquial expressions into standard expressions It leads to robust translation and efficient knowledge storage. Analysis knowledge for normalization is utilized to recover the ellipsis of function words such as particles, and to normalize some variant forms such as sentence-final forms into normal forms. Such knowledge helps the application of transfer knowledge to the input sentence.

The sentence "Watakushi wa Suzuki desu {I, particle, Suzuki, complementizer}" is translated into " I am Suzuki" by applying transfer knowledge such as the following:

```
X wa Y desu  => X' be Y'
```

However, in spoken Japanese, particles are frequently omitted. The sentence "Watakushi Suzuki desu" is natural spoken-Japanese. It is normalized to "Watakushi wa Suzuki desu", which has the omitted particle "wa" recovered, by applying the following analysis knowledge:

```
Pronoun  Proper-Noun  =>
    Pronoun  wa  Proper-Noun  (a set of examples)
```

The analysis module sends the information about the application of the analysis knowledge to the transfer module. The transfer module receives the information and applies the transfer knowledge to produce the English sentence " I am Suzuki"

By examples, this kind of analysis knowledge can also classify the particles to be recovered as shown below:

```
CN  Verb =>
    CN  o Verb
        ((hoteru{hotel}, yoyaku-suru{reserve}), ...),
    CN  ni  Verb
      ((kaigi{conference}, sanka-suru{participate}),...),
          :
```

This analysis knowledge allows the recovery of various particles such as,

```
"hoteru yoyaku-suru" -> "hoteru o yoyaku-suru"
"kaigi sanka-suru"-> "kaigi ni sanka-suru"
```

Analysis knowledge for normalization also has the advantage of making the scale of knowledge more economical and the translation processing more robust.

4.2.2 Analysis knowledge for structuring

Structuring is recognition of structure components of by insertion of a marker in order to apply transfer knowledge to each structure component. Analysis knowledge for structuring is applied to detect special linguistic phenomena such as adnominal expressions, wh-expressions, and discontinuities, so as to assign a structure to the SE.

Adnominal expressions appear with high frequency in Japanese, corresponding to various English expressions such as relative clauses, infinitives, pronouns, gerunds, and subordinate clauses. They can be detected by means of inflectional forms. Three components of adnominal expressions must be considered in the translation process: the modification relationship, the modifier, and

the modified. Analysis information for structuring is used to insert a marker at the boundary between the modifier and the modified. The following analysis knowledge can be constructed.

```
Adnominal-inflection   CN  =>
    Adnominal-inflection   Adnominal-marker   CN
                                (a set of examples)
```

This knowledge identifies adnominal relationships and separates the modifier from the modified so that transfer knowledge can be applied. When the transfer module receives the information about the application of this analysis knowledge, it applies the transfer knowledge needed to translate each component of the expression: the adnominal relationship, the modifier, and the modified. The scope of the modifier and the modified is determined by the total distance of each structure in which transfer knowledge is applied.

The following transfer knowledge about the adnominal relation determines the English expression by distance calculation with examples before and after the marker as follows:

```
X Adnominal-mark Y  =>
        Y' that X'   ((iku{go} , basu{bus} ), ...),
        Y' when X'  ((deru{attend} , hi{day} ), ...),
            :                    :
```

For example, analysis knowledge is applied to "Kyoto eki e iku basu{Kyoto station particle go bus}" , and the revised SE "Kyoto eki e iku Adnominal-marker basu" is produced. Then, by the application of the above transfer knowledge about the adnominal relation and the following transfer knowledge about the modifier and modified, the translation result "the bus that goes to the Kyoto station" is produced.

```
X e Y  =>   Y' to X',
Kyoto eki => Kyoto station,
iku => go,          basu => bus
```

5 TDMT Prototype System

A prototype Japanese to English system constructed to confirm the feasibility and effectiveness of TDMT is running on a Genera 8.1 LISP machine [Furuse and Iida 92].

Due to the restriction of the sequential mechanism, a method for driving the necessary process at the required time has not been completely achieved. However, the following control mechanism is used to obtain the most efficient processing possible.

- As much as possible, translation is attempted by first applying only transfer knowledge; when this fails, the system tries to apply analysis knowledge.

- Transfer knowledge is applied at the most concrete level as possible, that is, in the order of string, pattern, and grammar level.

In order to achieve flexible processing which exchanges necessary translation information, a parallel implementation is under study based on the results from the prototype system.

The knowledge base has been built from statistical investigation of the bilingual corpus, whose domain is inquiries concerning international conference registration. The corpus has syntactic correspondences between Japanese and English. We have established transfer and analysis knowledge as follows:

- string-level transfer knowledge (about 500 items)
- pattern-level transfer knowledge (about 300 items)
- grammar-level transfer knowledge (about 20 items)
- analysis knowledge (about 50 items)

6 Evaluation

We have evaluated the TDMT prototype system, with the model conversations about conference registration consisting of 10 dialogs and 225 sentences. The model conversations cover basic expressions. Table 1 shows the kinds of knowledge that were required to translate the model conversations.

Table 1 Knowledge Necessary to Translate Model Conversation
(total number of sentences - 225)

	sentences	rate
string only	73	32.4%
pattern and string only	90	40.0%
grammar-level transfer knowledge needed	21	9.3%
analysis knowledge needed	41	18.2%

At present, the prototype system can produce output quickly by the example-based framework.

200 of the sentences are correct, providing a success rate of 88.9%. The coverage by string- and pattern-level knowledge is wider than expected.

Table 2 shows the main causes of incorrect sentences.

Table 2 Causes of Incorrect Sentences
(total number of incorrect sentences - 25)

	occurrences
(1) inability to get such TEs as elided objects	9
(2) selection of incorrect TEs	8
(3) error in adverb position	4
(4) incorrect declension	1
(5) incorrect tense	1
(6) etc	2

The second factor shows that an elaboration of distance calculation and an enrichment of examples are needed. The first, third, and fourth factors are caused by the shortage of generation knowledge. The fifth factor is caused by the shortage of analysis knowledge. These facts show that the cooperative control that flexibly communicates various kinds of knowledge including context and generation knowledge, and various kinds of frameworks such as a rule-based and a statistical framework are useful to improve the translation performance.

7 Related Research

The example-based approach was advocated by Nagao [Nagao 84]. The essence of this approach is (a) retrieval of similar examples from a bilingual database and (b) applying the examples to translate the input. Other research has emerged following this line, including EBMT [Sumita and Iida 91], MBT [Sato and Nagao 90], and ABMT [Sadler 89]. EBMT uses phrase examples and will be integrated with conventional rule-based machine translation. MBT and ABMT use example dependency trees of examples and translate the whole sentence by matching expressions and by a left-to-right search of maximal matching. TDMT utilizes an example-based framework for various process as the method of selecting the most suitable TE, and combines multi-level transfer knowledge. On the other hand, MBT and ABMT utilize uni-level knowledge only for transfer.

8 Concluding Remarks

TDMT (Transfer-Driven Machine Translation) has been proposed. The prototype TDMT system which translates Japanese to English spoken dialogs, has been constructed with an example-based framework. The consistent description by example smoothes the cooperation between transfer and analysis, have shown the high feasibility. Important future work will include the achievement of flexible translation which effectively control the translation process. Also important is the implementation of TDMT in distributed cooperative processing by a parallel computer and incorporating various kinds of processing such as rule-based and statistical framework into the cooperation mechanism.

Acknowledgements

I would like to thank the members of the ATR Interpreting Telephony Research Laboratories for their comments on various parts of this research. Special thanks are due to Dr. Kohei Habara, the chairman of the board of ATR Interpreting Telephony Research Laboratories, Dr. Akira Kurematsu, the president of ATR Interpreting Telephony Research Laboratories, for their support of this research.

References

[Furuse and Iida 92] Furuse, O., and Iida, H. : An Example-based Method for Transfer-driven Machine Translation, Proc. of the Fourth International Conference on Theoretical and Methodological Issues in Machine Translation, (1992).

[Ikehara et al. 87] Ikehara, S., Miyazaki, M., Shirai, S., and Hayashi, Y : Speaker's Recognition and Multi-level-Translating Method Based on It, Trans.IPS Japan Vol.28 No.12., IPSJ , pp.1269-1279, (1987), (in Japanese).

[Nagao 84] Nagao, M. : A framework of a mechanical translation between Japanese and English by analogy principle, in Artificial and Human Intelligence, ed. Elithorn, A. and Banerji, R., North-Holland , pp.173-180, (1984).

[Ohno and Hamanishi 84] Ohno, S. and Hamanishi M. : Ruigo-Shin-Jiten, Kadokawa, (1984), (in Japanese).

[Sadler 89] Sadler, V. : Working with Analogical Semantics, Foris Publications (1989).

[Sato and Nagao 90] Sato, S. and Nagao M. : Toward Memory-Based Translation, Proc. of Coling '90, (1990).

[Sumita and Iida 91] Sumita, E., and Iida, H. : Experiments and Prospects of Example-based Machine Translation, Proc. of the 29th Annual Meeting of the Association for Computational Linguistics, (1991).

[Tsujii and Fujita 91] Tsujii, J. and Fujita, K. : Lexical Transfer based on Bi-Lingual Signs -Towards Interaction during Transfer, : Proc. of the 5th Conference of the European Chapter of the Association for Computational Linguistics, (1991).