Syntactic Analysis of Coordinate Noun Phrase

CHEOL HO KIM
Institute for Defense Information Systems, Seoul, Korea

GIL CHANG KIM
Korea Advanced Institute of Science and Technology, Taegon, Korea

Abstract

One of the major applications of the machine translation system is translation of scientific and technological writings. Compared with other kinds of texts, coordinate structures are frequently used in these texts. Coordination has received a great deal of attention in the linguistic literature. This paper describes an attempt to analyse Korean sentences into conceptual representations, especially focusing on the coordinate noun phrase. To obtain more natural translation results, proper determination of the scopes of coordinate structures and elimination of ambiguities are important.

We study the various types of coordination in order to determine scope in coordinate noun phrases and propose a scheme to eliminate the ambiguities in syntactic analysis using heuristic methods applied to surface forms of sentences and distinctive features of the Korean coordinate noun phrase.

1. Introduction

The Korean analysis system which has been developed in a joint KAIST-NEC machine translation project is a multilingual machine translation system between Korean, Japanese and Korean, and is a part of the VENUS machine translation system (Muraki, 1986) developed by NEC. From a Korean input sentence, the analysis system extracts a language-independent conceptual structure as an internal meaning representation. The generation system synthesizes a Korean sentence from a conceptual structure (Kim et al., 1988; Lee and Kim, 1989).

Machine translation has focused on the scientific and technological writings. These tend to use coordinate structures frequently for conciseness. Coordinate noun phrases are the main cause of ambiguities in syntactic analysis in a machine translation system because of the many possibilities concerning scope. In this paper, we formulate a characteristic of coordinate noun phrases and syntactic information. We also propose a scheme based on these heuristics to eliminate ambiguities in determining the scope of coordinate structures, especially focusing on coordinate noun phrases.

The optimal grammar formalism is not defined for syntactic analysis of Korean. However, dependency grammar is frequently used for the syntactic analysis of relatively free word-order languages like Korean, Japanese (Muraki, 1988), Finnish (Japprin et al., 1986) and Greek (Sgall and Paneva, 1987). Dependency grammar represents the syntactic relationship between two grammatical elements as that of modifier and modified (or head and dependent). To represent semantic relationships, case grammar is combined with dependency grammar in our Korean analysis system.

1.1 Stages of Analysis

The analysis system, where a Korean input sentence is analysed and transformed into its corresponding language-independent conceptual structure is organized into three phases.

Phase I (Morphological Analysis) analyses an input sentence. All morphemes are identified as keywords of a dictionary. The corresponding entries in the dictionary are loaded. Several morphemes are then combined into semi-phrases using the connectivity information. These semi-phrases are the basic units in the subsequent analyses.

Phase II (Syntactic Analysis) constructs a language-independent semantic structure from a semi-phrase list. The syntactic analysis is based on the dependency grammar framework (Muraki et al., 1985; Kodama, 1987).

Phase III (Conceptual Analysis) transforms a semantic structure into its corresponding conceptual structure by using the knowledge base and the structural transformation rules.

1.2 Problems in Analysis of Coordinate Structure

Problems in the syntactic analysis of coordinate structures have long been a topic of interest to both the language, and computational linguists because it causes many ambiguities. Eliminating ambiguities in coordinate structures is one of the crucial problems in machine translation. If the ambiguities are not cleared up, one cannot expect to produce machine translations of reasonably good quality. There are several problems in the syntactic analysis of coordinate structures.

Since a coordinate structure consists of several constituents with the same syntactic role in a sentence, coordinate noun phrases connected by conjuncts should be correctly detected to precisely interpret the relationship within the coordinate elements. In syntactic analysis, incorrect detection of the coordinate elements may cause many dependency relations which lead to ambiguities. Different from English, Korean has various coordinating conjunctions to construct coordinate structures. Using the patterns of coordinate structures from surface expressions in Japanese, an attempt to formulate heuristics has been made (Shudo et al., 1986). Korean is very similar to Japanese..
In linguistic features, but is considerably different from Japanese in the surface expression of coordinate structures. Thus, it would be helpful to analyse the patterns of coordinate noun phrases. The patterns of coordinate noun phrases in Korean would be surveyed in this paper and are different from those in Japanese. Based on the patterns, information for the syntactic analysis of coordinate structures is derived.

Determining the analysis of a coordinate noun phrase is mainly dependent on the precise meaning of each noun. In machine translation, a detailed description of the meaning of each noun is not usually available. Although the semantic information is not available, the patterns of surface expressions and the available syntactic information are used to determine the analysis of a coordinate noun phrase and eliminate ambiguities.

When two or more conjunctions reside within one noun phrase, it is very difficult to decide the scope of each conjunction to give the right meaning. To resolve this problem, the lexical item, the level of a coordinating conjunction which represents the scope of each conjunction, can be assigned and used in syntactic analysis.

This paper describes the characteristics of coordinate structures in Korean, and the heuristic information to determine the scope of coordinate noun phrases in syntactic analysis phase of machine translation system. We propose a scheme for eliminating ambiguities using this heuristic information. These heuristics are based on the characteristics of sentence structures and the syntactic features of conjunctions. Korean sentences are represented in the Yale Romanization.

2. Coordinate Structure in Korean

While English is a rigid word-order subject verb object (SVO) language, Korean word-order is relatively free. Although the normal order is subject object verb (SOV), any noun phrase in the same clause can appear anywhere as long as the verb comes at the end. According to the word-order typology (Cenner, 1981), Korean is a flexible SOV language with a variety of post-positional particles, and a long O (object) which may contain prepositional modifiers between Subject and Verb (V). A coordination element used for the long object causes ambiguity in determining the scope of coordination.

2.1 Coordinate structures of Korean Compared with English

2.1.1 Classification of Coordination. Coordination is divided into physical coordination and clausal coordination. Coordination in English is constructed using the conjunctions 'and', 'or', etc. in phrase and clausal coordination. It is not easy to recognize the syntactic function of the coordination elements in a sentence.

(1) Coordination in English:
(a) Young boys and girls (noun phrase)
(b) Young and old people (adjective phrase)
(c) He climbed up and over the wall (prepositional phrase)
(d) Kim sang and Sandy danced (clause)

(2) Coordination in Japanese:
(a) Daitokoro - or - shingyoudani (noun phrase)
(b) Genbouso - or - jitsukato (propositional clause)

We analyse the phenomena and compare with the medians.

(3) Coordination in Korean:
(a) Calycheul - or - simbogoki (noun phrase)
(b) Kim - or - Sandy (prepositional phrase)

2.1.2 Coordination Scheme. There are two sorts of coordinate constructions in English (Sag et all., 1981). In one, there can be only two conjuncts. In the other, there is no limit to the number of conjuncts permitted. They are exhibited in (4a) and (5a). The first is for the arbitrary-length coordination structures, and the second for the binary ones. In (4b) and (5b) we list the values that the CONJ feature has in English for the two constructions. One further parochial component of coordination is needed in order to be able to express the claims it makes concerning the structure of English.

We need a LP (Linear Precedence) statement to express the ordering constraints that hold across the various types of conjunct characterized by distinct values for the feature CONJ. These LP statements can be collapsed into a single schema as (6).

(4) Iterating Coordination Schema:
(a) X &gt;&gt; [CONJ a0, [CONJ a1, ...]]
(b) a &gt; <(and, NIL), <(and, and), ... <(neither, or), <(or, NIL), <(NIL, or)>

Here, H Head

CONJ Conjunction
a0 All conjuctions before the last conjunction
a1 The last conjunction
NIL No conjunction

connective postpositional particles, connective adverbs and punctuation marks. Connective postpositional particles are "wa/kwa" ("kwa is an allomorph of 'wa'), 'hako', 'ina', 'ima', 'imye', 'ita', 'iken', 'iko', 'i lang', 'ey(laka)", and so on. The connective postpositional particles have the meanings of enumeration, selection, and together with as well as the pure function of connecting noun phrases. "wa/kwa" is proper to the pure function. Meanwhile, 'ima', 'imye' and 'ita' suggest enumeration, 'ina', 'iken' and 'iko' suggest selection and 'i lang' suggests together with. Connective adverbs are 'to(mun)', 'michi', 'hokum', 'kuikiko', etc. Punctuation marks are comma(,) and centered period(·). The coordinating conjunction for a verb phrase is connective inflection. Sentences are connected by the sentential adverbs.

A frequency analysis of coordinating conjunctions in noun phrases is the abstracts from the Journal of Korean Information Science Society, vol 11(1)-15(3) is shown in Fig. 1.

<table>
<thead>
<tr>
<th>Part of Speech</th>
<th>Conjunction</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pospostion Parale</td>
<td>wa/kwa</td>
<td>173</td>
</tr>
<tr>
<td></td>
<td>hako</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>ina</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>imun</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>michi</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>niken</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>kuikiko</td>
<td>0</td>
</tr>
<tr>
<td>Advach</td>
<td>-</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Punctuation Mark</td>
<td>-</td>
<td>341</td>
</tr>
</tbody>
</table>

Fig. 1: Frequency of conjunctions in coordinate noun phrase

2.2.2 Coordination Patterns. The patterns of coordinate noun phrases and their frequencies are shown in Fig. 2. So far we have observed seventeen patterns of coordinate noun phrases connected by conjunctions including punctuation marks. Of the coordinate noun phrases 85% are connected by 'wa' and 'michi'.

Using the similarities among coordination elements in these patterns, we can reduce the ambiguity in coordinate noun phrases. Recognizing the common words or expressions in coordination elements makes easy to determine the scope of coordination (Shudo et al., 1986). From this information, considerable reduction in ambiguity is expected. We can also recognize the expressions which determine the scope of coordination, such as 'i lang' after the last coordination element. Using these surface expressions and patterns from Fig. 2, we develop the heuristics in Section 3.2.
2.3 Hierarchy in Coordinating Conjunctions

When there are two or more coordinating conjunctions within one noun phrase, we must decide which conjunctions mark the same level of coordination.

(9) tangeunseong-uy [(sicem-congcong) ky-syncung mich
[yuung-muwoeng] puwunbyul bad sawi ssa-um
atollicum-ul kuyalha-na.

(A comparatively simple algorithm is developed for determining startpoints and endpoints and classifying voiced and voiceless sounds in speech.)

In the above example, the coordinate noun phrases 'sicem-congcong ky-syncung' and 'yuung-muwoeng puwunbyul', constructed by 'mich', have nested coordinate noun phrases 'sicem-congcong' and 'yuung-muwoeng' respectively. This is at a higher level in the hierarchy than 'mich' in this case. This hierarchy is similar to the operator precedence for arithmetic expressions. Without careful recognition of coordinating conjunctions, we may have many interpretations as in analysing an arithmetic expression.

Explained in Section 2.2, there are a diversity of coordinating conjunctions in Korean, unlike English. Combinations of them are used in a sentence. Therefore, the lexical item is required to represent the level of coordinating conjunction in Korean, and is used to determine the scope of a coordinate structure. The level of conjunction is represented as 1, 2, and 3 in Fig. 3, where 1 denotes the widest scope.

3. Heuristics in Coordinate Noun Phrases

3.1 Ambiguities in Coordinate Noun Phrases

Coordinate noun phrases often appear in abstracts of scientific and technological papers. It is important to analyse them correctly, especially to determine the scope of coordination because they often lead to proliferation of analyses. The prepositional particle 'wa' 'kwa' plays almost the same role as the English 'and' in connecting noun phrases.

(10) kulayphy mohyang-4 [phulokalymuy hapseng
kwa sayang-uy kungung-ey iyongtoy swa
in-tyu.

(Graph model is used to synthesize the program and verify the specifications.)

In example (10), two coordination elements 'phulokalymuy hapseng' and 'sayang-uy kungung' are connected by the coordinating conjunction 'kwa'. The possible coordination elements from the first one are 'phulokalymuy hapseng' and 'hapseng', and the possible coordination elements from the second one are 'sayang-uy kungung' and 'sayang'. Thus, there are four analyses originating from different scopes of the conjunction. Figure 4 depicts the possible ambiguities from these analyses.

Since coordinate structures include many elements with the same syntactic role in a sentence, there is no grammar formalism which defines the relationship between elements.

Fig. 2 Patterns of coordinate noun phrases. NP = noun phrase

<table>
<thead>
<tr>
<th>Number of Elements</th>
<th>Pattern</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Elements</td>
<td>NP-\text{or} NP(=\text{or})</td>
<td>140</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>NP</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>NP-\text{or} NP</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP-\text{or} NP</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP-\text{and} NP</td>
<td>1</td>
</tr>
<tr>
<td>More than 3 Elements</td>
<td>NP-\text{and} NP</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP-\text{and} NP</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP-\text{and} NP</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP-\text{and} NP</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP-\text{and} NP</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>NP-\text{and} NP-\text{and} NP</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 3 Levels of conjunction

<table>
<thead>
<tr>
<th>Level</th>
<th>Conjunction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>mich, neori, heon, tabko</td>
</tr>
<tr>
<td>2</td>
<td>walkssu, lako, chan, ih, is, ilmo, tabksu</td>
</tr>
<tr>
<td>3</td>
<td>tuyo, mi, in, ih, di, ide, ey,</td>
</tr>
</tbody>
</table>
among them. No computational model has been known to resolve the scope determination problem in interpreting coordinate structures so far. Thus, we develop heuristics based on syntactic information in the analysis phase as well as information from surface patterns. And we show that the heuristics can be applied to determine the proper analysis of coordinate noun phrases.

3.2 Heuristics in Determination of Scope

To determine the proper analysis of the several possible ones we should, in general, consider not only information at the surface and syntactic level but also at the semantic level. However, it is impossible to disambiguate completely because natural language has ambiguity in itself. And just surface level information is sometimes used to determine the analysis of coordinate noun phrases in Japanese (Nagao et al., 1985; Tsuji et al., 1984).

For the above reasons, we try to determine the analysis of coordinate noun phrases by using heuristics derived from surface and syntactic information. Heuristics 1–4 below have been applied to Japanese and can also be applied to Korean. In addition, some heuristics specific to Korean, with examples, are also described below.

(Heuristic 1) Postpositional particle ‘wa’ may also be used as a case marker. If it appears in the position:

\[ \text{Noun} + \text{Verb} + \text{Noun}, \]

\[ \text{Noun} + \text{Adjective} + \text{Noun}, \]

the postpositional particle ‘wa’ is not a coordinating conjunction as a case marker, if the verb or adjective is one of those which require an argument with the surface case marker ‘wa’, and there are no extra words between the ‘wa’ and the verb or adjective. There may be two possible interpretations, one in which ‘wa’ is a case marker and ‘Noun + Adjective (Adjective)’ forms a relative clause that modifies the second noun, and another in which ‘wa’ is a conjunction in a coordinate noun phrase.

(11) yaweche-bi-wa ra-un kap

produced name  different value

(The value which is different from the predicted value.)

This heuristic is not for determining the analysis of scope in a coordinate noun phrase. (11) can be interpreted as coordinate noun phrase. But the heuristic prevents it from being so interpreted.

(Heuristic 2) If two ‘wakuno’ appear in the position:

\[ \text{Noun}_1 + \text{wa} + \ldots + \text{Noun}_2 + \text{wakuno} + \text{Noun}_3, \]

the boundary of the second element of the coordination is always Noun_2. The second ‘wakuno’ plays a role of delimiter for the second element of the coordination. Coordination elements are represented by ‘]’.

(12) [donkapsi-kwa man yin-si-kei-sen-yun-wa et-e-nan

The difference between theoretic value and simulation result

kye-wakuno-wa wakuno chail-hul

The second ‘wakuno’ is optional, but it often appears when the second coordination element is a long noun phrase and its scope is ambiguous without it.

(Heuristic 3) The scope of the coordinating conjunction ‘-i’ is a single noun.

(13) please help

han hankul . . .

(Design and implementation of Hangul system on personal computer.)

(Heuristic 4) In \[ X_1 + \text{Conj} + X_2 \] when the first word of \[ X_1 \] and \[ X_2 \] is the same or the last word of \[ X_1 \] and \[ X_2 \] is same, the first and the second coordination element are \[ X_1 \] and \[ X_2 \] respectively. Here the \[ X_1 \] is series of words other than conjunction, and Conj is a conjunction.

(14) ces-toy-mon an likyeye-yey [konglukke-ke

This model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.

konglukke-ke]

The model is divided into a machine-dependent part and a machine-independent part.
The design and implementation of HW and SW protocol for Local Network Adapter is described.

Heuristic 6: Coordinating conjunction such as 'mich', 'hong', and 'kalik' are always used as the last conjunction.

Heuristic 7: In the coordination pattern:
Noun, + noun, + Conjunction + Noun, + noun, + noun,
the conjunction is 'Noun, + Conjunction + Noun, and the second is 'Noun, + noun, + Noun, + Conjunction + Noun, + noun, + noun,'. Examples of success and failure for this heuristic follow. The successful analysis is represented by '☑'; the other is underlined.

Heuristic 8: In Noun, + Noun, + ... + Noun, + Conjunction + X, Noun, + Noun, + ... + Noun, is considered a conjunction. This heuristic may be ambiguous when there are more than two coordinations in a sentence. (Heuristic 7) is applied to (18) also.

Heuristic 9: A series of nouns is treated as a single noun.

Heuristic 10: In case of X, + Conjunction + X, + noun, the second coordinate is X, the same holds for the case marker 'kon' instead of 'noun'.

To solve the problem in (15), we give higher priority to (Heuristic 9) than (Heuristic 10). These heuristics are based on different levels of information (some are based on surface lexical items, some are based on syntactic information) and may lead to different decisions about the analysis of scope. Heuristics 1-6 almost always succeed, and 7-10 often lead to wrong decisions. Although the surface structure of English is different from that of Korean, heuristics based on surface lexical items have been used (Shibata et al., 1995) to correctly determine the analysis of scope in coordinate phrases in English without deep semantic analysis.

4. Syntactic Analysis of Coordinate Structure

4.1 Analyzing Coordinate Noun Phrase
The syntactic analysis phase determines coordination elements related to conjunctions, applies heuristics to input chart, and, if necessary, determines the level of coordination. Figure 5 shows the overall stages of the coordinate noun phrase analysis.

4.1.1 Loading Chart Information. A chart is the output of morphological analysis which is loaded to analyze the syntactic dependency in coordinate structure. Each edge of this chart represents a semi-phrase or word. Each edge is linked to its corresponding partial dependency structure (PDS). A PDS contains the result of morphological analysis where word corresponding to each edge are analyzed. The structure of a PDS is a tree structure where each node is a set of feature-value pairs. The contents of these pairs are loaded from the dictionary.

4.1.2 Detecting Candidate Conjunctions. The coordinating conjunction is marked first at this stage. Since Korean has a pronominal modifying structure, detection of candidates proceeds from the coordinating conjunction to the right.

In "A-wa B-ja C-ey D-han B-eul", the first element of coordination connected by "wa" is determined as A. B and E are alternatives for the second element. To choose the correct one, we apply the non-crossing principle, and then compare meaning classification of nouns for B and E respectively with that of A. Thereby, the noun with the nearest value of classification to A is selected as the candidate for the coordination element.

4.1.3 Determining Coordination Level. To determine the coordination level of a conjunction in a sentence, with two or more conjunctions, the lexical information on each conjunction is used. Marks denoting the start node and the end node of a coordinate structure are given at the beginning of this stage in order to construct the dependency relation.

4.1.4 Determining the Scope. Heuristics with a high success rate are applied first to the chart to determine the extent of the coordination element. If the extent is not determined by the heuristics, semantic features are enforced to do it. Figure 6 shows the applying sequence of heuristics that is the priority of heuristic.

4.2 Constructing Dependency Relation
As the syntactic analysis goes on, edges covering more than one semi-phrase are formed. In other words, these new edges cover more than one edge which covers a narrower range. When an edge connecting the left-most and right-most nodes of a chart is formed, syntactic analysis is terminated.

4.2.1 Selecting the Head. A possible pair of a syntactic head and its dependent is detected on the basis of their features.

4.2.2 Predicting Conceptual Relations. A set of permissible conceptual relations between head and dependent are predicted, using their positional features, phrase-structural features, case-structural features, and so on.

4.2.3 Verifying Conceptual Relations. The knowledge-base inference mechanism is activated using the conceptual primitives in corresponding conceptual information for each head and dependent of the predicted permissible relation. Then the most appropriate conceptual relation is selected and verified.

4.2.4 Constructing Semantic Dependency. Finally, we can construct a dependency structure between the head and its dependent using the features if the knowledge base returns that semantic interpretation of the two are consistent. In other words, if a consistent conceptual relation is found between their conceptual primitives.

5. Conclusion
It is important to eliminate ambiguities of scope in coordination in a machine translation system. This paper has been described as a coordinate noun phrase disambiguation model which determines the scope of coordinating elements. Especially, we suggested that heuristics from surface-level information worked very well to eliminate ambiguities without recourse to deep semantic analysis.

Compared with English, Korean has complex conjunctions such as connective postpositional particles, connective adverbs, and punctuation marks. Based on the patterns of the coordinate structures, we derived heuristics inherent to these expressions in Korean. By integrating surface and syntactic-level heuristic information, our analysis system efficiently produced a fairly
natural and preferable reading as output, without any extensive semantic processing.

Finally, a computational model for coordinate mean phrase analysis under the interlingual machine translation approach has been proposed, which makes it possible to determine the scope of coordinating elements correctly. We have considered only surface and syntactic information, which does not completely eliminate the ambiguities. Thus, further research on heuristics at the semantic and the pragmatic level may complement the weak points of this paper.

References


Literary and Linguistic Computing, Vol. 7. No. 3. 1992


