Passive Verb Sense Distinction in Korean WordNet

Eun-Ryoung Lee
Korean Language
Processing Laboratory
6203-1, Buid.A6, Pusan
National University, San 30
Jangjeon-dong, Geumjeong-gu
609-735 Busan,
Republic of Korea

easoon@pusan.ac.kr

Ae-Sun Yoon
Korean Language
Processing Laboratory
6203-1, Buid.A6, Pusan
National University, San 30
Jangjeon-dong, Geumjeong-gu
609-735 Busan,
Republic of Korea

Hyuk-Cheol Kwon
Korean Language Processing Laboratory
School of Electrical
& Computer Engineering
Pusan National University,
San 30, Jangjeon-dong,
Geumjeong-gu 609-735
Busan, Republic of Korea

Abstract

During semi-automatic translation of Princeton English WordNet (PWN) into Korean verbs, we noticed that the verbs of accusativity/inaccusativity alternation in English were mapped to two or more Korean verbs of different morpho-syntactic features and thus different senses. These mismatches in mapping show the need for reconstructing the lexical semantic structure of PWN for Korean verb wordnet, which enables distinguishing of lexical semantic features of each verb. The sense distinction of Korean verbs based on their morpho-syntactic features contributes also to improve the consistency of PWN and ensures the accuracy of Korean wordnet.

1 Introduction

Based on our experience in building a Korean wordnet (KorLex) this study aims to propose a useful method of verb sense distinction in constructing multilingual wordnets as well as to improve the consistency of any established wordnet, especially Princeton’s English WordNet 2.0 (hereafter PWN).

In the process of semi-automatic translation of PWN verbs into Korean, we observed that a certain group of English synset translates into two different Korean verb sets, which are a pair of accusative and passive verbs and a pair of accusative and inaccusative verbs. In general, accusativity and inaccusativity in the same English verb for example “whiten,” is often subject to ambiguous sense classification. However, in Korean, accusative, inaccusative, causative, and passive verb forms are generally derived from an active verb by attaching specific infixes, hence, they are morpho- semantically differentiated. We examine especially the pair of accusative and passive verbs in order to provide empirical linguistic criteria in (1) correcting semi-automatic translation errors and (2) positioning the modified word sense in the Korean verb hierarchy.

1 KorLex 1.0 contains only noun synsets (58,656), but it is being updated with about 13,000 verb synsets.

2 The inaccusative verbs, distinct from passive, are the verbs that do not take the objective.

3 There exists in Korean about two hundreds verbs of accusative-inaccusative alternation, called “neutral verb”: for example, ‘umijikada’ (move / move something), ‘idong-hada’ (go, travel / displace something).

This paper is structured as follows: Section 2 examines the way the PWN distinguishes word sense. Section 3 discusses why in Korean the sense of passive verb should be distinguished from its counterpart accusative verb. Section 4 proposes a solution to locating a distinguished passive verb sense in the Korean verb hierarchy.

2 Word Sense Distinction in PWN – Accusative/Inaccusative

PWN is a large-scale lexical database and human knowledge network. In one component it is the result of research on the mental lexicon in psychology, and in the other component of the increasing importance of semantic parsing in Natural Language Processing. In NLP, PWN is a pivot-model wordnet for other multilingual wordnets and a useful resource for cross-lingual translation, information retrieval, language education, and other tasks.

The unique set of synonymous words, called a synset, is the basic constituent of PWN. In PWN, verb synsets contain lexical items that share the same meaning but do not necessarily have the same syntactic structure. A synset is linked to other synsets by means of various semantic relations, including hypernymy, troponymy, antonymy, entailment, and cause relations. Moreover, except for these various semantic relations that a verb sense can have, one of the most important clues to disambiguating verb senses is contained in 35 sentence frames. Sentence frames hold the syntactic information that is dependent on each lexical element. Even though they carry some limitations, sentence frames also express the minimal selectional restrictions on the noun arguments with which the predicate is combined in a sentence. Also, PWN’s binary selectional restriction, ‘somebody’ [+animate] and ‘something’ [-animate], is pertinent to distinguishing the meaning of a verb, especially in the translation of PWN verbs to any non-English language.

2.1 Related studies

Previous studies on the PWN’s verb sense classification have focused, in the context of lexical-functional grammar, on integrating syntactic properties in the description of a verb lexicon. This is because the verb is of central importance in sentence parsing, and to facilitate this task, it is more convenient to convey the syntactic information according to the meaning of the verb in a particular verb lexicon. Consequently, how to arrange the syntactic information on a
verb lexicon and how to describe the verb lexicon in wordnet became important issues. In particular, Kohl et al. [1998] and Zickus [1994] identify the problems of verb sense distinction and classification of PWN, focusing on the verbs of case alternation based on Levin's work [1993].

The latter groups about 3,000 English verbs according to their diathesis alternation patterns. Results show us that if a verb is activated in two different syntactic frames, this has to be reflected in the verb sense description. For example, in 'He broke the glass plate' and 'The glass plate broke', the verb 'break' is accusative and intransitive, respectively. In this respect, Pustejovsky and Busa [1995] argue that the verbs of case alternation should be considered logically as polysemous and that the Levin's verb class is a solution for verb sense disambiguation, event though it is not fully explanatory for the semantic relatedness between the case alternation patterns of the same verb.

The result of Levin's verb classification is largely applied in VerbNet and FrameNet projects and sheds light upon the description of the verb lexicon for NLP purposes. In this respect, the syntactic analysis should be integrated to verb sense distinction in wordnet.

2.2 Accusative/Inaccusative Sense Distinction in PWN

A large-scale lexical database like PWN can contain errors and inconsistence because of manual construction. But there is another reason: the hierarchical structure. It is a very difficult to take into consideration and to reconcile both syntactic and lexical semantic features in a hierarchical structure. In verb.change, we can find three types of sense distinction in the accusative/inaccusative forms of the same verb. Table 1 shows sense distinction in relation to sentence frames in PWN.

In Type A, we pair the accusative and inaccusative forms of the lexical items 'dry,' 'dry out' and they are distinguished consecutively as two different word senses and have different top hypernyms in the hierarchical structure. Type B shows a pair of accusative/ intransitive use of the same verb. They constitute two different synsets and have the same top hypernym. In this case, the latter is {change 0}, of which the frame value is inaccusative: "become different in essence." Type C presents the case in which accusative and inaccusative are classified separately as with word sense. Here, the gloss of the synset is 'make or become___' and presents a dual meaning by "or." In verb.change, 114 verb senses appear as Type C.

2.3 Hierarchical Structure in Accusative/Inaccusative Verbs

We will examine the above three types for accusative/ inaccusative sense distinction according to their hierarchical structures in PWN. First, in Type A, {dry1, dry out5} has its hypernym {change1, alter1, modify11} and {dry0, dry out0}, {change0}. If a synset is defined as having an inaccusative value, its hypernym is also inaccusative.

If a synset has an accusative value, its hypernym is also accusative. For example, {dry1, dry out5} is defined as accusative, and its hypernym is { change1, alter1, modify11 }, which is also accusative. In contrast, for Types B and C, the top-hypernym is { change0 }, which has an inaccusative value. The syntactic and semantic nature of hypernym reveals that the hierarchical semantic relation found in Types B and C is arbitrary in comparison with that found in the type A.

It is undoubted that the semantic features are inherited by hyponym from hypernym, but the grammatical and syntactic features are not inherited in PWN. This is because a syntactic feature is something attributed to each lexical item, but not to a synset as a set of synonymous words. Noteworthy is Levin[1993]'s remark on this matter, that "syntactic patterns accompany systematically the semantic classification."

In Types B and C, the syntactic feature of the top-hypernym does not directly influence to the sense of its hyponyms. However, even if accusativity is a priori a syntactic feature in English, it is a property linked to the meaning of verbs. In this regard, we suggest that the verbs of type C should be revised for sense distinction and the hierarchical structure of type B and C should be modified.

2.4 Accusative/Inaccusative and Cause Relation

In 1.3, we pointed out that in Type A, the accusative and inaccusative of the same verb are classified respectively as constituting independent concepts. In addition, verbs in Type A case are linked to each other according to the cause relation in PWN. The cause relation relates not only the accusative and inaccusative concepts of the same verb, but also the causative and anticausative (or inchoative) concepts. Most of the cause relations in PWN appear in verb.change and the number of linked relations is only 70 pairs of synsets.

Moreover, we remark that the cause relation does not appear between the synsets of Types B and C. For instance, between {powderise0... } and { powderise1... }, there is no cause relation. The fact that the accusative and the inaccusative senses of the pair { powderise0... } and { powderise1... } are not linked by cause relation reveals, above all, the inconsistency of sense distinction and semantic relation. Therefore, the verbs appearing in Types B and C should be re-classified with their syntactico-semantic features— accusative/inaccusative, causative/anti-causative.

3 Morpho-Syntactic Properties of Korean Passive Verb in KorLex Text

In this section, we discuss in detail the verb sense of Type C and its Korean counterpart. The accusative/inaccusative verbs discussed in Section 1 will be examined here for concept mapping difficulties in translation. PWN verbs of Type C are translated in Korean either as a pair of (1) accusative/inaccusative or as a pair of (2) accusative/passive verb forms. In Korean, unlike English, the accusative/inaccusative/passive forms of a verb are morphologically distinct, which is ascribable to postpositional morphemes. They also have different syntactic and semantic fea-
Table 1: Accusative/Inaccusative Sense Distinction in PWN

<table>
<thead>
<tr>
<th>Type of taxonomy</th>
<th>Synset words</th>
<th>Gloss</th>
<th>Frame</th>
<th>Top hypernym</th>
<th>Korean counterpart</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type A</td>
<td>dry1, dry out5</td>
<td>remove the moisture from and make dry</td>
<td>8_0(Somebody —-s something) 11_0(Something—-s something)</td>
<td>change1, alter1, modify11</td>
<td>mal-hi-da, geonjo-hada</td>
</tr>
<tr>
<td></td>
<td>dry0, dry out0</td>
<td>become dry or drier</td>
<td>1_0(Something —-s)</td>
<td>Change0</td>
<td>maleuda geonjo-doeda</td>
</tr>
<tr>
<td>Type B</td>
<td>powderize1, pulverize1, powderise1</td>
<td>become powder or dust</td>
<td>8_0(Somebody —-s something) 11_0(Something —-s something)</td>
<td></td>
<td>galu-lo deoda</td>
</tr>
<tr>
<td></td>
<td>powderize0, powderise0, pulverize0, pulverise0</td>
<td>make into a powder by breaking up or cause to become dust</td>
<td></td>
<td></td>
<td>galu-lo-mandeulda, bbahda</td>
</tr>
<tr>
<td>Type C</td>
<td>grill0</td>
<td>cook over a grill</td>
<td>1_0(Something —-s) 8_0(Somebody —-s something)</td>
<td>Change 0</td>
<td>gub-da gu-wojida</td>
</tr>
</tbody>
</table>

3 In narrow sense, the passivisation of verb means the (1) but J-T, Lee[2001] reports the passive verbs of (2) show the highest frequency of passive forms in contemporary Korean. In our study, we consider all these three categories as passive verb.
Examples 3 and 4 show that passive verb can have more than one structure. Each passive verb has its restrictions regarding syntactic structure, thus Korean passivisation is not a syntax-dependent but a lexical-dependent process. Moreover, the fact that not all Korean accusative verbs can take the passive form shows that the sense of the verb determines the specific sentence structure. In this respect, we will argue that it is indispensable to distinguish the passive verb sense from the accusative verb sense in Korean.

### 3.2 Types of Error in Translation of PWN into Korean

In this section, we will narrow the scope of analysis to Korean counterparts of the verbs of Type C in Table 4. To find the errors, we extracted the English verb synsets in verb. change that have at least one sentence frame from (a) representing inaccusative value and at least one sentence frame from (b) representing accusative value.

<table>
<thead>
<tr>
<th>Table 2: Accusative/inaccusative sentence frames</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>(b)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

In addition, we checked whether the gloss contains more than two senses. Even though the sentence frames satisfy the condition of synset extraction, it is often not described explicitly in the gloss. It appears that in most cases the sentence frames more accurately catches the accusative/inaccusative senses than the gloss. We also counterchecked if the sentence frames of PWN are conformed by the Merriam-Webster online dictionary. As a result, we achieved 117 synsets having both accusative and inaccusative senses and discarded the 2 synsets containing idiomatic expression.

Automatic translation by the bilingual MRD often does not fully reflect the English verb meaning in Korean verbs when they contain both accusative and inaccusative senses at the same time.

Example 6. {burst1, burst0}

Sentence frame: Somebody ___ s something

Gloss: break open or apart suddenly and forceful

Automatic translation: teo-jida, payeol-hada, payeol-deoda

Missing Korean counterpart: teotteulida, payeol-sikida, payeol-deoda

Example 7. {steam0}

Sentence frame: Somebody ___ s something

Gloss: cook something by letting steam pass over it

Automatic translation: jjida

Missing Korean counterpart: jjyeo-jida

Example 6 illustrates the case in which the automatic translation gives only inaccusative forms and needs to be supplemented with the accusative verbs such as "teotteulida" and "payeol-sikida." Example 7 shows the case in which the automatic translation gives only the accusative verb and the inaccusative form is not selected. Sentence frames are important criteria to evaluate the automatic mapping correctness and to refine the selection of lexical items for KorLex. By these criteria, we correct the translation errors and achieve the results shown in Table 4.

In Section 2 we corrected the translation errors with the accusative/inaccusative values. In this section, we will examine the pair distribution of Korean verb forms shown in Table 4 and on the basis of passive sense distinction, we will propose a way to apply it to the KorLex verb hierarchy.

### 4 Building KorLex Verb Hierarchy

In Section 2 we corrected the translation errors with the accusative/inaccusative values. In this section, we will examine the pair distribution of Korean verb forms shown in Table 4 and on the basis of passive sense distinction, we will propose a way to apply it to the KorLex verb hierarchy.

#### 4.1 Sense distinction: accusative/passive form

In Table 4 we have four different distributions of verb pairs. Cases 1 and 2 show that the mapped Korean verbs are respectively accusative and inaccusative. Cases 1 and 2 illustrate the difference in lexicalization and conceptual disparity between languages. These are two of the common problems that one encounters in building multilingual wordnets. But Cases 3 and 4 reveal the problem of sense distinction, which requires a solution. Cases 3 and 4 arise from the sense distinction problem as well as the morpho-syntactic differences of lexic unit. When the accusative/inaccusative senses are encoded as one word sense in English, and if the Korean counterpart for the same concept is presented as a pair of accusative/passive verbs, accusative/passive verb forms should be semantically distinguished for the following two reasons ({steam0} in PWN is mapped to {jjida, jjyeojida} in Korean).

Example 6.

**na-neun gamja-leul jji-n-da**

I-nom potato-acc steam-present-ending

(I steam the potatoes)

Example 7.
Table 3: 115 accusative/inaccusative synsets from translation results

<table>
<thead>
<tr>
<th>Korean verb form form</th>
<th>accusative</th>
<th>inaccusative</th>
<th>accusative/passive</th>
<th>accusative/inaccusative</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>synsets</td>
<td>50</td>
<td>15</td>
<td>7</td>
<td>43</td>
<td>115</td>
</tr>
</tbody>
</table>

Table 4: 115 accusative/inaccusative synsets from correction results

<table>
<thead>
<tr>
<th>Korean verb form form</th>
<th>accusative</th>
<th>inaccusative</th>
<th>accusative/passive</th>
<th>accusative/inaccusative</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>synsets</td>
<td>20</td>
<td>2</td>
<td>34</td>
<td>65</td>
<td>115</td>
</tr>
</tbody>
</table>

4.2 Passive Verb classification and its place in KorLex

Now, how can we build the hierarchical structure for the distinctive senses of accusative and passive verb forms in the semantic net? In KorLex, since verbs hierarchy is biased with PWN structure, it is efficient to look for the appropriate place in relation to the top-hyponym and its hyponyms in PWN.

In Table 1, the top-hyponyms of Type A (pairs of accusative and inaccusative concepts) are the verb concepts that reflect the accusative/inaccusative values independently: \{change 0\} is inaccusative and \{change1, alter1, modify1\}, accusative. Given the 35 top hyponyms of verb.change, we can choose, following their syntactic and semantic features, an appropriate place in the hierarchies for the Korean passive or accusative verb. To facilitate the task, we can also take into consideration the existing cause relation between the top hyponyms of the accusative and the inaccusative sense. The Figure 1 represents the hierarchical structure of \{bake0\}.

\[
\begin{align*}
&=> change 0 \\
&=> change\_integrity 0 \\
&=> cook 0 \\
&=> bake 0
\end{align*}
\]

Figure 1: Hierarchy of \{bake0\}

In the hierarchy of \{bake0\}, all the hyponyms are of inaccusative value. We will use this structure to match the passive verb "gu-woeji-da" (bake-passive-ending), and the Korean counterparts for the hyponyms of \{bake0\} will be all of passive value. Now, we can separately build the accusative and the passive forms of "gub-da" (bake-ending) in the Figure 2.

In Figure 2, the left cell shows the hierarchy of the passive verb form of "guwoejida" and its hyponyms of passive value. The right cell represents the modified structure of the accusative verb form and its hyponyms of accusative value. Each counterpart of \{bake0\} now has a different semantic relation; they are linked by cause relation and "accusative-passive" tag in KorLex. As result, the English synset \{bake0\} is mapped to the accurate Korean synonym "guwoejida" and \{cook0\} is mapped to "ikda". Table 5 shows the results of sense distinction according to accusative/passive in Korean and the corresponding hierarchy modification.

Before sense distinction was performed, accusative and passive verbs coexisted in synsets. After the sense distinction, there were 74 synsets: 34 synsets representing the sense of accusative verbs, 34 synsets, passive verbs, and we added 6 new synsets located between the top node and the accusative verb synsets. This was done because it was necessary to represent the hierarchical structures of the passive and accusative verb concepts independently. In verb.change of PWN, there are 6 upper nodes, including \{decay\}, \{change\_integrity\}, \{change\_surface\}, \{change\_magnitude\}, \{change\_state\}, \{change\_by\_reversal\}, which reflect only the sense of inaccusative verbs. Therefore, we distinguished the senses of these nodes and created Korean verb synsets to appropriately represent the semantic hierarchies of accusative and passive verbs.

Concerning the hierarchy modification, we used two top synsets in verb.change: \{change1, alter1, modify1\} for the accusative verb synsets and \{change 0\} for the passive verb synsets translated respectively, 'byeonwha-sikida' and 'byeonwha-deoda.'

Conclusion and Future Work

While focusing on the morpho-semantic features in Korean accusative/passive verbs, we argued that unlike English verbs, the Korean passive voice is tributary to the lexical level and thus its lexical feature has to be reflected to constitute a distinct verb sense in KorLex. When a word
Figure 2: Reconstruction of hierarchy for Passive/Accusative verb of “gubda”.

Table 5: Results of sense distinction and hierarchy modification

<table>
<thead>
<tr>
<th>Sense distinction</th>
<th>Hierarchy modification (top node)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>accusative</td>
</tr>
<tr>
<td>Translation of PWN</td>
<td>34</td>
</tr>
<tr>
<td>KorLex</td>
<td>34(+6)</td>
</tr>
</tbody>
</table>

sense is differentiated by its distinctive semantic feature, it is also indispensable to reorganize the semantic relations. The morpho-semantic feature of a concept is heterogeneous throughout the languages, and should be considered seriously in the multilingual lexical databases, but at the same time can be a robust linguistic criterion for sense distinction and concept classification. The method proposed in this study can be useful to any multilingual lexical database, in the fact that it shows the possibility to improve the consistency of one wordnet via another wordnet.

However, we think that much work remains to be done, in particular, finding the distinctive semantic features of verbs to take into consideration for building Korean wordnet and determining consistent and reusable criteria for word sense distinction.

Acknowledgements
This work was supported by Korean National Research Laboratory Program (M1-0412-00-00 28-05-J00-00-014-10)

References