

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,
Taejeon, 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 linguistics 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 English, 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, 1988).

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 some heuristics from surface expressions 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, 1986), Finnish (Jappinen *et al.*, 1986) and Czech (Sgall and Panevova, 1987). Dependency grammar represents the syntactic relationship

Correspondence: Cheol Ho Kim, Institute for Defense Information Systems, P.O. Box 270, Cheongryang, Seoul 130-650, Korea. Fax +82-2-965-3295.

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-dependent 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

The analysis of coordinate structure has long been a topic of interest to theoretical 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 coordinate elements with the same syntactic role in a sentence, coordinate noun phrases connected by conjunctions 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 surface sentences and 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 (Comrie, 1981), Korean is a flexible SOV language with a variety of post-positional particles, and a long O(bject) which may contain prenominal modifiers between S(subject) and V(erb). A coordination element used for the long object causes ambiguity to determining the scope of coordination.

2.1 Coordinate structures of Korean Compared with English

2.1.1 *Classification of Coordination.* Coordination is divided into phrasal coordination and clausal coordination. Coordination in English is constructed using the conjunctions 'and', 'or', etc. in phrasal 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)

Phrasal coordination types in English are noun phrase, prepositional phrase and adjective phrase. To determine

the scope of coordination, the syntactic category of conjuncts must be understood. In Korean or Japanese, there are no phrasal coordinations except noun phrase coordination, which is constructed by using a post-positional particle or adverb. Clausal coordination is represented differently from phrasal coordination and is constructed by conjugation of verb and adjective, i.e. connective inflections for coordinated predicates. Because of this, we can easily recognize phrasal and clausal coordination in Korean and Japanese.

(2) Coordination in Japanese:

- (a) Detashori -to shingoushori
(noun phrase)
Data processing conjunction signal processing
(Data processing and signal processing.)
- (b) Genshou-wo kaishaku-shi, jissoku-to
Phenomena analyse inflection measurement
hikakushi-ta (clause)
compare
(We analyse the phenomena and compare with the measurement.)

(3) Coordination in Korean:

- (a) Calyocheli -wa sinhocheheli
(noun phrase)
Data processing conjunction signal processing
(Data processing and signal processing.)
- (b) Kim-i nolayhu-ko Sandy-ka
sing inflection
chwmuchu-nta (clause)
dance
(Kim sang and Sandy danced.)

2.1.2 *Coordination Schemata.* There are two sorts of coordinate construction in English (Sag *et al.*, 1984). In one, there can be only two conjuncts. In the other, there is no limit to the number of conjunctions 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 explore 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) $X \rightarrow H[CONJ \alpha_0], H[CONJ \alpha_1]^*$
 - (b) $\alpha \in \{<and, NIL>, <NIL, and>, <neither, nor>, <or, NIL>, <NIL, or>\}$
- | | | |
|-------|------------|--|
| Here, | H | Head |
| | CONJ | Conjunction |
| | α_0 | All conjunctions before the last conjunction |
| | α_1 | The last conjunction |
| | NIL | No conjunction |

- (5) Binary Coordination Schema:
 (a) $X \rightarrow H[CONJ \alpha_0], H[CONJ \alpha_1]$
 (b) α is in {<both, and>, <either, or>, <NIL, but>}
 (6) Linear Precedence Schema:
 $[CONJ \alpha_0] < [CONJ \alpha_1]$
 where $\alpha_0 \in$ {both, NIL, either, neither},
 and
 $\alpha_1 \in$ {and, but, nor, or}

In Korean, Binary Coordination Schema is a residual schema because it is included in Iterating Coordination Schema (Kim, 1987). The selection of pairs of conjunctions in (7) requires careful consideration of the syntactic constraints and the meaning. (7) and (8) show the Korean coordinate schemata extended by including connective adverbs. Noun phrase coordinations are combined as in (7) and their order in the combination is determined as in (8). This accomplishes the coordinate function.

- (7) Iterating Coordination Schema:
 (a) $X \rightarrow H[CONJ \alpha_0], H[CONJ \alpha_1]^*$
 (b) $\alpha \in$ {<wa, wa>, <hako, hako>, <ina, ina>, <ini, ini>, <ey(taka), NIL>, <imye, imye>, <ita, ita>, <iken, iken>, <iko, iko>, <ilang, ilang>, <tto(nun), tto(nun)>, <mich, mich>, <hokun, hokun>, <kuliko, kuliko>, <wa, NIL>, <hako, NIL>, <imye, NIL>, <ini, NIL>, <ina, NIL>, *<iken, NIL>, *<iko, NIL>, *<ilang, NIL>, <tto(nun), NIL>, <mich, NIL>, <hokun, NIL>, <kuliko, NIL>}

- (8) Linear Precedence Schema:
 $[CONJ \alpha_0] < [CONJ \alpha_1]$
 where $\alpha_0 \in$ {wa, hako, ina, iken, ini, imye, ita, ey(taka), iko, ilang, tto(nun), mich, hokun, kuliko}
 $\alpha_1 \in$ {wa, hako, ina, iken, ini, imye, ita, ey(taka), iko, ilang, tto(nun), mich, hokun, kuliko, NIL}

Let us examine the difference between 'and' and 'wa'. In Korean 'wa' plays the same role as 'and' in English. For example,

A-wa B-wa C-wa D(-wa)ka \rightarrow A, B, C-wa D-ka

is not same meaning (Kim, 1970). The latter means a group of A, B, and C confronted with D. It has different meaning that every A, B, C, and D are participated in coordination with same competency. On the other hand, 'A, B, C mich D' is used as the same meaning. As we see in the example, 'wa' in the middle of coordination cannot be omitted. The omission of them leads to a different meaning.

2.2 Types of Coordinate Structure

2.2.1 Coordinating Conjunctions. A coordinating conjunction is an expression which connects multiple noun phrases, verb phrases and sentences. Coordinating conjunctions for connecting noun phrases are

connective postpositional particles, connective adverbs and punctuation marks. Connective postpositional particles are 'wa/kwa' ('kwa' is an allomorph of 'wa'), 'hako', 'ina', 'ini', 'imye', 'ita', 'iken', 'iko', 'ilang', 'ey(taka)', 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, 'ini', 'imye' and 'ita' suggest enumeration, 'ina', 'iken' and 'iko' suggest selection and 'ilang' suggests together-with. Connective adverbs are 'tto(nun)', 'mich', 'hokun', 'kuliko', etc. Punctuation marks are comma(,) and centred 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 in the abstracts from the *Journal of Korea Information Science Society*, vols 11(1)-15(3) is shown in Fig. 1.

Part of Speech	Conjunction	Frequency
Postposition Particle	wa/kwa	173
	hako	0
	ina	6
Adverb	tto(nun)	5
	mich	62
	hokun	2
	kuliko	7
Punctuation Mark	,	76
	.	10
Total		341

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 phrase connected by conjunctions including punctuation marks. Of the coordinate noun phrases 85% are connected by 'wa' and 'mich'.

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 'tung' after the last coordination element. Using these surface expressions and patterns from Fig. 2, we develop the heuristics in Section 3.2.

Number of Elements	Pattern	Frequency
2 Elements	NP-wa NP(-wa)	140
	NP mich NP	48
	NP, NP	11
	NP+NP	8
	NP-ina NP	5
	NP(,) to(nun) NP	5
	NP hokun NP	2
	NP-ina NP-wa	1
	NP kuliko NP	1
More than 3 Elements	NP(, NP) ¹ -wa NP	23
	NP(, NP) ¹ mich NP	10
	NP(, NP) ¹ kuliko NP	6
	NP-wa NP mich NP	3
	NP-wa NP(, NP) ¹	2
	NP-wa NP(-wa NP) ¹	2
	NP-wa NP, kuliko NP	2
	NP+NP+NP	1
Total		270

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

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) taneumseng-uy [[sicem-congcem] kyelceng mich [yuseng-mwuseng] pwunlyu]-hul hal swu iss-nun alkolicum-ul kaypalha-nta.

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

In the above example, the coordinate noun phrases 'sicem-congcem kyelceng' and 'yuseng-mwuseng pwunlyu', constructed by 'mich', have nested coordinate noun phrases 'sicem-congcem' and 'yuseng-mwuseng' respectively. '.' 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 postpositional particle 'wa/kwa' plays almost the same role as the English 'and' in connecting noun phrases.

- (10) kulayphu mohyeng-i [phulokulaym-uy hapseng-kwa sayang-uy kumcung]-ey iyongtoy! swu iss-ta.

(Graph model is used to [synthesize the program and verify the specification].)

In example (10), two coordination elements 'phulokulaym-uy hapseng' and 'sayang-uy kumcung' are connected by the coordinating conjunction 'kwa'. The possible coordination elements from the first one are 'phulokulaym-uy hapseng' and 'hapseng', and the possible coordination elements from the second one are 'sayang-uy kumcung' 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

Level	Conjunction
1	mich, to(nun), hokun, kuliko
2	wa/kwa, hako, iten, iko, ina, stohan
3	imye, ini, ita, ilang, '1', '2'

Fig. 3 Levels of conjunction

(15) hankul teitha-uy [cenguy, cocak, ciluy *mich*
definition, manipulation, query
 pokosecakseng]-tung-uy
report generation
 kinung-ul ceykongha-nta.

(This system provides facilities for data definition, manipulation, query and report generation of Hangul data.)

(Heuristic 6) Coordinating conjunction such as 'mich', 'ttonun', and 'kuliko' are always used as the last conjunction.

(16) phothulan sosu phulokulaym-ey tayhan tongcek
 theysuthu-lul haycwu-nun thul-un
 [khontulol tulaipe, thesuthu sayngsengki,
control driver test generator
 simyulleyithe *mich* tongcek thesuthu]-lo
simulator dynamic tester
 kwusengtoy-e program behavior-ey kwan-han
 cengpo-lul ceykonghaycwu-nta.

(Automated tools for testing of Fortran source code include test drivers, test data generators, simulators, and dynamic testers. They provide information on program behaviour.)

(Heuristic 7) In the coordination pattern:

Noun₁+uy+Noun₂+Conj+Noun₃+uy+Noun₄,

the conjunct is 'Noun₁+uy+Noun₂', and the second is 'Noun₃+uy+Noun₄'. Examples of success and failure for this heuristic follow. The successful analysis is represented by '[]', the other is underlined.

(10) kulayphu mohyeng-i [phulokulaym-uy hapseng
 kwa sayang-uy kemcung]-ey iyongtoyul swu iss-ta.

*(17) [cencakyeysanki hoylo-uy selkyey-na

design of computer circuit
 kocang cintan]-uy kyengwu-ey

fault diagnosis
 kwangpemwi-hakey iyongtoyko iss-ta.

(This model is widely used for the design of computer circuits and for fault diagnosis.)

(Heuristic 8) In Noun₁+Noun₂+...+Noun_n+Conj+X, Noun₁+Noun₂+...+Noun_n is considered a conjunct. This heuristic may be ambiguous when there are more than two coordinations in a sentence. (Heuristic 7) is applied to (18) also.

(18) pon nonmwun-uy Local Area Network-nun
 [2kay-uy host processor, 1kay-uy IMP, 3kay-uy
 disk unit-wa yelekay-uy data link]-lo kwuseng-
 toyey iss-ta.

(The system configuration of this LAN includes two host processors, one IMP, three disk units, and several data links.)

*(19) LNA-uy [hatuweye *mich*

hardware
 sophuthuweye] phulothokhol [selkyey *mich*

software design
 kwuhyen] pangpep-ul kiswultha-yessta.

implementation

(The design and implementation of HW and SW protocol for Local Network Adapter is described.)

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

(20) phase-nun DCG-lul iyong-haye
 caksenghyessumye [pasing,

parser
 hankul naypwu kwuco sayngseng *kuliko*

Korean internal structure generation
 mwunpep pyenliwan]-uy kinung-ul ha-nta.
grammar conversion

(A parser, implemented using DCG, parses the list form, generates Hangul internal structure and converts the grammar from English into Korean.)

*(21) pon nonmwun-un [umseng *mich* hwasang]

audio video
 teitha censing sepisu-lul silsikan-ulo

data transmission service
 ceykongha-nun phulothokhol-ey kwanha-n
 kes-ita.

(The research reported in this paper is on the development of a communication protocol to support Audio/Video data transmission in real time.)

(Heuristic 10) In case of X₁+Conj+X₂+ul/lul, the second conjunct is X₂. The same holds for the case marker 'kai' instead of 'ul/lul'.

(22) phothulan-uy [khemphyuthe-uy hyoyulseng-kwa

efficiency of computer
 sayongca-uy phyenyuseng]-ul

user friendliness
 nophi-koca sacen penyekki-lul selkyeyha-
 yessta.

(To improve the efficiency and the user friendliness of Fortran in the computer, the pre-compiler was designed.)

*(15) hankul teitha-uy [cenguy, cocak, ciluy *mich*

pokose cakseng]-tung-uy kinung-ul
 ceykongha-nta.

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 (Ishikawa *et al.*, 1985) to correctly determine the analysis of scope in coordinate phrases in English without deep semantic analysis.

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

4. Syntactic Analysis of Coordinate Structure

4.1 Analysing Coordinate Noun Phrase

The syntactic analysis phase determines coordination elements related to conjunction, 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.

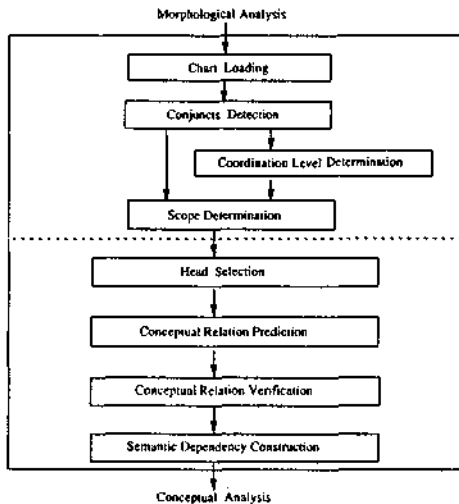


Fig. 5 Overall stages of syntactic analysis

4.1.1 Loading Chart Information. A chart is the output of morphological analysis which is loaded to analyse 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 analysed. 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 Conjuncts. The coordinating conjunction is marked first at this stage. Since Korean has a prenominal modifying structure, detection of candidates proceeds from the coordinating conjunction to the right.

In 'A-wa B-lul C-ey D-hannu E-hul', 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 postpositional 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 relations. 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 coordinations 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

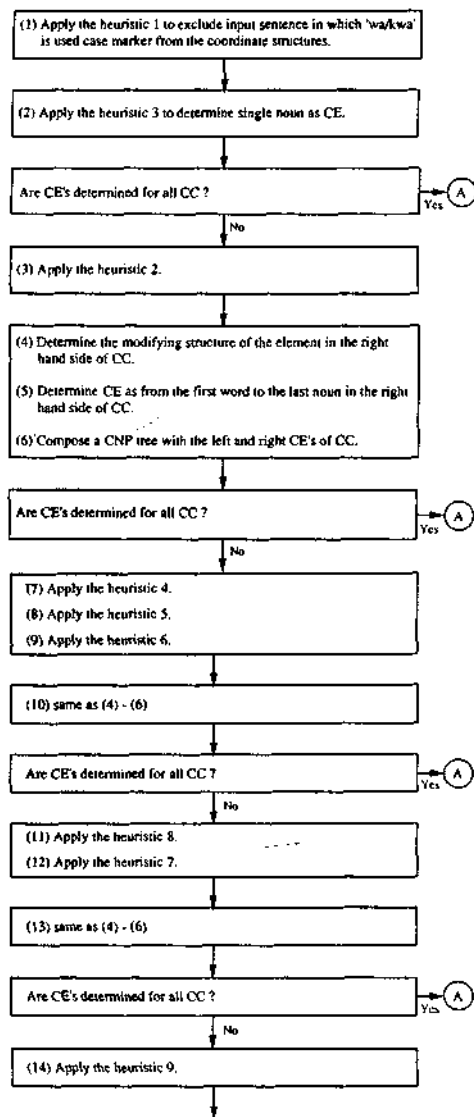
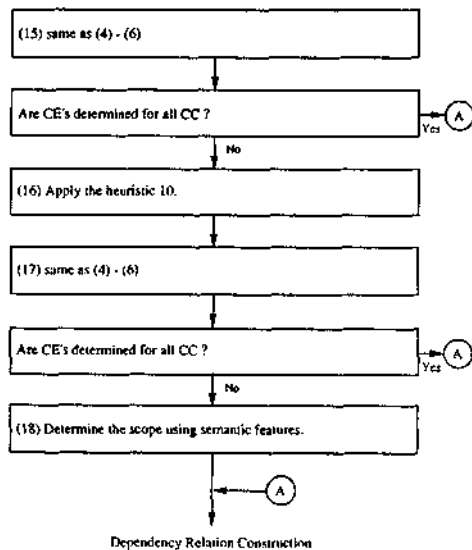


Fig. 6 Sequence of applying heuristics

natural and preferable reading as output, without any extensive semantic processing.

Finally, a computational model for coordinate noun 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



Dependency Relation Construction

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

- Comrie, B. (1981). *Language Universals and Linguistic Typology*. Basil Blackwell, Oxford, England.
- Ishikawa, G., Tsujii, J., and Nagao, M. (1985). Ambiguities in English Coordinate Phrase and its Processing. *Natural Language Processing*, 50.2: 1-7. Information Processing Society of Japan.
- Jappinen, H., Lehtola, A., and Valkonen, K. (1986). Functional Structures for Parsing Dependency Constraints, *COLING-86*, 461-3.
- Kim, Cheol Ho, Choo, Byung Hong, Kim, Gil Chang, Choi, Key-Sun, and Ichiyama, S. (1988). Generation of Korean from Conceptual Representation. '88 Fall Conference of Information Processing Society of Japan, pp. 947-8.
- Kim, Jin Soo (1987). *A Study on Korean Connective Post-position Particle and Word Ending*. Tower Press, Seoul, Korea.
- Kim, Wan Jin (1970). Clausal Conjunction 'wa' and Phrasal Conjunction 'wa'. *Linguistic Research*, 6.2: 1-10. Language Research Institute, Seoul National University, Korea.
- Koduma, T. (1987). *A Study on Dependency Grammar*. Kenkyusha, Tokyo, Japan.
- Lee, Jong Ilyeok and Kim, Gil Chang (1988). Voice Generation from Conceptual Representation: Syntactic, Semantic, and Pragmatic Aspects, *Literary and Linguistic Computing*, 3.4: 250-4.
- Muraki, K. (1986). VENUS: Two-phase Machine Translation System, *Future Generations Computer Systems*, 2: 117-19.
- Ichiyama, S., and Fukumochi, Y. (1985). Augmented Dependency Grammar: A Simple Interface between the Grammar and the Knowledge. 2nd European Conference of ACL, pp. 198-204.

- Nagao, M., Tsujii, J., and Nakamura, J. (1983). Coordinate Phrase in Technical Paper and its Processing, *Natural Language Processing*, 36.4: 1-9. Information Processing Society of Japan.
- Sag, I. A., Gazdar, G., Warsaw, T., and Weisler, S. (1984). *Coordination and How to Distinguish Categories*. Report No. CSLI-84-3, Stanford University.
- Sgall, P. and Panevova, J. (1987). Machine Translation, Linguistics, and Interlingua. 3rd European Conference of ACL, pp. 99-103.
- Shudo, K., Yoshimura, K., and Tsuda, K. (1986). Coordinate Structures in Japanese Technical Sentences, *Transactions of Information Processing Society of Japan*, 27.2: 183-90.
- Tsujii, J., Nakamura, J., and Nagao, M. (1984). Analysis Grammar of Japanese in the Mu-Project—A Procedural Approach to Analysis Grammar, *COLING-84*, 267-74.