Analyzing Korean Dialogical Sentences with Situational Knowledge

Hea-Kyung Jeong† and Yong-Seok Lee††,

Dept. of Computer Science , Chonbuk National University, Korea

Summary

Korean interface has been widely researched to provide a convenient and smart user interface in the Internet shopping mall area. Existing Korean interface systems have used only a morphological and/or a syntactic knowledge without considering semantics. Therefore, the system users must make a query with a complete Korean sentence without an ellipsis and an anaphora. For more natural interface, the system must be able to handle dialogical sentences using a situational knowledge. In this paper, we suggest Korean Interface that uses domain specific semantic information based on Sowa's Conceptual Structure: conceptual patterns, relational definitions, semantic structure conversion rules, and actors etc. First, we explain the roles of the information. And, we will show a methodology that translates several kinds of Korean sentences having same meaning into one semantic structure using the information. Also, we will show a method that transforms an incomplete Korean sentence into a complete semantic structure. The advantage of the suggested system can provide user with Korean interface kindly by processing a dialogical query sentence using a situational knowledge.

Key words:

Korean interface, dialogical sentences, a situational knowledge, domain specific semantic information, semantic structure.

1. Introduction

Since the need of natural interface is on the rise, a lot of researches have been going on actively[4],[5],[6],[7],[8] in the Internet shopping mall. In the Internet application area, there are some domestic goods searching engines using Natural Language processing technology: NLIDB, Pandora, Answerer, AIAGENT[2]. It is assumed that some commercial systems[3] using above engines are handled by extracting nouns of a morphological analysis level. As an example, the query "100 만원에서 150 만원 사이의 컴퓨터(1 to 1.5 million won computers)?" shows a right answer, while the similar type of the query "300 만에서 400 만 화소대의 디카(3 to 4 mega pixel Digital cameras)?" doesn't show a suitable one. It is assumed that it is the result from an exception processing; processing is affected by the monetary unit like "원(won)" rather than the meaning of a sentence.

Thus, end users have to input a complete query sentence without an elliptical phrase and an anaphora phrase because existing Korean interface systems use primarily a morphological information and/or a syntactic knowledge with no semantics.

Nevertheless, there is a strong probability for end users to use a situational knowledge when they enter a query for searching goods, but there is no method to solve this problem by existing method. For example, in Fig. 1 the query "CD 겸용인 것(Things that are a CD combined using)?" and "국내 제품으로(Domestic products)?" can not be handled after the query "인기있는 어학용 카세트 플레이어(Most popular cassette players for a language study)?".

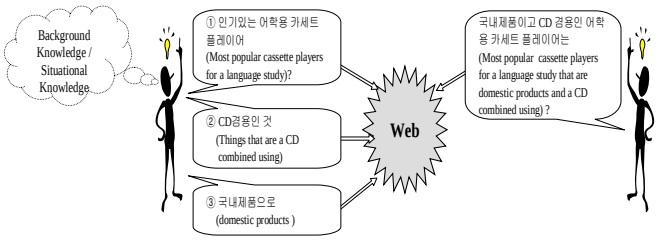

Fig. 1 Dialogical Query Sentence Complete Query Sentence

In this paper, our Korean interface targets on an environment that offers the information of goods through a dialogical query sentence rather than a complete query sentence.

Although an ellipsis and an anaphora are very frequent in the dialogical query sentence, a person does not feel trouble on a conversation, because of using a situational knowledge. For using a situational knowledge like a person, Korean interface could use a knowledge base and transform a user's query into a semantic structure. That is, Korean interface must be able to express a situational knowledge inside a computer, and to represent and to operate a meaning to solve an anaphora problem or an ellipsis problem. Therefore, Korean interface must be able to translate the result of an analysis to an exact semantic structure, and it have to use a knowledge base to solve an ambiguity, an anaphora and an ellipsis. In this paper, we call this Korean interface based on knowledge.

We propose a design of Korean interface based on knowledge using a domain specific semantic information that is based on Sowa's a conceptual structure[1]. There are several semantic knowledge such as conceptual

Manuscript received July 5, 2007
Manuscript revised July 25, 2007

patterns, relation definitions, semantic structure conversion rules, and actors, etc. First of all, we describe a role of the above information. And using the information, we will show the method that various types of Korean sentences can be changed to one semantic structure.

This paper is organized as follows. In section 2, we describe dialogical Korean query sentences and several kinds of Korean query sentences in electronic goods. In section 3, we explain the kinds and roles of knowledge bases, and in section 4 we describe a methodology to generate a semantic structure using knowledge bases and a situational knowledge. In section 5, we present the architecture of a semantic structure generator. Finally, conclusions are given in section 6.

2. Korean Query Sentences

In this chapter, we will look around several cases of dialogical Korean query sentences, and classify the types of Korean query sentences according to a processing unit.

2.1 Dialogical Query Sentences

A dialogical query sentence is a context-sensitive sentence when a user searches goods on the Internet shopping. A user does not choose goods through a query of just once, and wants to give another query continuously in connection with the result from a preceding query. Therefore, a characteristic of a dialogical query sentence is that the using of an ellipsis and an anaphora phrase is frequent.

Table 1 shows dialogical query sentences that can be compared with a previous query. There are four types according to query content.

Table 1: Examples of dialogical query sentences

Query No	Example
Query 1	"인기 있는 어학용 카세트 플레이어? (Most popular cassette players for a language study?)" "CD 겸용인 것 (Things that are a CD combined using)" "국내제품으로 (a domestic products)"
Query 2	"LG 에서 만든 세탁기? (Washing machines that made by LG?) " "트롬으로 (TROM) " "대용량인 것 (Things that are high-capacity) "
Query 3	"50 만원 대 TV 를 보여줘? (Show me five hundred thousand won price range TV?) " "그 보다 싼 것은? (Thing that is cheaper than it?) " "삼성 것 (Things that are SAMSUNG) "
Query 4	"요즘 인기 있는 김치 냉장고? (These day, most popular Kimchi refrigerators?) " "딤채 빼고 (except for DIMCHAE) "

First, a user makes a query to get other functions of different goods besides desired goods. For example, Query 1 shows a change of a focus from a cassette player to a CD player. Second, it is a case that appends additional functions to the goods that a user wants. For example, Query 2 is belonged to here. Third, it is a case that has opposite the goods that a user wants, Query 3 belongs to here. Fourth, it is a case that selects goods except for special conditions among lots of goods that a user wants. Query 4 belongs to this case**.**

2.2 Many different kinds of Korean query sentences

A user makes a query with several kinds of syntactic structures in order to search same goods, because a syntactic structure of Korean sentence is different according to a predicate. Fig. 2 is several query sentences that ask "LG 세탁기(LG washing machines)?" in Query 2.

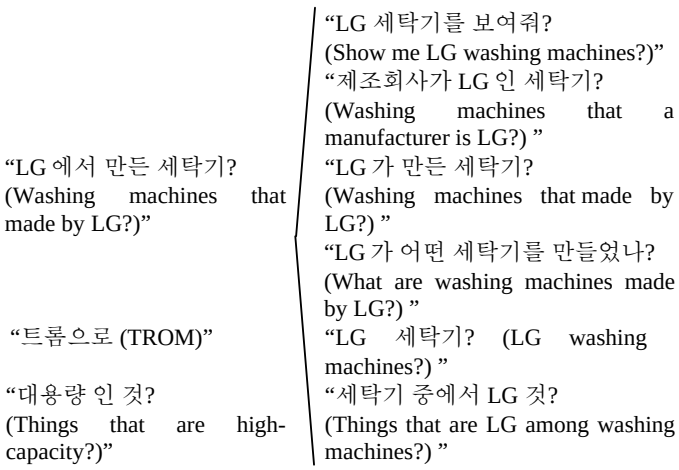

Fig. 2. Several query sentences that ask "LG 세탁기(LG washing machines)"

Table 2: Basic unit to analyze a Korean query sentence(from here, an italic style means a function word)

Basic Unit	Query Type		Remark
Predicate Phrase	N *을/를* V (V N)	① "LG 세탁기*를* 보여줘? (Show me LG washing machines?)"	
	N *이/가* V 한 M	② "제조회사*가* LG 인 세탁기? (Washing machines that a manufacturer is LG?) "	

	(M that N V)	③ "LG *가* 만든 세탁기? (Washing machines that made by LG?) "	Same Meaning
	including interrogative style(WH-type)	④ "LG *가* 어떤 세탁기*를* 만들었나? (What are washing machines made by LG?) "	
Modifier Phrase	N(*의*) M (M N)	⑤ "LG 세탁기? (LG washing machines?) "	
Adverb Phrase	N *중에서* M (M among N)	⑥ "세탁기 *중에서* LG 것?(Things that are LG among washing machines?) "	
Comparative Phrase	N *이/가* C *보다* A 한 M (M that A more than C)	⑦ "가격*이* 100 만원대 *보다* 싼 냉장고? (Refrigerators that are cheaper than one million won price range?)"	Different Meaning
		⑧ "가격*이* 딤채 냉장고 *보다* 싼 냉장고? (Refrigerators that are cheaper than DIMCHAE?)"	

Table 2 shows that various syntactic structures are classified according to a basic unit: predicate phrase, modify phrase, adverb phrase, and comparative phrase. First, a predicative phrase is a case that includes a predicate. It is to be classified into four items: a common verb type like "보여주다(show)", a common verb type like "만들다(make)", be verb type like"이다(be)", a predicate including an interrogative style like WH-clause. Second, a modifier phrase is a case that an uninflected word modifies other uninflected word. Third, an adverb phrase is a case that includes an adverb such as "중에서(among)". Fourth, a comparative phrase is a case that has a comparative expression such as "보다(more than/~er than)".

We can know that these sentences are expressed variously using a predicate phrase, a modifier phrase, and an adverb phrase to search an information on "Washing machines", but a meaning of these sentences are all same. In contrast to, two sentences including a comparative phrase have a same syntactic structure but a different meaning.

3. Knowledge bases

In this chapter, we describe various knowledge bases needed in electronic goods in order to generate a semantic structure. It is almost impossible to construct a common purpose knowledge base because a word has various meanings according to a situation and a classification of a meaning is vague according to an application area. But, if a domain is restricted within a specific application area such as the Internet shopping mall, we can build up knowledge base completely.

Table 3 shows the expression to one semantic structure from several kinds of query sentences that have same meaning (Fig. 2).

Table 3: Representation several syntactic structures having same meaning with one semantic

No	Syntactic Structure	Pseudo semantic structure	Semantic Structure
①	보여줘 / 세탁기를 / LG	Show / OBJ / Washing Machine / OF / Company:LG	
②	세탁기 / LG의 / 제조회사가	Be / OBJ / Company:LG / SBJ1 / Company / SBJ2 / Washing Machine?	
③	세탁기 / 만든 / LG가	Make / SBJ / Company:LG / OBJ / Washing Machine?	
④	만들었나 / 세탁기를 / 어떤 / LG가	Make / SBJ / Company:LG / OBJ / Washing Machine?	
⑤	세탁기는 / LG	Washing Machine? / OF / Company:LG	
⑥	S / 것 / LG / 세탁기 중에서	Washing Machine / OF(OWN) / Thing:? / OF / Company:LG	

In table 3, the pseudo semantic structures are to be generated as a combination of the basic unit which proposed in table 2, by receiving the syntactic structure of a query sentence as input. Such generated pseudo semantic structure includes several linguistic ambiguities that are unsolved by expressing a surface meaning of a sentence. Also, in a step that generates a pseudo semantic structure, it can not know a fact that the meaning of these sentences is all same.

We describe sorts and roles of knowledge bases that are necessary to change from a pseudo semantic structure to exactly one semantic.

Table 4:Korean to English Mapping Table for Table 3 and Table 5

Classifica_tion	Korean Word	English Word	Concept/Relation Type
Content Word	보여주	Show	[Show]
	세탁기	Washing Machine	[Washing Machine]
	LG	Company Brand	[Company:LG]
	이	Verb "Be"	[Be]
	제조회사	Manufacturer	[Manufacturer]
	만들	Make	[Make]
	어떤	What/Which	[Thing:?]
	것	Thing	[Thing]

	냉장고	Refrigerator	[Refrigerator]
	싸	Cheap	[Cheap]
	100 만원대	One Million Won Range	[Price:One Million Won Price Range]
	김치냉장고	Kimchi Refrigerator	[Kimchi Refrigerator]
	딤채	DIMCHE(Brand)	[Kimchi Refrigerator: DIMCHE]
Function Word	*을/를*	The Objective Case	(OBJ)
	이/가	The Subjective Case	(SBJ)
	의	The Possessive Case	(OF)
	중에서	An Adverb Phrase	(AMONG)
	보다	A Comparative Phrase	(THAN)

Inflected Word	Content Word	Function Word & Comment	
보여줘	보여주	*어*	The ending of word of an Imperative Form
인	이	ㄴ	The ending of word of a Modifier Form
싼	싸	ㄴ	
만든	만들	ㄴ	
만들었나	만들	*었나*	The ending of word of an Interrogative Form

Table 4 shows that Korean words are mapped to English words. For example, an inflected word "보여줘" consists of a content word "보여주" and a function word "*어*". A content word "보여주" corresponds to "show" in English and means to a concept type [show]. A function word "*어*" is the ending of word of an imperative form.

3.1 Conceptual patterns

A conceptual pattern restricts a relationship between a predicate and an object, a relationship between an object and an object in the real world.

Table 5: Conceptual patterns

No	Syntactic Structure	Conceptual Pattern
①	보여줘 / 세탁기 / LG	Show — ORI — Electronic Goods; Electronic Goods — OF — Company
⑦	냉장고 / 싼 / 100만원대 보여	1 Cheap — SBJ2 — Electronic Goods; SBJ1 — Price; THAN — Price:Number
⑧	냉장고 / 싼 / 딤채냉장고 보여	2 Cheap — SBJ2 — Electronic Goods; SBJ1 — Price; THAN — Electronic Goods

Table 5 shows conceptual patterns that are to define in electronic goods. For example, the syntactic structure of a sentence ① consists of a predicate and a modifier phrase. Among these, let's examine a modifier phrase. "세탁기 (Washing Machine)" is a subtype of "Electronic Goods" and LG is a company. As we represent a modifier relation into a relation "OF", the conceptual pattern of a sentence ① means that a concept [Electronic goods] and a concept [company] are related to conceptually".

The conceptual patterns need to look for matching candidates possible and to translate to a semantic structure while searching the pseudo semantic structure based on a tree structure. Also, these need to incorporate an ellipsis phrase and an anaphora in a dialogical sentence.

3.2 Relation Definitions

A relation definition shows a relationship between an object and an object in the real world. Table 6 is relation definitions that can be defined in electronic goods domain.

Table 6: Relation Definition.

No	Relation Definition
Rule 1	**Relation** OBJ(show, electronic goods) **is** [electronic goods:?]
Rule 2	**Relation** OF(x,y) **is** [make] - (OBJ) - [electronic goods:x] (SBJ) - [company:y]

Rule 1 defines that the object of a predicate is omitted. A predicate "보여줘(show)" in the sentence ① can be replaced by other predicates "찾아줘(find), 알려줘(inform), 원해요(want), 사고싶어요(want to buy), 검색해줘(search), 골라줘(choose), 필요해(need)". The common feature of these predicates is that the object of these predicates is restricted to electronic goods. . By contracting or expanding Rule 1 and Rule 2, the pseudo semantic structure of the sentence ① could be converted to a semantic structure or vice versa.

3.3 Type hierarchy

Concept types are organized in a hierarchy according to levels of generality. The type hierarchy needs to grasp a inherited relationship of concept types: a concept [washing machine] is the subtype of a concept [electronic goods].

3.4 Semantic structure conversion rule

This rule changes a semantic structure by reducing double subjects into single subject. There are lots of predicates that have double subjects in Korean language. The rule needs to make the predicate that take double subjects

simple as single subject or to change a pseudo semantic structure to a semantic structure.

Table 7: Semantic structure conversion rule

Syntactic Structure	Conceptual Pattern	Remark
(A) 세탁기는 가격이 싸다 (A washing Machine is cheap) 싸다 가격이 세탁기는	Cheap — SBJ2 — Electronic Goods SBJ1 — Price	Same Mean_ ing
(B) 세탁기의 가격이 싸다 (A Price of a Washing Machine is cheap) 싸다 가격이 세탁기의	Cheap — SBJ1 — Price — OF — Electronic Goods	

For example, table 7 shows a semantic structure conversion rule of an adjective "싸다(cheap)" that has double subjects. A sentence (A) has two nominative auxiliary words superficially, and a sentence (B) has one. Although a sentence (A) and (B) have a different syntactic structure, but the meaning of two sentences is same. We define a semantic structure conversion rule to change the conceptual pattern that is created from a parse tree (A) into (B).

3.5 Actor

An actor is an automatic execution. For example, in the sentence ⑧ that include a comparative phrase we can not know how much is the corresponding price to "refrigerator". We must retrieve a refrigerator table in electronic goods database to search the price of this "DIMCHAE refrigerator". The function to search the price is called an actor.

4. Semantic Structure Generation

This chapter describes a process that translates a dialogical query sentence into a semantic structure using knowledge bases described in previous chapter. Various query sentences with a same meaning is also processed in batch.

The procedure of a semantic structure generation is as following.

1. Step that incorporates by solving a anaphora and an elliptical phrase.
2. Step that collects applicable patterns while searching a tree
3. Step that prioritizes patterns
4. While applying patterns one by one, if an actor exists then execute the actor,
   else go to step 2 .
5. If the semantic structure that is generated at step 4 does not agree with a defined
   semantic structure then move to step 2, else generate a SQL.

We describe why necessary each step and explain a process to handle each step by an example.

[Step 1] Incorporating step: Solving an anaphora and processing an ellipsis phrase

An anaphora phrase is used frequently in dialogical query sentences and the query sentences included an adverb phrase such as "*중에서*(among)". A resolution of this anaphora phrase regards nouns of just a previous sentence as situational knowledge and then pushes these nouns to a stack. In here, a size of a stack accommodates maximum 100 sentences.

In a dialogical query sentence, each sentence is connected with each other rather than each sentence has an independent meaning. However, a semantic relation between sentences limits in just previous sentence, and a processing of an ellipsis uses only the sentence.

For example, it describes a method to find an anaphora phrase of the sentence "Thing that is cheaper than it?" after querying the first sentence in the query 3.

First of all, in the first sentence, push a concept [TV] and a concept [price:fifty hundred thousand won] to a noun stack for using as a situational knowledge. That is, there are two concepts by order [TV] and [Price] in stack. And then, we use conceptual patterns and situation knowledge to grasp anaphora phrases of second query sentence.

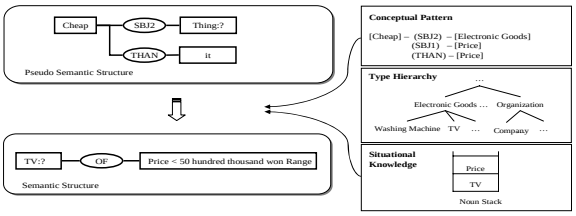

Fig. 3 Step that incorporates an anaphora phrase in Query 3

Fig. 3 shows a process that obtains specific information that "그(it)" indicates "price" and "것(thing)" is "TV" by using a conceptual pattern "싸다(cheap)" and a situational knowledge.

[Step 2] Applicable pattern collection while searching a tree

This step collects candidates that may be matched using a conceptual pattern. For example, the first sentence in Query 3 has two patterns: a predicate phrase and a modifier phrase.

[Step 3] Prioritizing patterns

When matching candidates are existed more than two, a priority order of a pattern is to decide which candidate is applied first.

In the example of step 2, a method 1 applies a predicate phrase after applying a modifier phrase first. A

method 2 applies a modifier phrase after applying a predicate phrase first.

In this paper, a common predicate among predicate phrases is handled first: a query type “N 을/를 V(N+V)”. Because it is omitted by the rule 1 of a relation definition, a predicate of this type can simplify a semantic structure. We apply a method 2: a modifier phrase pattern after omitting the predicate of “보여주다(Show)” in a predicate phrase pattern.

[Step 4] Applying patterns one by one if an actor exists then execute the actor and go to step 2.
When a comparative target is not given by a certain value in a sentence including a comparative phrase, its value is obtained by searching a database.

Fig. 4 is a process that applies an actor <refrigerator> to obtain the price of DIMCHAE refrigerator in a sentence ⑧

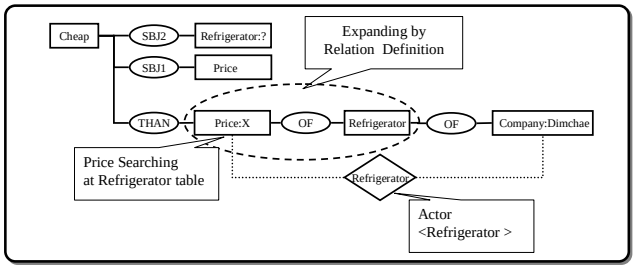

Fig. **4** Execution of Actor <Refrigerator>

[Step 5] If a defined semantic structure does not match a generated semantic structure then move to step 2, else generate a SQL.

If the semantic structure that is generated at step 4 matches a defined semantic structure then go a SQL generation module, else move to step 2.

5. Architecture of a Semantic Structure Generator

This Chapter describes a semantic structure generation module for Korean interface for based on knowledge

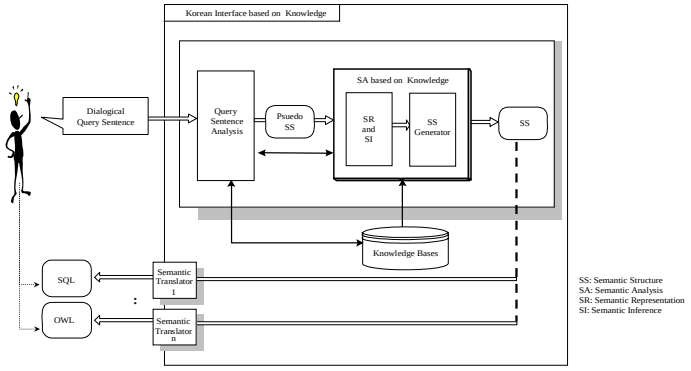

Fig. 5 Architecture of Korean interface based on knowledge

A semantic structure generation for Korean interface based on knowledge has two phases: a query sentence analysis and a semantic analysis based on knowledge. The bold line area in fig. 5 shows a semantic analysis based on knowledge.

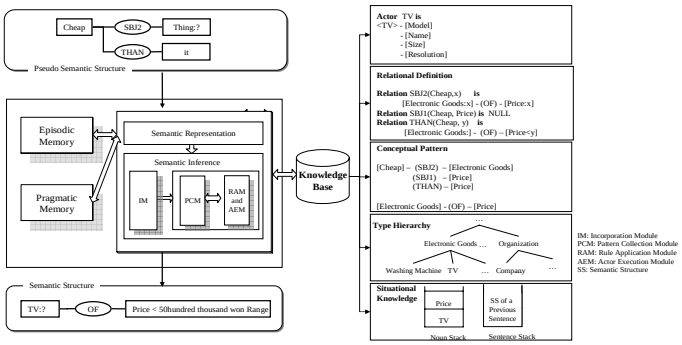

Fig.6**.** Architecture of a Semantic Structure Generator

The semantic analysis translates a pseudo semantic structure into a semantic structure using knowledge such as an actor, a relation definition, a type hierarchy and a conceptual pattern, and through a situational knowledge of the query sentence. For example, fig. 6 shows knowledge and an inference in order to generate the semantic structure of the second query sentence in the Query 3. When a semantic structure is generated, it is needed two kinds of memories. Fist, Pragmatic Memory means memory space that memorized domain knowledge now in use. Second, Episodic Memory means memory space that memorized the new knowledge obtained through a user's input. A semantic inference has to provide the inference operators that may produce the results wanted by processing domain knowledge. There are inference operations such as a specification module, a pattern gathering module, a rule application module and an actor execution module that a user inputs.

6. Conclusions

We described a design of Korean interface based on knowledge to search goods user wants in Internet shopping mall. Because of difficult to build up a general knowledge, we explained on electronic domain.

We proposed a method that transforms an incomplete dialog sentence into a complete semantic structure using a situational knowledge and domain knowledge. We also described a method that translates several syntactic structures into one semantic structure in batch processing.

The future, we will research a method that produces a SQL from a semantic structure generated in this paper.

References

[1] Jon F. Sowa, "Conceptual Graph Summary," In:Timothy E. Nagle, Jancie A. Nagle and Laurie L. Gerholz. (eds.), Conceptual Structures:Current Research and Practice, Ellis Horwood, pp. 3-52, 1992.

[2] http://www.diquest.com, http://www.unisoft.co.kr, http://www.quark.co.kr http://www.answerer.co.kr, http://ontoline.com

[3] http://www.gseshop.co.kr/jsp/main.jsp

[4] 4. I. Androutsopoulos, G. Ritchie, and P. Thanisch, An Efficient and Portable Natural Language Query Interface for Relaional Databases, Proceedings of the 6th International conference on Industrial & Engineering Applicational Conference on industrial & Engineering Applications of Artificial Intelligence and Expert Systems, 1993.

[5] 5. M Bates, M. G. Moser, D. Stallard, "The IRUS transportable natural language database interface," in Expert Database Systems, Benjamin Cummings, 1986.

[6] 6. Jae Moon Kim, Jae Hong Kim, Sang Jo Lee, "Design and Implementation of Korean Natural Language Query System Using Keyword" Proceedings of the 19th Korea Information Science Society Conference, vol. 22, No. 1, pp. 957-960, 1995.

[7] 7. Jinseok Chae, Seonggi Kim, Sukho Lee, "Design and Implementation of a Korean Natural Language DB Query System," Journal of the Korea Information Science Society, Vol. 20, No. 6, pp. 810-820, 1993.

[8] 8. Hyun-Seok Park, "Representing & Analyzing QLFs for Korean ATIS Queries," KRCOGSCI, Vol. 10, No. 5, pp. 17-25, 1999.

[9] 9. In-Cheol Park, "A Study on Generation of Conceptual Graph using Korean Sentence Patterns," PhD Thesis, Chonbuk National University, 1998.

HeaKyung Jeong received the B.E. and M.E. degrees in Computer Science and Statistics from Chonbuk National Univ. in 1990 and 1992, respectively. He is currently pursing his Ph.D. degree in Chonbuk National Univ. her research interests are Korean Language Processing, Natural Language Interface and Ontology.

YongSeok Lee received the B.E. degree in Electronic Engineering from Seoul Univ. in 1977, M.S. degree in Computer Science from KAIST, Ph.D. degree in Intelligent and Information System Engineering from Tokusima University of Japan in 1995 respectively. From 1979 to 1983, he was a senior researcher at Korea Research Institute of Standards. Since 1983, he has been a professor in the Dept. of Computer Science, Chonbuk National University. His research interests are Korean Language Processing, and Information Retrieval.