Building a Large-Scale Commonsense Knowledge Base by Converting an Existing One in a Different Language

Yuchul Jung[1], Joo-Young Lee[2], Youngho Kim[1], Jaehyun Park[2],
Sung-Hyon Myaeng[1,*], and Hae-Chang Rim[2]

[1] School of Engineering, Information and Communications University,
119, Munjiro, Yuseong-gu, Daejeon, 305-732, Korea
{enthusia77, yhkim, myaeng}@icu.ac.kr
[2] Department of Computer Science and Engineering, Korea University 1,
5-ka, Anam-dong, Seongbuk-Gu, Seoul 136-701, Korea
{jylee, jhpark, rim}@nlp.korea.ac.kr

Abstract. This paper describes our effort to build a large-scale commonsense knowledge base in Korean by converting a pre-existing one in English, called ConceptNet. The English commonsense knowledge base is essentially a huge net consisting of concepts and relations. Triplets in the form of Concept-Relation-Concept in the net were extracted from English sentences collected from volunteers through a Web site, who were interested in entering commonsense knowledge. Our effort is an attempt to obtain its Korean version by utilizing a variety of language resources and tools. We not only employed a morphological analyzer and existing commercial machine translation software but also developed our own special-purpose translation and out-of-vocabulary handling methods. In order to handle ambiguity, we also devised a noisy concept filtering and concept generalization methods. Out of the 2.4 million assertions, i.e. triplets of concept-relation-concept, in the English ConceptNet, we generated about 200,000 Korean assertions so far. Based on our manual judgments of a 5% sample, the accuracy was 84.4%.

1 Introduction

This paper describes a hybrid English-Korean Machine Translation (E-K MT) method for making a Korean ConceptNet (K-ConceptNet) based on English ConceptNet [1]. ConceptNet is an easily usable, freely available commonsense knowledge base and natural-language-processing toolkit which supports many practical textual-reasoning tasks over real-world documents including topic-gisting, affect-sensing, analogy-making, and other context-oriented inferences. The knowledge base is a semantic network presently consisting of over 1.6 million assertions of commonsense knowledge covering the spatial, physical, social, temporal, and psychological aspects of everyday life. The Open Mind Common Sense (OMCS) Project [2] started common

[*] Corresponding Author.

A. Gelbukh (Ed.): CICLing 2007, LNCS 4394, pp. 23 – 34, 2007.

sense knowledge gathering with the help of the general public from the year 2000. As of today, the knowledge base consists of over 729,000+ sentences that were inputted from a template-based web interface; it uses strict templates to make it easier to parse the sentences into the forms used in ConceptNet. As part of the OMCS project, ConceptNet [1] was developed based on the OMCS knowledge.

By applying a set of automatic processes (such as extraction, normalization, and relaxation) to the semi-structured English sentences of the OMCS corpus, ConceptNet corpus was generated. ConceptNet's semantic network can be visualized like Fig. 1. For example, concepts can be represented in semi-structured English by composing a verb (e.g. 'drink') with a noun phrase ('coffee') or a prepositional phrase ('in morning')

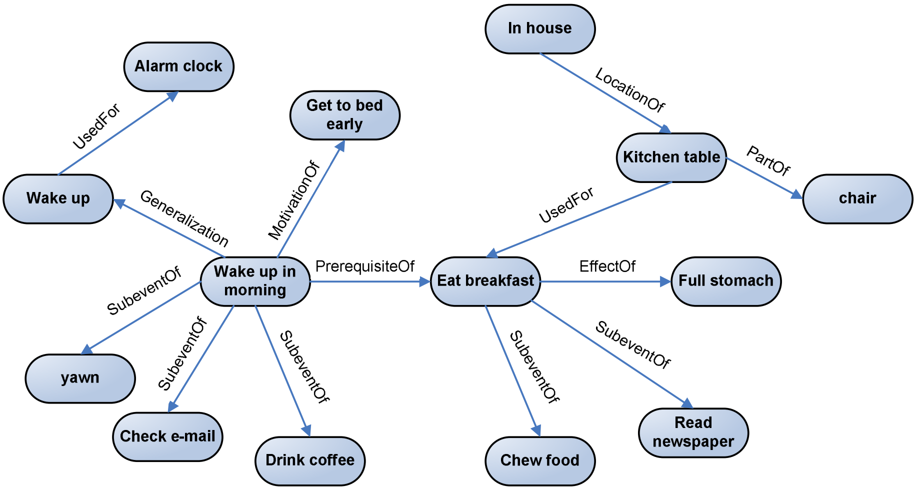


Fig. 1. ConceptNet's semantic network of commonsense knowledge, excerpt from [1]

Among the many avenues we should explore with ConceptNet is an investigation on its usefulness across cultural boundaries. First of all, it is not so clear whether the granularity of the concept nodes and the types of the 20+ relationships in ConceptNet are appropriate for commonsense computing in a country using a different language. On the flip of the coin is that ConceptNet is not immediately usable for most practical applications in Korea because they involve texts in the Korean language. Although a variety of interesting ideas have been proposed for using ConceptNet, it is not clear whether they are applicable to problems in the Korean context. If the original ConceptNet is "translated" into Korean, promising commonsense applications can appear in the Korean language domain. Besides, the existence of the knowledge base in two different languages would be in and of itself useful for applications across the two cultural boundaries.

As an effort to understand the effects of different culture and language within the common sense semantic network of OpenMinds, the GlobalMind[1] project, a multilingual OMCS, has been launched. Currently, a web site is available with an OMCS style knowledge input interface and a visual browser for word inference involving English, Korean, Japanese, and Chinese languages. Our K-ConceptNet construction effort and the GlobalMind project are complementary to each other.

Given the needs, the first task to be embarked on is “Translating” ConceptNet in English to Korean. However, the size of over 1.6 million assertions in ConceptNet makes the task of translation a formidable one, if it is done manually. Although our approach, a combined method which uses a commercial E/K translation S/W (EasyMan E/K translator[2]) and our rule-based translation module for translation, does not produce perfect translation results, it is imperative to employ the method that will at least help reducing the cost of translation. Actually, the commercial E/K translation software produces a large number of mistranslations – awkward or incorrect translations – because it does not take advantage of the OpenMind’s strict template nor they generate Korean sentences with the structure of the template. To fill up the chasm, we have developed a rule-based translation module that can handle English ConceptNet corpus driven from OMCS sentences. Our manual evaluation of 5% sample among 200,000 E-K translated results shows a reasonably high accuracy of about 84.4%.

2 The Method

Ideally speaking, Korean ConceptNet should be built from a Korean OpenMind corpus. That is, collecting Korean commonsense knowledge from Korean people is probably the most natural way. Before launching an OMCS style web-site to build a Korean OpenMind corpus, we wanted to investigate the potential of a method for automatically building Korean ConceptNet using already existing English ConceptNet. The result can be combined with common sense knowledge directly obtained by running a Web site.

Researchers attempted to construct a Korean WordNet using exsiting WordNet [3] and Korean MRD [4]. In addition, Moon [5] used hypernym information of a Korean dictionary in combination with Korean translation of the English WordNet. A manual pruning was done during the noun construction for accuracy. However, this approach is very complex and time-consuming because it requires lots of manual pruning processes that rely on linguists’ vocabulary. Another research for constructing a Korean WordNet based on the English WordNet [6] used a bilingual dictionary to link the senses of Korean nouns to the synsets of English WordNet. They built several heuristics for word sense disambiguation (WSD) and combined each heuristic with a decision tree. The approach achieved over 90% of accuracy.

The nature of user-inputted commonsense sentences, OMCS [2], is quite different from that of WordNet. Thus, existing approaches of Korean WordNet construction are

[1] GlobalMind Web Site: http://globalmind.media.mit.edu
[2] EasyMan E/K translator: http://www.clickq.com/

not directly applicable to Korean ConceptNet construction. In addition, there exists no comparable resource to the best of our knowledge.

Our approach has some unique procedures compared to the previous Korean WordNet construction approaches because the coverage of translation is beyond simple nouns; a concept in English ConceptNet can be a noun, compound-noun, phrase, or number.

As in Fig. 2, our approach is divided into largely three parts.

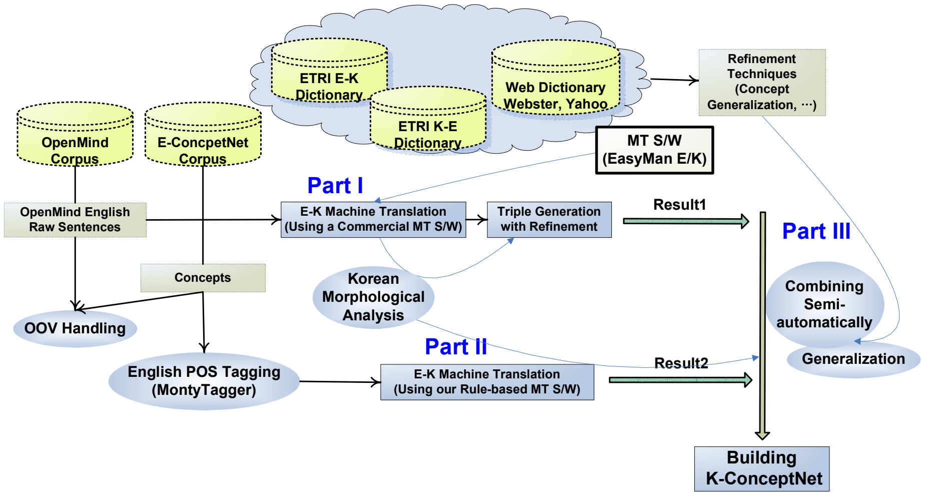

ETRI E-K Dictionary
ETRI K-E Dictionary
Web Dictionary Webster, Yahoo
Refinement Techniques (Concept Generalization, ···)
OpenMind Corpus
E-ConcpetNet Corpus
MT S/W (EasyMan E/K)
Part I
OpenMind English Raw Sentences
E-K Machine Translation (Using a Commercial MT S/W)
Triple Generation with Refinement
Result1
Part III
Concepts
Korean Morphological Analysis
Combining Semi-automatically
OOV Handling
Part II
Generalization
Result2
English POS Tagging (MontyTagger)
E-K Machine Translation (Using our Rule-based MT S/W)
Building K-ConceptNet

Fig. 2. Overall Architecture

(1) **Part I: Translating the English OpenMind corpus into Korean, and converting the result to Korean ConceptNet**

This part uses commercially available machine translation (MT) software, EasyMan E/K translator, to translate English OpenMind raw sentences (e.g. "*Ants are social insects*"). After the E-K translation phase, triples (first order logic style) are generated to be compared with the results of the second part described below (e.g. *<Is-A> <"개미/NNG">,<사회/NNG+ 적/XSN+ 이/VCP+ ㄴ/ETM 곤충/NNG>* where the second argument corresponds to "ant" and the third to "social insect".)

(2) **Part II: Translating the English ConceptNet into a Korean ConceptNet**

The second part is based on the ConceptNet-specific rule-based MT software implemented by us. About 130 rules for E-K translation have been extracted based on our elaborate analysis of the ConceptNet corpus. Simply speaking, the translation follows English translation patterns that most Korean people would agree. The rules can perform English-Korean translation based on part-of-speech (POS) tagged information as in Table 1, and they cover more than 95.2% of the whole English ConceptNet corpus. We implemented this because the commercial MT software generated too many incorrect translation results in the target language.

(3) **Combing the results of Part I and Part II**
The results of the two translation approaches are combined by an algorithm that includes concept generalization. The purpose of the algorithm is to generalize the results of the first and second parts into a more acceptable Korean ConceptNet.

Because OpenMind sentences have words not found in dictionaries (we call them "out-of-vocabulary (OOV) words"), which are usually broken words or newly coined words, they have been handled through our auto-correction word list (pairs of frequently occurring typos and their correct expressions) and a Web dictionary. Since the corpus is in a highly structured short sentence form, and the first sense among the senses of an ambiguous word is correct, we hypothesized that we would have relatively clean translations compared to other texts such as news paper, novel, etc.

2.1 Translating the English OpenMind Using a Commercial MT Software

This part is an attempt to reuse a large amount of commonsense knowledge in English OpenMind and build Korean ConceptNet. We generate Korean translation of the sentence in English OpenMind and subsequently Korean ConceptNet from it. Before selecting EasyMan as our E/K MT software that shows the best translation result, we tested three E/K MT software packages: EasyMan, EnGuide4.0, and Smartran5.0. Although translated sentences are not always complete, we assume that triplets in Korean can be extracted through a set of procedures as follows.

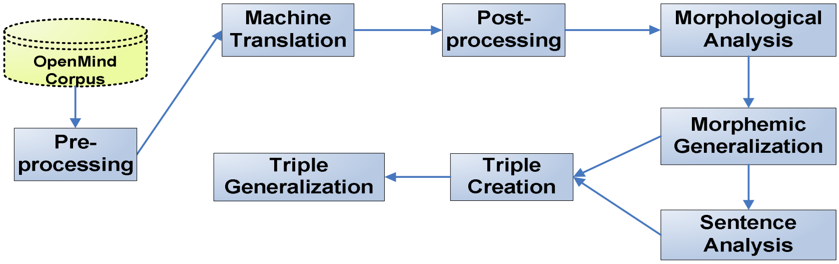


Fig. 3. Concept Generation after MT Translation [Part I]

Fig. 3 shows the overall process of OpenMind translation and concept generation. Because machine translation is still an active research area awaiting a breakthrough, English-Korean translation results of OpenMind have many incorrect sentences. Our simple experiment reveals that the errors are mostly caused by complex sentences, which include those with double quotation marks and long sentences. To alleviate these problems, the OpenMind sentences were preprocessed by the following schemes.

- If a sentence length (number of words in a sentence) is greater than N (currently, N = 30), we remove that sentence. Basically, the goal of OpenMind translation approach is not converting whole OpenMind but gathering as many correct, appropriate Korean sentences as possible. Therefore it is better to reject sentences that may generate translation errors than to achieve 100% coverage.
- Sentences beginning with some specific patterns are not meaningful because they were collected by prompting users with fill-in-the-blank templates to restrict the structure of sentences. The repeating patterns are removed in the translation results. For example, the first part before the colon in the sentence, "*Things that are often found together are: water, people, boat,* " is a template, and only the part after that is extracted and translated.

The Korean sentences generated from the translation process are tagged with part of speech (POS) and then parsed using [7] and [8]. Because of translator errors, some of the translated Korean sentences have a grammatically invalid or awkward structure. In the parsing step, those sentences that have a paring failure are dropped.

Similar to ConceptNet, the concepts of K-ConceptNet are generated from the sentences by using regular expressions, POS information, and syntactic structure information. The difference is that ConceptNet uses shallow paring (chunking) information, whereas K-ConceptNet uses full parsing information. Since the Korean language has free word order unlike English, it is hard to analyze the relationship between two arguments that are extracted from a sentence by using chunking information only.

The procedure for creating a concept from a Korean sentence is as follows:

(1) Pre-defined regular expressions are applied to a sentence. The sentence is tagged with POS, and regular expressions are defined with a lexicon and POS patterns. Since Korean is a very inflective language, we can increase the coverage of the regular expressions by using the POS patterns. Each regular expression is defined with a related predicate. If a sentence is matched with one of the regular expression patterns, arguments are extracted from the pattern, and a concept is generated with the arguments and the pattern-related predicate. Fig. 4 shows examples of translation and concept generation by using regular extractions.

(2) When there is no matched expression pattern, a subject and a predicate of a sentence are extracted by using parse tree information. Then, we remove unnecessary words such as '대부분의(almost all)', '어떤(some, certain)' from each subject and predicate that were extracted, and create a concept with the remained part.

(3) In the next step, created concepts are generalized by word replacement. For example, we replace words like '당신(you)' and '우리(we)' by a general word '사람(person)'.

2.2 Translating the English ConceptNet with Heuristic Translation Rules

The second part is to translate the predicates in English ConceptNet into Korean predicates. Because the OpenMind corpus were already generalized, parsed, and optimized into predicates in English ConceptNet [1], we translate these predicates

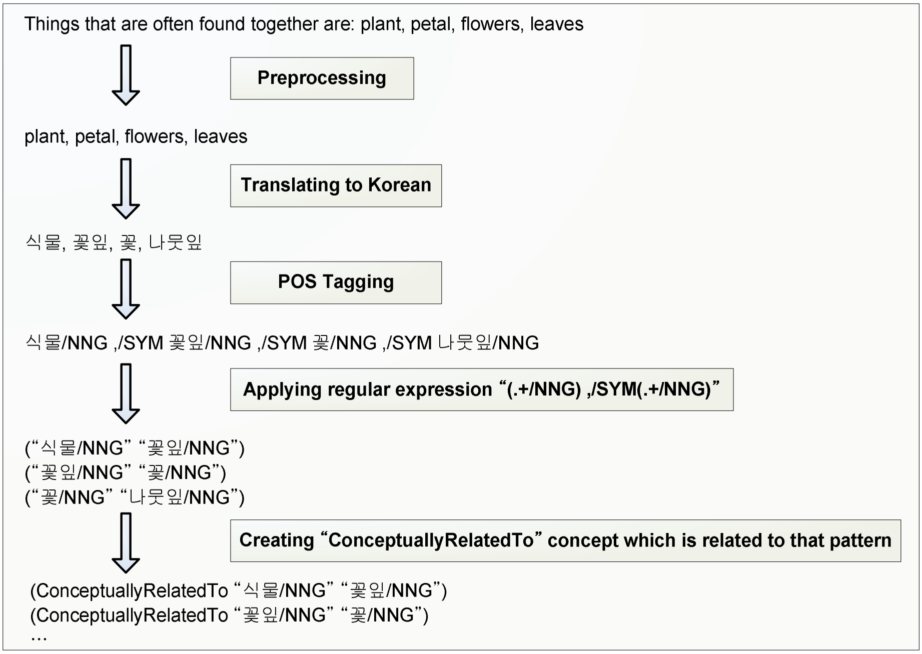

Things that are often found together are: plant, petal, flowers, leaves
Preprocessing
plant, petal, flowers, leaves
Translating to Korean
식물, 꽃잎, 꽃, 나뭇잎
POS Tagging
식물/NNG ,/SYM 꽃잎/NNG ,/SYM 꽃/NNG ,/SYM 나뭇잎/NNG
Applying regular expression "(.+/NNG) ,/SYM(.+/NNG)"
("식물/NNG" "꽃잎/NNG")
("꽃잎/NNG" "꽃/NNG")
("꽃/NNG" "나뭇잎/NNG")
Creating "ConceptuallyRelatedTo" concept which is related to that pattern
(ConceptuallyRelatedTo "식물/NNG" "꽃잎/NNG")
(ConceptuallyRelatedTo "꽃잎/NNG" "꽃/NNG")
...

Fig. 4. An Example for Concept Generation

into Korean for building Korean ConceptNet. To implement our rule-based machine translation (MT) E/K software for Korean ConceptNet building, we have designed the following 5-step procedures (Fig. 5). This is facilitated by the intuition that existing concepts within English ConceptNet are words which can be directly translated by using English-Korean dictionary, simple phrases, or sentences that are interpretable using POS tagged pattern (e.g. "bike," "falling off a bike," and "you get hurt," respectively).

(1) OOV handling: Only the two major types of OOV problems (broken words and newly coined words) were considered because a complete OOV handling requires too much of time-consuming manual efforts. In the current work, about 42.5% (4,430) of the whole OOV words (10,425), which were identified based on ETRI E/K dictionary, have been corrected automatically by using the OneLook dictionary[3], Online Webster dictionary[4], and Yahoo Web E/K dictionary[5].

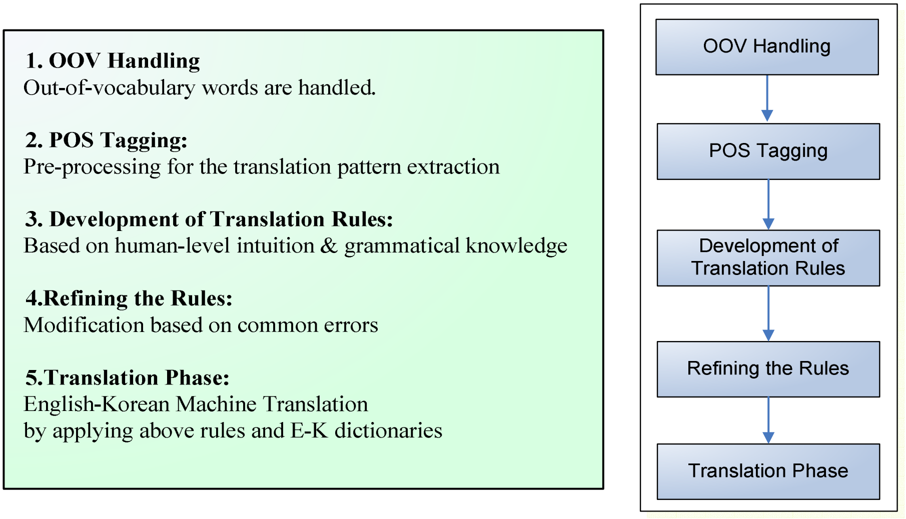
(2) POS tagging: We chose MontyTagger[6], a rule-based part-of-speech tagger based on Eric Brill's transformational-based learning POS tagger [9] which uses a Brill-compatible lexicon and rule files. Through the POS tagging process based on MontyTagger, we could build a base-line for starting a MT.

[3] http://www.onelook.com/

[4] http://www.webster-dictionary.org/

[5] http://kr.dic.yahoo.com/search/eng/

[6] http://web.media.mit.edu/~hugo/montytagger/



1. OOV Handling
Out-of-vocabulary words are handled.

2. POS Tagging:
Pre-processing for the translation pattern extraction

3. Development of Translation Rules:
Based on human-level intuition & grammatical knowledge

4.Refining the Rules:
Modification based on common errors

5.Translation Phase:
English-Korean Machine Translation
by applying above rules and E-K dictionaries

Fig. 5. Rule-based MT [Part II]

(3) Development of translation rules: Based on the result of the previous tagging, a set of E-K translation rules were defined by human's intellectual work. As in Table 1, about 95.2% of the concepts were covered by about 130 translation rules.

(4) Refining the rules: Although a set of translation rules has been developed, there is a potential for POS tagging errors. After checking hundreds of manual POS pattern checking, we have revised the errors to minimize rule-based translation errors.

(5) Machine translation of sentences: By using a machine readable E-K dictionary, which is previously developed by ETRI for a general MT system, we have translated 95.2% of the English concepts in the English ConceptNet. A sample of translation results are shown in Table 2.

2.3 Combining Two Translation Results

To combine the translated Korean concepts that were generated separately by the commercial MT software and our rule-based MT, we have employed a morphological analyzer [7] and heuristics for concept generalization.

For example, if a word 'diagram' is translated to '그림/NNG (picture)' by our rule-based MT module and to '도표/NNG (figure)' by the commercial MT software, this kind of conflict should be resolved. In the subsequent generalization process, we used word in the synonym list is extracted from Korean WordNet[10] manually. The synonym list contains 50 pairs of synonym words.

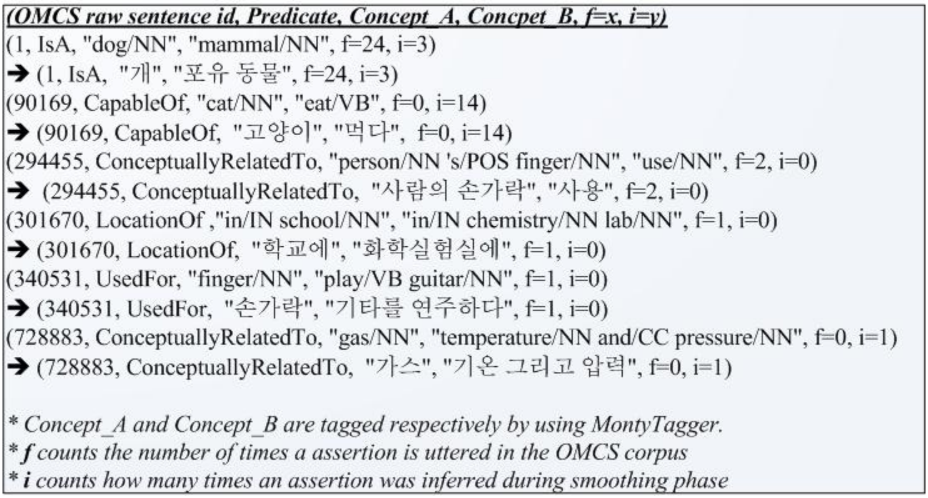
In addition, we remove *josa* , which is a case marker in Korean, playing a role of function word like a preposition or a particle. For example, in a phrase '사람의 손(hand of a person)', the *josa* '의' corresponds to a preposition 'of'. Basically,

Table 1. Extracted POS Patterns of English Concepts

No.	Pattern	Example	Count	Accumulated Ratio	Num. of Concept (*)
1	NN/	dog	25,506,691	51.603%	2,550,691
2	JJ/NN/	social insect	403,140	59.758%	2,953,831
3	NN/NN/	george washington	320,010	66.233%	3,273,841
4	VB/	fly	285,844	72.016%	3,559,685
5	VB/NN/	hold thing	167,418	75.403%	3,727,103
6	NN/POS/NN	person ʻs way	123,588	77.903%	3,850,691
7	JJ	small	59,268	79.102%	3,909,959
8	CD/NN/	4 person	58,467	80.285%	3,968,426
…	…	…	…	…	…
30	VB/NN/POS/NN/	carry person ʼs lunch	9,513	89.591%	4,428,406
31	JJS/NN/	biggest lizard	9,464	89.782%	4,437,870
32	NN/CD/	january 2000	8,717	89.958%	4,446,587
…	…	…	…	…	…
91	VBP/RB/VB/	do not eat	1,757	94.255%	4,458,986
92	VB/IN/NN/POS/NN/	fit in person ʻs garage	1,742	94.291%	4,660,728
…	…	…	…	…	…
127	CD/POS/NN/	one ʻs environment	1,058	95.250%	4,708,141
128	VB/NN/POS/NN/IN/NN	lose person ʻs car under sofa	1,043	95.271%	4,709,184

(*) This means the number of concepts with a duplicate counting permitted.

Table 2. Translation Results



(OMCS raw sentence id, Predicate, Concept_A, Concpet_B, f=x, i=y)

(1, IsA, "dog/NN", "mammal/NN", f=24, i=3)

➔ (1, IsA, "개", "포유 동물", f=24, i=3)

(90169, CapableOf, "cat/NN", "eat/VB", f=0, i=14)

➔ (90169, CapableOf, "고양이", "먹다", f=0, i=14)

(294455, ConceptuallyRelatedTo, "person/NN 's/POS finger/NN", "use/NN", f=2, i=0)

➔ (294455, ConceptuallyRelatedTo, "사람의 손가락", "사용", f=2, i=0)

(301670, LocationOf ,"in/IN school/NN", "in/IN chemistry/NN lab/NN", f=1, i=0)

➔ (301670, LocationOf, "학교에", "화학실험실에", f=1, i=0)

(340531, UsedFor, "finger/NN", "play/VB guitar/NN", f=1, i=0)

➔ (340531, UsedFor, "손가락", "기타를 연주하다", f=1, i=0)

(728883, ConceptuallyRelatedTo, "gas/NN", "temperature/NN and/CC pressure/NN", f=0, i=1)

➔ (728883, ConceptuallyRelatedTo, "가스", "기온 그리고 압력", f=0, i=1)

** Concept_A and Concept_B are tagged respectively by using MontyTagger.*

** **f** counts the number of times a assertion is uttered in the OMCS corpus*

** **i** counts how many times an assertion was inferred during smoothing phase*

'사람의 손' and '사람 손(person hand)' have the same meaning in Korean, and by removing the *josa*, we can combine them. As a result, we have built 200,000 E-K translated assertions.

3 Manual Evaluations of E-K Translated Concepts

An evaluation of the translation quality was carried out with randomly selected 5% of the E-K translation results consisting of 200,000 K-ConceptNet assertions. Each translation is graded by one of the four ranks (described below) by two graduate students, who are Korean native speakers, and their grading measures are given below:

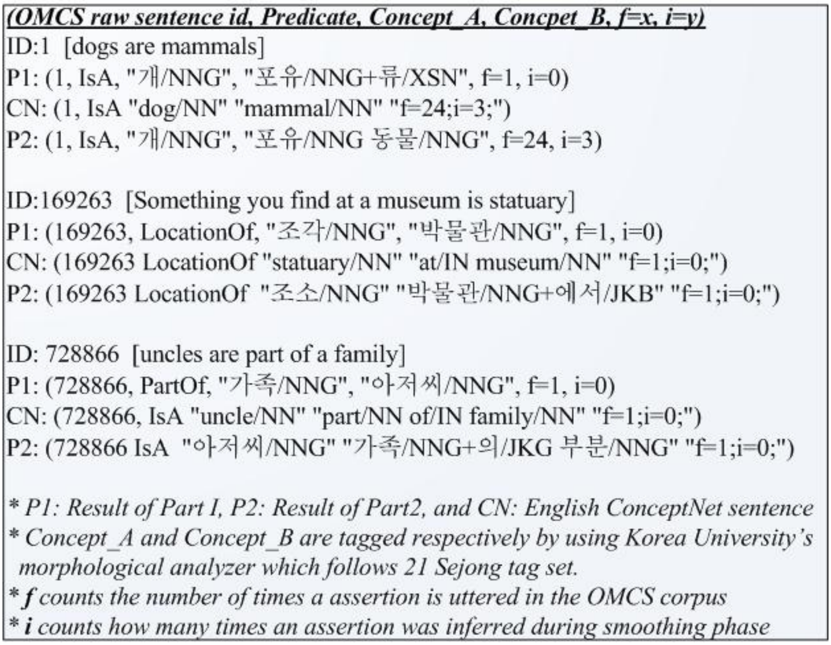
[A] Perfect: No problems in translation. The meaning of the sentence is very clear and no grammatical error of word translation exists.
[B] Good: Easily understandable translation with a minor grammatical error.
[C] Acceptable: The meaning of the sentence can be understood only after several times of reading.
[D] Nonsense: Hard to understand or very ambiguous translation with many errors

Table 3. Translation Accuracy

Category	Rank	Num. of Concepts	Percentage
Good	A	4841	50.07%
	B	1873	19.37%
	C	1446	14.96%
Bad	D	1508	15.60%
Total	A+B+C+D	9668	100%

From this evaluation, we obtained the accuracy of 84.4% assuming that D is a failure (Table3). Based on our analysis the translation errors were due to the lack of context information, insufficient coverage of translation rules, or word sense ambiguities. During the evaluation, the evaluators looked at the English raw sentences of English

Table 4. K-ConceptNet Examples used in Evaluation



(OMCS raw sentence id, Predicate, Concept_A, Concpet_B, f=x, i=y)

ID:1 [dogs are mammals]
P1: (1, IsA, "개/NNG", "포유/NNG+류/XSN", f=1, i=0)
CN: (1, IsA "dog/NN" "mammal/NN" "f=24;i=3;")
P2: (1, IsA, "개/NNG", "포유/NNG 동물/NNG", f=24, i=3)

ID:169263 [Something you find at a museum is statuary]
P1: (169263, LocationOf, "조각/NNG", "박물관/NNG", f=1, i=0)
CN: (169263 LocationOf "statuary/NN" "at/IN museum/NN" "f=1;i=0;")
P2: (169263 LocationOf "조소/NNG" "박물관/NNG+에서/JKB" "f=1;i=0;")

ID: 728866 [uncles are part of a family]
P1: (728866, PartOf, "가족/NNG", "아저씨/NNG", f=1, i=0)
CN: (728866, IsA "uncle/NN" "part/NN of/IN family/NN" "f=1;i=0;")
P2: (728866 IsA "아저씨/NNG" "가족/NNG+의/JKG 부분/NNG" "f=1;i=0;")

* *P1: Result of Part I, P2: Result of Part2, and CN: English ConceptNet sentence*
* *Concept_A and Concept_B are tagged respectively by using Korea University's morphological analyzer which follows 21 Sejong tag set.*
* ***f** counts the number of times a assertion is uttered in the OMCS corpus*
* ***i** counts how many times an assertion was inferred during smoothing phase*

OpenMind, derived English ConceptNet triples, and its translation as well. Table 4 is selected examples that are used in the evaluation.

4 Concluding Remarks and Future Work

We proposed a method for building a Korean ConceptNet by translating English ConceptNet and the original OMCS sentences. The method combines two different sources of translation evidence, i.e., translations from commercial MT software and from a rule-based MT approach. In addition, several NLP techniques have been incorporated, such as OOV handling, POS tagging, automatic rule refinement, morphological analysis, and concept generalization. Finally, based on the challengeable approach, we generated 200,000 K-ConceptNet assertions with reasonably high accuracy and time efficiency.

Through our experiments, we developed a firm belief that our approach can be adoptable to the development of ConceptNet in other languages if machine readable language resources are available and translation patterns from English to the target language can be easily extractable. Although detailed pre-processing and post-processing should be differentiated according to the languages, the overall approach can be generally applied language-independently without too much manual work.

For future work, we have a plan to integrate our work with Korean language part of GlobalMind to extract commonsense knowledge automatically from the Web. For further extension of ConceptNet, we are interested in extracting commonsense knowledge from the existing World Wide Web because a great deal of commonsense is contained in those semi-structured or free text web pages.

For the robustness of Korean ConceptNet, we still need further helps from the general public. As a way to build & evaluate Korean commonsense knowledge, we have launched a web-site[7] where our machine translated results are opened to everybody who access to the web page. Anyone can evaluate existing E-K translated concepts by looking at the original English sentences and participate in inputting corrected commonsense knowledge in Korean.

References

1. Liu, H., and Singh, P.: ConceptNet – A Practical Commonsense Reasoning Tool-kit. BT Technology Journal, (2004), 211-226.
2. Singh, P., Lin, T., Mueller, E. T., Lim, G., Perkins, T., and Li Zhu, W.: Open Mind Common Sense: Knowledge acquisition from the general public. *Proc. of the 1st Int. Conf. on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*, (2002), 1123-1237.
3. C. Fellbaum.: WordNet, *an electronic lexical database*. MIT Press, (1998)
4. Hangeul Society: *Urimal Korean Unabridged Dictionary,* Eomungag, (1997) (in Korean)
5. Moon, Y. J.: Methodology and Techniques for the Design of Korean Noun WordNet. *Proc. of the Natural Language Processing Pacific Rim Symposium,* (1997), 465-469

[7] Korean ConceptNet Web Site: http://k-conceptnet.icu.ac.kr

6. Lee, C. K., *et al.*: Automatic WordNet mapping using word sense disambiguation. *Proc. of Joint SIGDAT Conference on EMNLP/VLC,* (2000), 142-147.
7. Lee, D.-G.: Probabilistic Models for Korean Morphological Analysis and Part-of-Speech Tagging. Ph. D. thesis, Korea University, (2005).
8. Park, S.-Y.: Probabilistic Feature-based Parsing Model for Korea Syntactic Analysis. Ph. D. thesis, Korea University, (2005).
9. Brill, E.: Some advances in rule-based part of speech tagging. *Proc. of the 20th National Conf. on Artificial Intelligence*, (1994), 722-727.
10. Jung, Y.I., Yoon, A.-S., and Kwon, H.-C.: Disambiguation Based on Wordnet for Transliteration of Arabic Numerals for Korean TTS, *Proc. of Computational Linguistics and Intelligent Text Processing (CICLing)*, (2006), 366-377.