Spoken-Language Translation Method Using Examples

Hitoshi IIDA, Eiichiro SUMITA and Osamu FURUSE
ATR Interpreting Telecommunications Research Laboratories
2-2 Hikaridai
Seika-cho, Kyoto 619-02, JAPAN
{iida, sumita, furuse}@itl.atr.co.jp

1 Introduction

Conventional approaches to machine translation are mostly concerned with written text, such as technical documents. This paper addresses the problem of spoken-language translation and explains the method and its capability to handle spoken language.

2 Seven requirements for spoken-language translation

The following new design features are critical for success in spoken-language translation:

1. **Incremental processing**
   Incremental processing is required so as to handle fragmental phrases or incomplete utterances and to realize a real-time response. This has a very close relation with item 5 below.
2. **Handling spoken language**
   Fragmental phrases, isolated phrases, a gradient of case role changing, complex topicalization, metonymical phrases, idiomatic expressions for etiquette, and inconsistent expressions in one utterance are main characteristics of spoken language. They strongly depend on dialogue situations.
3. **Handling euphemistic expressions**
   Under the influence of social position or situation, euphemistic expressions appear in various scenes in various forms.
4. **Deterministic processing**
   Neither pre-editing nor post-editing can be relied on in a speech translation system. Interactive disambiguation by speakers does not necessarily converge a correct interpretation.
5. **Sufficient speed to avoid to break communication**
   As an interpreter intervenes between speakers, real-time response is required to keep smooth turn taking.
6. **High-quality translation**
   This is necessary in order to ensure correct information exchange between speakers.
7. **Recovering from speech recognition errors**
   There are various aspects to recovering from speech recognition errors, for example in correcting phoneme sequences, syllable sequences, word sequences (including compound words and collocations).

3 Meeting the seven requirements

3.1 Incremental processing

This is an essential technology if one is to build an incremental translation system like a simultaneous interpreter, and the proper way to grasp a chunk of a translation unit corresponding to some chunk in a target language is to extend 'constituent boundary parsing' to bottom-up-type parsing [Furuse96].

3.2 Recovering from errors

A certain recovery method is now under consideration: a re-entrizing model for phoneme candidates by means of searching the correct phonemes using modification depending on recognition error characteristics in an example-based framework [Wakita95]. This approach provides a recovery effect in handling phoneme or syllable sequences, and the effect depends on the particular speakers because of individual error characteristics.

3.3 Requirements covered by EBMT/TDMT

The remaining requirements are handled effectively by an example-based approach as explained here.

In NLP systems, especially for spoken language, many possibile syntactic structures are produced. It is an important and difficult process to choose the most plausibile structure. Conventional approachs, such as knowledge-based one, cannot easily handle continuous phenomena: gradation of case role changing; derivation of a metonymical relation; and relationship between a topicalized word and the main predicate.

We have proposed Example-Based Machine Translation (EBMT) to deal with these difficulties[Sumita92-a]. The EBMT method prepares a large number of translation examples; the translation example that most closely matches the input expression is retrieved; and the example is mimicked.

When applying EBMT to sentence translation, the sentence must be analyzed by matching transaltion patterns of phrases [Furuse94]. This model is in a sense "driven by transfer", and we call it Transfer-Driven Machine Translation (TDMT).

3.3.1 Handling spoken language

Spoken language includes many phenomena; here, however, we concentrate on the following ones:

(1) "wa" is a Japanese topic marker and, in general, this marker can be replaced by other case particles. But some usages cannot be identified as to case role because of gradation of case role changing. Moreover, if there are double topic markers in a sentence, they cannot be replaced by other particles[1]. The first sentence in our Japanese-to-English (JE) translation "snapshot" (Figure 1), for example, is properly translated in our TDMT prototype system.

   (i) "Chikatetsu-wa ichiban-chikai eki-wa doko desu-ka."
   ('subway-topicalized,' 'the nearest,' 'station-topicalized,' 'where,' 'be-question')

(2) Two sentences are mixed in one utterance. The first is pended, then immedaitely the second sentence starts without conjunction.

   (ii) "Shiharai-wa ginkou-furikomi-o o-machi-shite-orimasu."
   ('payment-topicalized,' 'bank-transfer-objective,' 'wait-for-polite-modest')

3.3.2 Handling euphemistic expressions

(1) There are various types of expressions for politeness, modesty, and euphemism. Such expressions are used depending on social roles. The fourth sentence in our Japanese-to-Korean (JK) translation snapshot (Figure 2) is a sample of this type, which is properly dealt with by TDMT.

   (iii) "Yoyaku-wo kakunin-sasete-itadaki-masu."
   ('reservation-objective,' 'confirm-modest')

   (iv) "Go-dengon-wo o-tutae-moushiage-masu."
   ('message-polite-objective,' 'inform-honorific')

[1]In this paper, sample Japanese sentences are written alphabetically and surrounded by double quotes, and the corresponding English words with usage modifiers follow in parenthesis.

3.3.3 Deterministic processing

Conventional MT methods provide multiple translation candidates but no information to use in selecting among them, or else just the first possible sentence that is generated.

On the contrary, EBMT generates all the possible candidates combining suitable phrases. It also provides proper scores to each candidate using a similarity calculation. The scores realize "deterministic" translation.

3.3.4 Speed

[Furuse96] has improved a matching mechanism over translation patterns. By accepting input in left-to-right order and dealing with best-only substructures, the explosion of structural ambiguity is restrained and an efficient translation of a lengthy input sentence can be achieved. Preliminary experimentation has shown that average translation times are reduced from 1.15 seconds to 0.55 seconds for input of 10 words in length and from 10.87 seconds to 2.04 seconds for input of 20 words in length. The incorporation of incremental morphological analysis and generation [Akamine95] into the new-version TDMT, is promising for achieving incremental (simultaneous) translation for a practical spoken-language translation system.

If instantaneous response is required, the rest dominant process is retrieval of the closest translation patterns from bulk collection. It is effectively solved by using a massively parallel algorithms and machines [Sumita95-a, Sumita95-b, Oi93].

3.3.5 Quality

First, a well-known difficult problem in Japanese to English translation was selected as a test. The Japanese noun phrase of the form "noun + NO + noun" using the Japanese adnominal particle "NO" is an expression whose meaning is continuous. A translation success rate of about 80% has been demonstrated in a Jacknife test [Sumita92-a]. Also, for other Japanese and English phrases, similar effectiveness in target word selection and structural dsiambiguation has been demonstrated[Sumita92-b].

We have evaluated a experimental TDMT system, with 825 model sentences about conference registration. These sentences cover basic expressions in an inquiry dialogue. The success rate is 71% for a test data set consisting of 1,050 unseen sentences in the same domain.

```
source : "ラスベガスには夜の九時までには着きたいんです"
target : "I would like to arrive at Las Vegas by nine o'clock at night"

source : "九時十五分のバスに乗られましたら夜の八時までには着きますが"
target : "If you get on the bus at nine fifteen , you will arrive by eight o'clock at night"

source : "八時ですか"
target : "At eight o'clock ?"

source : "じゃあだいたい十一時間掛かるんですね"
target : "Well , it takes eleven hours approximately , right ?"

source : "いいえ時差がありますから十二時間掛かります"
target : "No , there is the time difference and it will take twelve hours"
```

Figure 1: JE translation snapshot by TDMT

```
source : "こちらでホテルの予約はお願いできますか"
         (Hi, is it possible to make hotel reservation from here?)
target : "여기에서 호텔 예약은 부탁할 수 있습니까?"

source : "ご紹介はできるんですがご予約の方はお客様ご本人で直接していただきます"
         (OK, what we do is to give you all the information you need and then
          we ask you to go ahead and make the call yourself.)
target : "소개는 가능한데 예약은 손님 본인에서 직접 해 주시겠습니다."

source : "じゃあとりあえず市内のホテルを紹介してください"
         (OK, I'm looking for a central location, if possible.)
target : "그러면 우선 시내 호텔을 소개해 주십시오."

source : "宿泊費があまり高くなくて交通の便利なところがいいんですけれども"
         (Not too expensive, and it shouldn't take too long to get to the major sights from there.)
target : "숙박비가 너무 비싸지 않고 교통의 편리한 데가 좋은데요."
```

Figure 2: JK translation snapshot by TDMT

4 JE & JK prototype systems

The TDMT system is being expanded so as to handle travel arrangement dialogues including the topics of hotel reservation, room services, troubleshooting during hotel stays, various information queries, and various travel arrangements. At present the JE system has about a 5,000-word vocabulary and a transfer knowledge from 2,000 training sentences. The JK system is half this size. While some modules, such as morphological analysis and generation, are language-specific, the transfer module is a common part of every language pair. Through JE and JK implementation, we believe that the translation of every language pair can be achieved in the same framework using TDMT. On the other hand, it has turned out that the linguistic distance between source and target languages reflects the variety of target expression patterns in the transfer knowledge. Table 1 shows the number of target expression patterns corresponding a Japanese particles in JE and JK. These numbers are counted from the current TDMT system's transfer knowledge, and the numbers of examples are token numbers (i.e., not including duplications).

5 Discussion

5.1 Integration of Speech and Language

A mechanism for spontaneous speech translation must be consistent with a mechanism for handling associative knowledge, such as translation usage examples and word co-occurrence information for memory-based processing, and with a mechanism for logical structure analysis according to detailed rules for each processing phase in the Transfer-Driven MT processing. Under the process, a study should be carried out on building a stochastic language model using both syntactic and semantic information for speech understanding.

5.2 Related Research

On the other hand, some studies hope to build spoken language translation systems using a certain interlingua method. A semantic parser is a typical example of this method. In particular, "semantic pattern based parsing" in JANUS, CMU's speech to speech translation system [Woszczyna93, Levin95] uses frame based semantics with a semantic phrase grammar and the operation of the parser is viewed as "phrase spotting." Another one is MIT's multilingual

Table 1: Japanese particle translation in JE and JK translation

Japanese Pattern	JE		JK	
	Example	Target patterns	Example	Target patterns
X wa Y	224	30	66	1
X ga Y	140	15	40	1
X no Y	226	36	88	2
X o Y	147	15	41	1
X ni Y	154	22	55	5
X de Y	120	25	33	5

GALAXY: a human-language interface to online travel information [Goddeau94]. The system makes use of 'semantic frame representation' so as to paraphrase a recognized speech input utterance into a concrete and simple expression that conforms with one of the system's internal representations and makes the utterance meaning easy to handle. However, in extracting the meaning of an input sentence, many default values are required so as to execute heuristic inferences. The inference is too powerful in explaining a speaker's intention and the propositional content of the utterance by one key word or phrase. Such a method may work well in a certain domain, but less scalability may be revealed when making a larger prototype system.

VERBMOBIL is a typical translation system for face-to-face dialogue [Wahlster93]. This system adopts English as a dialogue language for human-machine interface and makes use of DRT-based semantic representation units.

6 Conclusion

TDMT has been proposed as a general technique for spoken-language translation. We have applied TDMT to two language pairs, i.e., Japanese-English, and Japanese-Korean, as a first step toward multi-lingual translation. Also, we are planning to integrate speech recognition with TDMT for achieving effective and efficient speech translation.

References

[Akamine95] Akamine, S. and Furuse, O.: Einichi-taiwabun-hon'yaku niokeru zenshinteki-nihongobun-seisei (Incremental generation of Japanese Sentence in English to Japanese Dialogue Translation), in Proc. of 1st NLP convetion, pp.281-284 (1995), (in Japanese).

[Furuse94] Furuse, O. and Iida, H.: Constituent Boundary Parsing for EBMT, in Proc. of COLING'94, pp. 105-111 (1994).

[Furuse96] Furuse, O. and Iida, H.: Incremental Translation Utilizing Constituent Boundary Pattern, in Proc. of COLING'96 (1996).

[Goddeau94] Goddeau, D., et al. : GALAXY: A HUMAN-LANGUAGE INTERFACE TO ON-LINE TRAVEL INFORMATION, in Poc. of ICSLP94, pp.707-710 (1994).

[Iida93] Iida, H. : Prospects for Advanced Spoken Dialogue Processing, IEICE TRANS. INF. and SYST., VOL. E-76-D, No.1, pp. 2-8 (1993).

[Levin95] Levin, L. , et al. : Using Context in Machine Translation of Spoken Language, in Proc. of TMI-95, pp. 173-187 (1995).

[Nagao84] Nagao, M. : A Framework of a Machine Translation between Japanese and English by Analogy Principle, in Artificial and Human Intelligence, eds. A. Elithorn and R. Banerji, North-Holland, pp. 173-180 (1984) .

[Oi93] Oi, K. et al. : Toward Massively Parallel Spoken Language Translation, in Proc. of the Workshop on Parallel Processing for AI, IJCAI'93, pp. 36-39 (1993).

[Sumita92-a] Sumita, E. and Iida, H. : Example-Based Transfer of Japanese Adnominal Particles into English, IEICE TRANS. INF. and SYST., VOL. E-75-D, No.4, pp. 585-594 (1992).

[Sumita92-b] Sumita, E. and Iida, H. : Example-Based NLP Techniques - A Case Study of Machine Translation - , Statistically-Based NLP Techniques - Papers from the 1992 Workshop, Technical Report W'92-01, AAAI Press (1992).

[Sumita95-a] Sumita, E. and Iida, H. : Heterogeneous Computing for Example-Based Translation of Spoken Language, in Proc. of TMI-95, pp. 273-286 (1995).

[Sumita95-b] Sumita, E. and Iida, H. : Heterogeneous Computing for Example-Based Translation of Spoken Language, in Proc. of TMI-95, pp. 273-286 (1995).

[Wahlster93] Wahlster, W. : Verbmobil: Translation of Face-To-Face Dialogs, in Proc. of MT-Summit IV, pp. 127-135(1993).

[Wakita95] Wakita, Y. et al. : Phoneme Candidate Re-entry Modeling Using Recognition Error Characteristics over Multiple HMM States, in Proc. of ESCA Workshop on Spoken Dialogue Systems, pp. 73-76 (1995).

[Woszczyna93] Woszczyna, M., et al. : RECCENT ADVANCES IN JANUS: A SPEECH TRANSLATION SYSTEM, in Proc. of EUROSPEECH'93, pp. 1295-1298 (1993).