

# Integrated Knowledge Bases of Semantic Networks for Automatic Translation of Ambiguous Words

Yoo-Jin Moon\* · Young-Ho Hwang\*\*

단어의 자동번역을 위한  
의미 네트워크의 통합 지식베이스

문 유 진\* · 황 영 호\*\*

## Abstract

Automatic language translation has greatly advanced due to the increased user needs and information retrieval in WWW. This paper utilizes the integrated knowledge bases of noun and verb networks for automatic translation of ambiguous words in the Korean sentences, through the selectional restriction relation in the sentences. And this paper presents the method to verify validity of Korean noun semantic networks that are used for the construction of the selectional restriction relation by applying the networks to the syntactic and semantic properties. Integration of Korean Noun Networks into the SENKOV system will provide the accurate and efficient knowledge bases for the semantic analysis of Korean NLP.

---

\* Dept. of MIS, Hankuk Univ. of Foreign Studies Korea, yjmoon@hufs.ac.kr

\*\* Honam University Korea, yhhwang@honam.ac.kr

## 1. Introduction

Automatic language translation has greatly advanced due to increased user needs and information retrieval in WWW. Most of languages have quite a lot of polysemous words. So automatic translation of ambiguous words(ATAW) in Korean has been one of the most popular research themes in NLP and information retrieval. In order to solve the problem, semantic networks for verbs and nouns are combined into an integrated knowledge base. If the semantic networks for verbs and nouns are combined effectively, they will play a great role in solving the ATAW problem.

There are several kinds of semantic networks for nouns and verbs --- WordNet, Levin verb classes and VerbNet., German WordNet, Korean Noun WordNet and SENKOV(Semantic Networks for Korean Verbs) etc.[1, 2, 5, 7]. There have been controversies about their validity for ATAW and difficulties to prove it. It has been a difficult task to prove that semantic networks have been built with valid hierarchical classes and the semantic networks work for the semantic analysis of sentences properly.

It is why the networks are based on dictionaries, concept, recognition and heuristic methods[1, 4, 5, 8, 11, 12]. This paper presents the method to verify

validity of the semantic networks that are used for the construction of the selectional restriction relation by applying the networks to the syntactic and semantic properties. And this paper utilizes the integrated knowledge bases of Korean noun and verb networks for automatic translation of ambiguous words in the Korean sentences, through the selectional restriction relation (subjects, objects and predicates etc.) in the sentences.

## 2. Literature Review

### 2.1 English Semantic Networks

WordNet is a kind of semantic networks for English nouns and verbs, and Levin verb classes are semantic networks for English nouns. WordNet[7] is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adverbs and adjectives are implemented in terms of synonym sets. Each synonym set represents one underlying lexical concept. WordNet presently contains about 120,000 word forms. WordNet may be viewed as the semantic networks which represent hypernyms of English word senses in the form of IS-A hierarchies.

9 Verbs of Putting : arrange, put, place
9.1 Put Verbs : put, set, place, pose, position, lay
Verbs of Putting in a Spatial Configuration
9.3 Funnel Verbs : funnel, move through a funnel, pour through a funnel
:
9.7 Spray/Load Verbs : spread, distribute, extend, cause to spread

[Figure-1] A Part of Levin Verb Classes

Levin verb classes[5, 6] contain various syntactically relevant and semantically coherent English verb classes. It takes a semantic classification structure and incorporates the syntactic relationship into the semantic relationship for verbs. Levin classifies approximately 3,000 verbs into 49 verb classes and the verb class groups meaningfully related verbs together. [Figure-1] illustrates a part of Levin verb classes. However, there is little hierarchical organization compared to the number of classes identified.

English WordNet for nouns does not play a big role by itself for ATAW. As well, a set of Levin verb classes does not play a big role by itself for ATAW. Because a sentence makes sense by joining nouns and verbs etc. Thus this paper tries to integrate semantic noun networks into semantic verb networks.

## 2.2 Korean Semantic Networks

Korean Noun Networks are sets of IS-A hierarchies for Korean nouns[1]. The IS-A hierarchies consist of nodes and edges. The nodes represents synonym sets of Korean nouns and English WordNet. And the edges represents hypernymous relations among nodes. SENKOV system[9] classifies about 700 Korean verbs into 46 verb classes by meaning. [Figure-2] illustrates a part of SENKOV verb classes. It has been implemented on the basis of the definition in a Korean dictionary, with top nodes of Levin verb classes, hierarchies of WordNet and heuristics. It attempts to incorporate syntactic relation into the semantic relation for Korean verbs, and distinguishes the intransitive verb from the transitive verb. [10] proves validity of SENKOV verb classes by applying them to the selectional restrictions among adverbs and verbs in the sentence. In this paper, Korean Noun Networks are utilized to automatically extract sets of hypernymous concepts.

9 Verbs of Putting : arrange, put, place.
9.1 Put Verbs
두다, 놓다, 배열하다, 설치하다, 놓다, 배치하다, 맡기다, 쌓다, 저장하다, 매달다, 걸다, 기대다, 쉬다, 앉히다, 세우다
9.3 Funnel Verbs
집어넣다, 뜨다, 치다, 긁어내다, 접다, 우겨넣다, 밀어넣다, 퍼넣다
:
9.7 Spray/Load Verbs
9.7.1 뿌리다, 튀기다 (intransitive verbs)
9.7.2 뿌리다, 튀기다, 첨병이다 (transitive verbs)
9.8 Fill Verbs
채우다
9.9 Butter Verbs
9.9.1 바르다, 덮다 (agent:1, patient:2)
9.9.2 바르다, 붙이다, 찍우다 (agent:1, patient:3)

[Figure-2] A Part of SENKOV Verb Classes

Most of the verb semantic networks have been classified on the basis of the semantic properties of the verbs. In order to verify valid hierarchical classification of the networks that are used for the selectional restriction relation, [10] considered the syntactic properties as well as the semantic properties of the verbs. That is, the verbs belonging to the same verb class should share the syntactic and semantic properties.

Korean Noun Networks do not play a big role by themselves for ATAW. Also, SENKOV system does not play a big role by themselves for ATAW. Especially when a word has more than one meaning, the word in a sentence can be interpreted by associating the other words in the sentence. Because a sentence makes sense by joining nouns and verbs etc. Thus, this paper attempts to integrate Korean Noun Networks into

SENKOV system.

### 3. Verifying Validity of Noun Semantic Networks

There are many factors to evaluate the validity of the noun semantic networks in NLP. One of the evaluating factors is whether they have valid hierarchical classes for the selectional restriction relation. The valid hierarchical classes of the noun semantic networks will provide proper selectional restriction relations that are required for correct semantic analysis of sentences. In this section the noun semantic networks are considered from the viewpoint of the selectional restriction relation in order to utilize the networks for automatic translation of ambiguous words.

1st sense of water water => food, comestible, edible => substance, matter => object, inanimate object => entity
--

[Figure-3] IS-A Hierarchies of "water" in English WordNet

### 3.1 Validity of English Noun WordNet for Korean

From the viewpoint of the selectional restriction relation, the validity of English Noun WordNet to Korean hierarchical classes is studied.

For example, a predicate "먹다(eat)" in the Korean sentence "메리가(Mary) 사과를(apple) 먹다(eat) - Mary eats an apple" takes the argument of an object, which is a kind of food. English Noun WordNet contains a synset "food, comestible, edible" which can be used as the selectional restriction of the predicate "먹다(eat)". For example, [Figure-3] illustrates one of the IS-A hierarchies for "water" in English Noun WordNet.

A predicate "마시다(drink)" in the Korean sentence "메리가(Mary) 물을(water) 마시다(drink) - Mary drinks water" takes the argument of an object which is a kind of liquid stuff such as beverage and drinking water. However, English Noun WordNet doesn't contain a synset "beverage, drinking water" which

is one of the selectional restrictions of the predicate "마시다(drink)".

Thus English Noun WordNet does not contain all the valid hierarchical classes for Korean ATAW. In addition, it is not possible for the English Noun WordNet to cover the selectional restriction relation in Korean properly.

### 3.2 Validity of Korean Noun WordNet to Korean

From the viewpoint of the selectional restriction relation, the validity of Korean Noun WordNet to Korean hierarchical classes is studied in this section.

A predicate "먹다(eat)" in Korean takes the argument of an object, which is a kind of food. Korean Noun WordNet contains a synset "food, comestible, edible" which can be used as the selectional restriction of the predicate "먹다(eat)" in Korean. For example, [Figure-4] illustrates one of the IS-A hierarchies for "water" in Korean Noun WordNet.

1st sense of 물(water) water (물, 음료수) => beverage, drinking water => food, comestible, edible => substance, matter => object, inanimate object => entity
---

[Figure-4] IS-A Hierarchies of "물(water)" in Korean WordNet

A predicate "마시다(drink)" in Korean takes the argument of an object which is a kind of liquid stuff such as beverage and drinking water. Korean Noun WordNet contains a synset "beverage, drinking water" which is one of the selectional restrictions of the predicate "마시다(drink)".

By similarly applying Korean Noun WordNet to the selectional restriction relation, we know that Korean Noun WordNet contains the valid hierarchical classes for the selectional restriction relation in Korean.

If English Noun WordNet can work for the selectional restriction relations of Korean sentences, we can utilize English Noun WordNet for Korean ATAW without building Korean Noun WordNet. But in section 3.1 we see that we can't utilize English Noun WordNet for Korean ATAW. It means that for the semantic analysis of Korean sentences the proper knowledge base is Korean Noun WordNet, but not English WordNet. Thus this paper utilizes Korean Noun WordNet for Korean ATAW in chapter 4.

#### 4. Integration of Semantic Networks For Translation of Ambiguous Words

Integration of semantic networks will contribute to the semantic analysis of NLP and speech recognition. This paper tries to resolve the automatic translation of ambiguous words in machine translation utilizing integration of noun semantic networks into verb semantic networks for Korean. [Figure-5] illustrates a part of Database for Integration of Semantic Networks (DISNet) [10].

Until now DISNet has been built for about 600 Korean verbs which are a verb list of SENKOV system in the index of [9]. Steps to construct DISNet are as following.

Firstly, Korean dictionaries and LDOCE(Longman Dictionary Of Contemporary English) have been utilized to extract selectional restriction relations of the verbs.

Secondly, corpus has been utilized to

extract selectional restriction relations of the verbs.

Thirdly, verbs are categorized according to the selectional restriction relations of the verbs and SENKOV system.

Fourthly, the selectional restriction relations of the verbs are mapped to Korean Noun Networks.

[Figure-5] describes a part of DISNet. SENKOV verb class 9.1 contains the verb "걸다(hang, stake, run) . The verb "걸다" has three slots and their corresponding values as follows:

POS(part-of speech) : a transitive verb

SYN (synonym set) : S+V+O+L

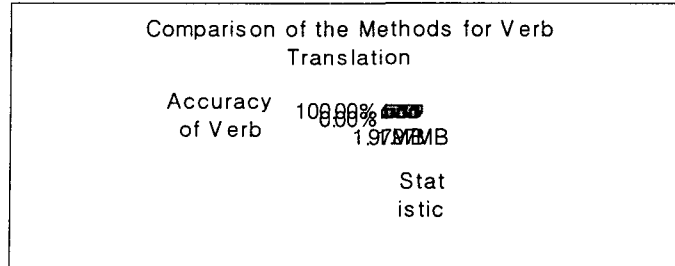
SUBCAT (subcategorization): [S - nc 1.2.1 (person, individual, human) · where the values of the SUBCAT are integrated with the hierarchical classes of Korean Noun Networks.

That is, the selectional restriction of the subject for the verb "걸다" is 'person, individual, human' (noun class 1.2.1) or 'animal, animate being, brute' (noun class 1.2.2), and that of the object is 'life' or 'object, inanimate object, thing' (noun class 1.3) etc. Values of SUBCAT are collected from corpus[2, 3, 12] and mapped to Korean Noun Networks.

<p>9.1 걸다 (hang, stake, run)</p> <p>[POS] : [transitive verb]</p> <p>[SYN] : [S+V+O+L]</p> <p>[SUBCAT] :</p> <p style="padding-left: 20px;">[S - nc 1.2.1 (person, individual, human)</p> <p style="padding-left: 40px;">nc 1.2.2 (animal, animate being, brute)</p> <p style="padding-left: 20px;">V - hang, stake, run</p> <p style="padding-left: 20px;">O - nc 1.3 (object, inanimate object, thing)</p> <p style="padding-left: 40px;">(Eng., hang + nc 1.3 + prep. + L)</p> <p style="padding-left: 40px;">- 목숨(life) (Eng., run + a risk)</p> <p style="padding-left: 40px;">- nc 7.5.3 (money and other possessions, medium of exchange)</p> <p style="padding-left: 40px;">(Eng. stake + nc 7.5.3)</p> <p style="padding-left: 40px;">- nc 2.3.2.8.11 (telephone, telephons) (Eng. call)</p> <p style="padding-left: 20px;">L - nc 5.6 (location) ]</p> <p>9.1 : SENKOV verb class 9</p> <p>POS : part of speech</p> <p>SYN : syntactic structure</p> <p>SUBCAT : subcategorization information</p> <p>nc : hierarchical class of Korean Noun Networks</p> <p>S : subject, V : verb, O : object, L : location</p> <p>Eng. : English</p> <p>*) Values of SYN and SUBCAT are collected from corpus.</p>
--

[Figure-5] A Part of DISNet

[Table-1] Comparison of the Methods for Verb Translation



For example, the verb "걸었다"(the past form of "걸다"(hang, stake, run)) in the Korean sentence "사장이 벽에 그림을 걸었다." might be translated into the English word hung rather than either staked or ran.

사장이(S: owner) 벽에(L: on the wall) 그림을(O: picture) 걸었다(V: ?).

The predicate of the sentence can be translated into one of three English verbs --- hang, stake, run. The object of the sentence is "picture" which belongs to the noun class 1.3. According to DISNet in [Figure-5], the predicate of the sentence may be translated into the English verb "hang." Thus the above sentence is translated into "The owner *hung* a picture on the wall."

However, the verb "걸었다" in the following Korean sentence might be translated into "stake".

할머니는(S: grandmother) 짝수에(on the evens) 오만원(O: 50,000 won) 걸었다(V: ?).

The predicate of the sentence can be translated into one of three English verbs --- hang, stake, run. The object of the

sentence is "50,000 won<sup>21)</sup>" which belongs to the noun class 7.5.3. According to DISNet in [Figure-5], the predicate of the sentence may be translated into the English verb "stake." The above sentence is translated into "A grandmother staked 50,000 won on the evens."

We call the ATAW method utilizing DISNet presented in this paper as the DISNet Method. The DISNet Method for ATAW has been applied to the KEMT(Korean-English Machine Translation) System for verb translation of the middle school textbooks. While the statistics-based method performs the verb translation with about 70.8% of accuracy, the DISNet Method performs the verb translation with about 88.2% of accuracy in the verb translation as illustrated in [Table-1].

DISNet provides the accurate and efficient knowledge base for automatic translation of ambiguous words. Also, DISNet can play an important role in both computational linguistic applications and psycholinguistic models of language processing.

Applicable areas of DISNet are

21) unit of Korean currency



disambiguation of nouns and verbs for NLP and machine translation, writing aid, speech recognition, conversation understanding, the abridged sentences, human-computer interface, and extraction of co-occurrence information and structure information in the information retrieval and summary.

## 5. Conclusions

It has been a difficult task to prove that semantic networks have been built with valid hierarchical classes and the networks for the semantic analysis of sentences properly. It is why the networks are based on dictionaries, concept, recognition and heuristic methods. This paper presented the method to verify validity of Korean noun semantic networks by applying the networks to the syntactic and semantic properties. And this paper utilizes the integrated knowledge bases of noun and verb networks for automatic translation of ambiguous words in the Korean sentences, through the selectional restriction relation in the sentences.

The presented method can be utilized to prove validity of the semantic networks for the other languages. And integration of Korean Noun Networks into the SENKOV system will provide the accurate and efficient knowledge bases for the semantic analysis of NLP.

Future works are to extend the DISNet to all of the Korean verbs and to apply them to NLP.

## References

- [1] 문유진, 의미론적 어휘개념에 기반한 한국어 명사 WordNet의 설계와 구축, 서울대학교 박사 학위논문, 1996.
- [2] 신중호 외, "클러스터링 기법을 이용한 동사 분류," 인지과학회 학술 발표논문지, 1999.
- [3] Gonzalo, J., Chugur, I. and Verdejo, F., "Sense Clusters for Information Retrieval: Evidence from SemCor and the EuroWordNet InterLingual Index," Proc. of SIGLEX Workshop on Word Senses and Multi-linguality in ACL-2000, 2000.
- [4] Hernert, P., "KASSYS : A Definition Acquisition System in Natural Language," Proc. of COLING-94, Aug. 1994, pp. 263-267.
- [5] Levin, B., English Verb Classes and Alterations : A Preliminary Investigation, The MIT Press, 1997.
- [6] Levin, B. and Hovav, M., Unaccusativity : At the Syntax-Lexical Semantics Interface, The MIT Press, 1996.
- [7] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K., "Introduction to WordNet : An On-line Lexical Database," in Five Papers on WordNet, CSL Report, Cognitive Science Laboratory, Princeton

University, 1993.

- [8] Montemagni, S. and Vanderwende, L., "Structural Patterns vs. String Patterns for Extracting Semantic Information from Dictionaries," Proc. of COLING-92, Aug. 1992, pp.546-552.
- [9] Moon, Y., "Implementation of SENKOV System : A Knowledge Base for Semantic Analysis," Information Systems Review, 2000, pp.245-253.
- [10] Moon, Y., "Construction of Semantic Networks for the Language Information Processing," Proc. of International Symposium on Advanced Intelligent Systems, 2001, pp.42-46.
- [11] Nyberg, Eric H. etc., The KANT Translation System: From R&D to Large-Scale Deployment, LISA Newsletter, vol.2:1, 1998.
- [12] Roland, D., "Verb Subcategorization Frequency Differences between Business-News and Balanced Corpora: The Role of Verb Sense," Proc. of the Workshop and Comparing Corpora in ACL-2000, 2000.

## ■ 저자소개



### 문유진

한국외국어대학교를 졸업하고 Pennsylvania State University의 전산학과에서 이학석사학위를 취득하였

으며, 서울대학교 컴퓨터공학과에서 공학박사학위를 취득하였다. 호남대학교 컴퓨터공학과와 한남대학교 경영정보학과 교수로 재직하였으며, 현재 한국외국어대학교의 경영정보학과 교수로 재직 중이다. 그리고 University of Pennsylvania에서 Post-Doc. 연수를 하였다. 관심분야는 인공지능, 자연 언어처리, 지식베이스 및 휴먼인터페이스, 전자상거래 등이다.



### 황명호

서울대학교에서 박사학위를 취득하였고, 미국 펜실베이니아대학 워튼스쿨 (Wharton School)에서 선

임연구위원(Senior Fellow)을 역임하였으며, 현재 호남대학교 사회과학부 교수로 재직중이다.