2B-6

COMMENTS ON KOREAN ANALYSIS SYSTEM SHARING JAPANESE PARSER

Key-Sun CHOI and Kazunori MURAKI
C&C Information Research Laboratories, NEC Corp.

1. Introduction

This paper presents how the Korean syntactic analysis system has been developed for the NEC/PIVOT system. Korean is similar to Japanese in linguistic levels such as morphology (word-formation), syntax and semantics. Focusing on these similarities, our development of Korean system has been started by investigating what are not identical between two languages and adjusting them. Because the same phenomena of two languages can be handled in the same method, many parts of knowledge base and rules can be shared.

In the following sections, identical and different points between two languages will be discussed. Then, for each different point, the possible linguistic level for its handling will be pursued. This consideration will be linked to the problem what method is appropriate for machine translation system between two languages: direct, transfer, or pivot.

2. Comparison of Syntactic Structures between Japanese and Korean

The following two points had been considered before developing Korean analysis system under PIVOT system [1]:

- Korean is the closest language to Japanese.[2]
- Japanese analysis system has been already developed for PIVOT system.

First, similarities between two languages will be discussed.

2.1. Identical Points

Generally, the following common points can be indicated [2]:

(1) Word order is of SOV, and partially free-ordered: in literary style, predicate should be located at the end of a sentence, but other constituents can be ordered freely under some constraints.
(2) Postpositional characteristics: this is a characteristics of agglutinative languages. Postpositions are attached after word stems or nouns in surface, and they indicate the syntactic and semantic roles of word stems or nouns attached before them.
(3) Modifiers should come before modifiees.
(4) Ellipsis can occur easily.
(5) Honorific expressions are well-developed.

The above points of (1)-(4) show that a syntactic parser is sharable, if the following knowledge is separable from parser: how to map surface forms of postpositions into their corresponding deep roles in syntactic and semantic levels.

If the roles of functional words including postpositions show one-to-one correspondences between two languages, direct translation is enough to translate them each other. In order to check this possibility, differences should be checked seriously.

2.2. Different Points

Among several kinds of functional words, case markers, auxiliaries and their case shifts will be investigated.

2.2.1. Case Markers

Here, among several kinds of postpositions, the correspondencies of obligatory and optional case markers between two languages will be shown. Since obligatory case markers are selected depending on predicates, directly translated Korean ones of Japanese predicates are selected from entries of several Japanese-Korean dictionaries [3,4,5]. In

case of optional case markers, they indicate semantic roles independently of predicate, its investigation has used example sentences of the above dictionaries and Japanese grammar books written in Korean [2]. As shown in figure 1, this investigation resulted that one-to-one correspondence cannot be hold, and such similar result is also found in [6]:

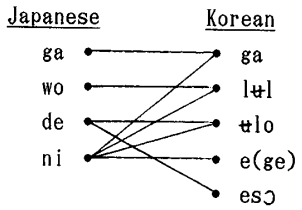

Figure 1. Comparison of Case Markers between Japanese and Korean.

Here, "e(ge)" represents two case markers "ege" and "e" which are differentiated by semantic features: [+ANIMATE] and [-ANIMATE].

2.2.2. Auxiliaries and Case Shifts

Auxiliaries show different phenomena between two languages in three points. Firstly, direct Korean translation of some Japanese auxiliary verbs show different case shifts in Korean. Case shift means how obligatory case markers are shifted to others by attaching auxiliaries to a predicate. Typical examples are auxiliaries which occur in "passivization" or "causativization" phenomena. These phenomena of Korean show peculiar characteristics [7] and different case shifts from Japanese or English. This handling should be based on lexicon, for there is no rule for verb to take passivizer or causativizer in Korean. Because their surface forms are different depending on more detailed subfeatures and some verbs cannot take passivization or causativization, these handling should be based on lexicon and more detailed categorized subfeatures of auxiliaries.

Secondly, although some Japanese auxiliary verbs have representative corresponding Korean words, their distribution is different because of different constraints. For example, many Korean verbs reject the use of direct translation of Japanese ones. The direct Korean translation "badda" for a Japanese auxiliary "morau" is not used in the same usage. Instead of "badda", "juda" is usually used which is a directly translated word of Japanese "kureru". This phenomena can be overcome by lexicon or some transfer method.

Finally, there is no directly translated auxiliaries. They should be expressed in other part-of-speech or idiomatic expression. For example, consider "nareru" of Japanese. "tsukai-nareru" is translated to "son-e igda" in Korean whose direct Japanese translation is "te-ni nareru", but Japanese "ki-nareru(着慣れる)" is translated to Korean "ib-osɔ gildɯlyɔjida" whose direct Japanese translation is "kite-kara nareru". This phenomena requires idiomatic handling in case of Korean.

3. Conclusion

Identical points between two languages imply that identical parser can be used for both languages, for they are global syntactic phenomena. Different points indicate that some functional words cannot be translated directly. This leads the fact that direct translation is impossible between Japanese and Korean.

By changing only the finite number of functional words and sharing the same syntactic parser by minor tuning, Korean analysis system has been developed.

References

[1] Muraki, K. and S. Ichiyama (1988.9) 「PIVOT:多言語間翻訳システム概要」、昭和63年電子情報通信学会秋季全国大会予稿集.

[2] Kim, K.H. (1987) 標準日本語文法、進明出版社, Seoul.

[3] 金星出版社、小学館 (1987) New Ace 日韓中辞典、金星教科書、Seoul.

[4] 東亜出版社編集部 (1979) 現代日韓辞典、東亜出版社、Seoul.

[5] 日本国際交流基金 (1987) 基礎日本語学習辞典、時事英語社、Seoul.

[6] 田窪行則 (1987.6) 「誤用分析3：自転車を乗って行きました」、日本語学、明治書院、pp.84-89.

[7] Lee, J.H. et al. (1988.9) "Sentence-Structure Planning for Natural Expressions of Korean", 昭和63年情報処理学会秋季全国大会予稿集.