

Processing Idiomatic Expressions in Korean*

Chung Chan

Dongseo University

Jong-Bok Kim

Kyung Hee University

Canonical idioms are in general syntactically opaque and semantically non-compositional. However, there also exist many idioms whose component parts are rather morpho-syntactically transparent and semantically decomposable. This paper identifies two types of Korean idioms and review their basic properties. Based on these empirical studies, the paper then proposes a constraint-based approach that can allow us to process these two different types of idioms.

Keywords : Korean idioms, decomposable, non-decomposable, constraint-based, HPSG

* We thank anonymous reviewers for the constructive comments and suggestions. This work was supported by the Korea Research Foundation Grant (KRF-2005-042-A00056).

† 교신저자: 김종복. 경희대학교 영어학부, 연구분야: 통사, 의미론, 전산언어학
E-mail: jongbok@khu.ac.kr

Introduction

Idioms are well-known for their opaque or fixed properties in the sense that they resist undergoing any syntactic operations and that their meaning cannot be predicated from the meanings of their component parts (Katz 1973, Chomsky 1980, Machonis 1985, MaCawley 1983, van der Linden 1992, among others). For example, the English idiom *kick the bucket* has nothing to do with kicking or with buckets. It simply means 'to die' and is opaque to syntactic operations:

- (1) Non-decomposable idioms
 - a. Pat kicked the bucket.
 - b. *Pat kicked several buckets.
 - c. *The bucket, Pat kicked.
 - d. *The bucket was kicked by Pat.

Does this mean that idioms must be registered in the dictionary as a whole unit? As noted in Numberg et al. (1994), unlike such semantically 'non-decomposable' idioms, there also exist 'decomposable' idioms. For example, an idiom like *spill the beans* allows internal quantification, topicalization, passivization, and raising as well:

- (2) Decomposable idioms
 - a. Pat spilled the beans.
 - b. Don't look at me. I didn't spill a bean.
 - c. The beans, Pat already spilled.
 - d. The beans were spilled by Pat.
 - e. The beans continue to appear to be certain to be spilled.

Such decomposable idioms can undergo various syntactic processes as shown in (2b)

- (2d), supporting the idea that they cannot be registered in the lexicon just as a unit. Further, in terms of semantics, the components of such decomposable idioms have relatively more transparent semantic roles and behave like syntactic constituents in a certain way. For example, in (2), *spill* can be considered to mean something like ‘divulge’, and *the beans* something like ‘secrets’ (cf. Nunberg et al. 1994, Jackendoff 1997, McGinnis 2002).

Considering every language has idiomatic expressions, we expect Korean may also have ‘decomposable’ and ‘non-decomposable’ idioms. In this paper, we will look into the basic properties of these two different types of Korean idioms and provide an analysis that can give us answers to how we store idioms in the lexicon and then process them in syntactic environments. In what follows we will first identify the two types of Korean idioms and review their syntactic and semantic properties. We then offer a syntactic analysis of these decomposable idioms within the framework of construction-based HPSG (Head-driven Phrase Structure Grammar).

Two Types of Korean Idioms

Non-decomposable Idioms

One salient feature of idioms is that their meanings are not composable from their parts. For example in (3), it is impossible to induce the idiomatic meaning of ‘not given’ from the strings *kkwueng kwue mekun* ‘roasting and eating a pheasant’ (cf. Park 2003):¹⁾

(3) Mary-ka mikwuk-ulo tolaka-senun, kkwueng kwue mek-un sosik-ita

1) The glosses used in this paper are as follows: Acc (accusative), Comp (complementizer), Cop (copula), Decl (declarative), Nom (nominative), Mod (modifier), Nmlz (nominalizer), Pass (passive), Top (topic), etc.

Mary-Nom America-to return-after pheasant roast ate-Mod. news-Cop
'After Mary returned to America, there has been no news from her at all (we didn't hear from her at all).'

The components of such an idiom are strictly fixed in the sense that they do not allow an inflectional ending, case variation, let alone the word replacement:

- (4) a. *talk kwue mek-un sosik (word replacement)
chicken roast eat-Mod. news
b. *kkwueng kwue tusi-n sosik (honorific inflection)
pheasant roast eat.Hon.-Mod. news
c. *kkwueng kwue mek-ess-ta-un sosik (tense-mood inflection)
pheasant roast ate-Past-Decl.-Mod. news
'literal only: the news that (someone) roasted and ate a pheasant'
d. *kkwueng-ul kwue mek-un sosik (overt case marking)
pheasant-Acc roast ate-Mod. news

In (4a), replacing the constituent 'pheasant' by a similar word 'chicken' makes the idiom a nonsense. The verbal part *mek-un* cannot take its honorific form *tusi* as in (4b) or a long quotative form *mek-ess-ta-nun* as in (4c). Note that even the noun 'pheasant' functions as an object of the predicate *mek-un*, it cannot have an accusative case marking as shown in (4d).

Such fixed properties of idioms can be observed in other cases too:

- (5) a. mwul cha-n ceypi
water kick-Mod. swallow
'Idiomatic: very neat/swift/slim looking'
'Literal: a swallow that kicks water'

- b. kkwueta nohun poli-calwu
borrow put-Mod. barley-sack
'Idiomatic: dumb like a wallflower'
'Literal: a barley sack borrowed'
- c. totholi khi cay-ki
acorn height measure-Nominalizer
'Idiomatic: not much difference'
'Literal: to measure the height of acorns'

We can easily notice that the components of these idioms are strictly fixed and thus can neither have an overt case marking nor occur with a tense marking:

- (6) a. mwul-ul cha-n ceypi (overt case marking)
water-Acc kick-Mod. swallow
'Literal only: a swallow that kicks water'
- b. kkwueta noh-ass-ten poli-calwu (verbal inflection change)
borrow put-Past-Adnom barley-sack
'Literal only: a borrowed barley sack'
- c. totholi-lul khi-lul cay-ki (overt case marking)
acorn-Acc height-Acc measure-Nmlz
'Literal only: to measure the height of acorns'

The idioms we have seen in this section are considered to be 'non-decomposable' since their parts are opaque from morphological or syntactic processes, implying that they are atomic units which lack internal structures. These fixed phrasal expressions or frozen forms will be listed in the lexicon as a whole, paired with their idiomatic meanings.

Decomposable Idioms

Unlike such opaque or non-decomposable idioms, the language also has grammatically transparent or decomposable idioms. Many of the Korean idioms belong to this type even though there are some differences in the degree of idiomaticity or decomposability (cf. Park 2003). In most syntactically flexible idioms, each component of the decomposable idiom plays a certain role in determining the whole idiomatic meaning. For example, in the idiom (7a), the NP *kkoli* 'tail' can be considered to mean something like 'seduction', while *chi-ta* 'wag' something like 'make'. These meaning components compositionally contribute to the idiomatic meaning 'seduce':

- (7) a. *kkoli-lul chi-ta*
tail-Acc wag-Decl
'Idiomatic: seduce'
'Literal: wag the tail'
- b. *kan-i khu-ta*
liver-Nom big-Decl
'Idiomatic: be very confident/brave-Decl'
'Literal: have a big liver'

As is well-known, most of the idioms have a metaphoric usage and their meanings are conventionalized. That is, the literal meaning of each component is accommodated to a certain idiomatic meaning by linguistic convention. It seems that the easier the accommodation is, the more readily the components can go through some syntactic processes. Most of the decomposable idioms allow their components (usually nominal parts) to go through syntactic processes such as scrambling, optional case marking, and external modification. However, many of them do not allow syntactic processes such as internal modification, topicalization, pronominalization, and passivization. And none of them allow cleft and afterthought. In this section, we will observe which

idioms can go through such syntactic processes and discuss what makes the difference.

Scrambling: Most of the decomposable idioms allow scrambling of their components while keeping their idiomatic meaning:

- (8) a. Mary-ka [komwusin-ul kekkwulo sinesta].
 Mary-Nom rubber.shoes-Acc backwards wore
 ‘Idiomatic: Mary betrayed the love.’
 ‘Literal: Mary wore rubber shoes backwards.’
 b. Komwusin-ul Mary-ka kekkwulo sinesta.
- (9) a. Mary-ka namphyen-hanthey simha-key [pakaci-ul kulkessta].
 M-Nom husband-to hard gourd.dipper-Acc scraped
 ‘Idiomatic: Mary nagged her husband hard.’
 ‘Literal: Mary scraped a gourd dipper to her husband.’
 b. Mary-ka namphyen-hanthey pakaci-lul simha-key kulkessta.
- (10) a. Mary-ka ku sihem-eyse cengmal [miyekkwuk-ul mekessta].
 M-Nom the exam-in really seaweed.soup-Acc ate
 ‘Idiomatic: Mary really failed the exam.’
 ‘Literal: Mary ate seaweed soup during the exam.’
 b. ?Mary-ka ku sihem-eyse miyekkwuk-ul cengmal mekessta.

In Korean, as illustrated in (11) when two NPs have the identical case marking, their linear order generally cannot be switched with each other.

- (11) a. Sacang-i ku chinkwu-lul phal-ul capassta.
 president the friend-Acc arms hold
 ‘The president hold the friend by the arm’
 b. *Sacang-i phal-ul ku chinkwu-lul capassta

The same observation can be made in the decomposable idioms as in (12):

- (12) a. Sacang-i ku chinkwu-lul mok-ul calassta.
president the friend-Acc neck-Acc cut
'Idiomatic: The president fired the friend.'
'Literal: The president cut the friend's neck'
b. *Sacang-i mok-ul ku chinkwu-lul calassta.

Case and Particle: In most decomposable idioms, case markings are optional as in most non-idiomatic nominals.²⁾

- (13) a. Mary-ka komwusin(-ul) kekkwulo sin-ess-ta.
M-Nom rubber shoes-Acc backwards wear-Past-Decl
'Idiomatic: Mary betrayed the love.'
b. Mary-ka ku sihem-eyse miyekkwuk(-ul) mek-ess-ta.
M-Nom the exam-in seaweed soup-Acc eat-Past-Decl
'Idiomatic: Mary failed the exam.'
c. John-i kan(-i) khu-ta.
J-Nom liver-Nom big-Decl
'Idiomatic: John is very brave.'

Unlike the non-decomposable idioms, a component of the decomposable idioms can cooccur with a delimiter like *man* 'only' or *to* 'also' when given a proper context:

- (14) a. ku-lul mok-man calu-n kes-i anila,...
he-acc neck-only cut-Mod kes-Nom not, ...

2) The null case marking often occurs in an informal speech style, making examples like (13) more natural when the main verb ends with an informal declarative ending like *-e*.

'Idiomatic: (He) not only fired (him), but also...

b. kan-to khu-ci, kuleskey hako.

liver-also big-Comp, so do

'Idiomatic: Seeing that, he is also brave,...'

Modification: Most of the decomposable idioms allow their components to be modified by an adverbial expression:

(15) a. Mary-ka komwusin-ul wancenhi kekkwulo sinessta.

M-Nom rubber shoes-Acc completely backwards wore

'Idiomatic: Mary completely betrayed the love of her sweetheart.'

b. Mary-ka namphyen-hanthey pakaci-ul simhakey kulkessta.

M-Nom husband-to gourd dipper severely scraped

'Idiomatic: Mary severely nagged her husband.'

c. Mary-ka ku sihem-eyse miyekkwuk-ul twupen mekessta.

M-Nom the exam-in seaweed soup-Acc twice ate

'Idiomatic: Mary failed the exam two times.'

d. Sacang-i ku chinkwu-uy mok-ul tan khal-ey calassta.

president the friend-Gen neck-Acc one slash-at cut

'Idiomatic: The president fired the friend all of a sudden.'

However, adnominal modification is in general not allowed, which can be expected if we assume that the modified nominal component of the idiom has no semantic meaning:

(16) a. Mary-ka kemceng komwusin-ul kekkwulo sinessta.

M-Nom black rubber shoes-Acc backwards wore

'Literal only: Mary wore the black rubber shoes backwards.'

b. Mary-ka namphyen-hanthey kun pakaci-ul kulkessta.

M-Nom husband-to big gourd dipper scraped
'Literal only: Mary scraped a big gourd dipper to her husband.'

In (15) and (16), the intended idiomatic meanings are all predicative (e.g., 'betray', 'nag', 'fail', and 'fire'), and thus only external modification with an adverbial is allowed. However, an internal modification with an adnominal form is also allowed when the combined meaning of the adnominal and noun is suitable to the whole predicative idiom meaning. For example, when the adnominal *kum* 'big' in (16b) is replaced with *simhan* 'severe', the expression carries the idiomatic meaning as shown in (17), suggesting that the modification compatibility with an adnominal is in principle a matter of semantics:

(17) Mary-ka namphyen-hanthey simhan pakaci-lul kulkessta
M-Nom husband-to severe gourd dipper scraped
'Idiomatic: Mary severely nagged her husband.'

Considering the semantic compatibility, it is not difficult to find examples in which the nominal component of the idiom is modified by an adjectival element:

- (18) a. Mary-ka ta toyn il-ey sikhemen cay-lul ppwuleyssta.
M-Nom all done work-at pitch-black ash-Acc scattered
'Idiomatic: Mary completely bungled the work all done.'
'Literal: Mary scattered pitch-black ash over the work all done.'
- b. Mary-ka cinan pam-ey emcheng kun cito-lul kulyessta.
M-Nom last night-at very big map-Acc drew
'Idiomatic: Mary wetted the bed soaked last night.'
'Literal: Mary drew a very big map last night.'

What we can conjecture from the adnominal modification is that when the

nominal of an idiom can be linked to a metaphoric meaning as in English idioms like *spill the beans* and thus can have a certain meaning, it can be modified by an adjectival expression.

One interesting fact about the decomposable idioms is that there also exist idioms selecting an adverbial element as the obligatory complement. For example, the adverb *kekkwulo* ‘backwards’ in (19a) and the adjective *ttukewun* ‘hot’ in (19b) are both required to induce the intended idiomatic meanings:

- (19) a. Mary-ka {komwusin-ul (kekkwulo) sinesta}.
 M-Nom rubber shoes-Acc backwards wore
 ‘Literal only without the adverbial: Mary wore rubber shoes.’
- b. Mary-ka {(ttukewun) mas-ul poassta}.
 M-Nom hot taste-Acc saw
 ‘Idiomatic with the adnominal: Mary had bitter experiences.’
 ‘Literal only without the adnominal: Mary tasted something.’

Passivization: The nominal object of some decomposable idioms can be realized as the subject of passive, further evidencing the syntactic independence of decomposable idioms:

- (20) a. Machimnay ku chinkwu mok-i cal-i-essta.
 at last the friend neck-Nom cut-Pass-Past
 ‘Idiomatic: At last the friend was fired.’
- b. Ta tyon il-ey cay-ka pwulye-ci-essta.
 all done work-at ash-Nom scatter-Pass.-Past
 ‘Idiomatic: The work all completed was bungled.’

However, not all decomposable idioms allow passivization as shown below:

- (21) a. *Kkoli-ka (Mary-eyuyhay) ku namca-eykey chie-ci-essta.
 tail-Nom M-by the man-to wag-Pass.-Past
 ‘Literal only: The tail was wagged (by Mary).’
- b. *Miyekkwuk-i (Mary-eyuyhay) ku sihem-eyse mek-hi-essta.
 seaweed soup-Nom M-by the exam-in eat-Pass.-Past
 ‘Literal only: The seaweed soup was eaten by Mary during the exam.’

Here note that examples in (21a) and (21b) are awkward even in a non-idiomatic reading. This suggests that the contrast between (20) and (21) seems to have little to do with the idiomaticity. The contrast rather seems to be related to semantic or pragmatic factors.

Topicalization: The possibility of topicalizing a component of the decomposable idioms varies among the speakers. Even though the judgment is unstable, topicalization of the nominal elements in many idioms are allowed as shown below:³⁾

- (22) a. ??Miyekkwuk-un Mary-ka ku sihem-eyse mekessta.
 seaweed soup-Top M-Nom the exam-in ate
 ‘Literal only: As for the seaweed soup, Mary ate it during the exam.’
- b. ??Ku chinkwu-uy mok-un sacang-i calassta.
 The friend-Gen. neck-Top president-Nom cut
 ‘Literal only: As for the friend's neck, the president cut it.’
- c. ??Cay-nun Mary-ka ta toyn il-ey ppwuleyssta.
 ash-Top M-Nom all done work-at scattered
 ‘Literal only: As for the ash, Mary scattered it on the work all done.’
- b. ??komwusin-un Mary-ka kekkwulo sinesta.
 rubber shoes-Top M-Nom reversely wore

3) See Schenk (1995) for a similar proposal for English idioms.

‘Literal only: As for the rubber shoes, Mary wore them backwards.’

The nominal component of the decomposable idioms usually does not carry a specific referent. For example, when (22a) is used in an idiomatic sense, *miyekkwuk* ‘seaweed soup’ does not refer to a specific object but rather refers to a more abstract individual like ‘failure’. The paraphrases of the intended idiomatic meanings will then be something like the following:

- (23) a. #As for failure, Mary did it in the exam.
b. #As for my friend's job, the president fired him.
c. #As for bungle, Mary did it on the work.
d. #As for love, Mary betrayed her sweetheart.

What we can notice is that it is relatively harder to connect the topic and the contents of the subsequent comment clause.

Topicalization is not always impossible in the decomposable idioms. Most speakers allow topicalization in idioms like the following with no difficulty:

- (24) a. Yangtali-nun Mary-ka cal kelchinta.
both-legs-Top M-Nom well lay-over
‘Idiomatic: Mary is good at opportunism.’
‘Literal: Mary lays both legs (over both sides).’
b. Kkoli-nun Mary-ka mence chiesta.
tail-Top M-Nom first wagged
‘Idiomatic: Mary made seduction first.’
‘Literal : Mary wagged her tail first.’

Though there seems to be no clear reason why the examples in (24) are better than (23), but here again the easiness of accommodation seems to take place. That

is, it is relatively easier to accommodate the literal meaning of *yangtali* 'both legs' to an idiomatic meaning like 'opportunism', and *kekoli* 'tail' to 'seduction'. Also note that in more fine-grained meaning paraphrases, the connection between the topic and its comment in (25) is much more natural than the connection in (23):

- (25) a. As for the opportunism, Mary is very good at it.
b. As for the seduction, Mary first made it.

These observations suggest that the lack of topicalization in most of the decomposable idioms has little to do with syntax but is more closely related to extragrammatical factors such as language processing.

Pronominalization: A nominal in most decomposable idioms cannot be replaced with a pronominal. In examples below, the idioms with the pronominal replacement carry only literal meanings, which brings about the communication breakdown or nonsense joke readings.

- (26) A: Mary-ka ku sihem-eyse miyekkwuk-ul mekessta.
M-Nom the exam-in seaweed soup-Acc ate
'Mary failed in the exam.'
B: #John-to kuke mekesse. (kuke = miyekkwuk)
J-Also it ate
'Literal only: John also ate it.'
- (27) A: Mary-ka komwusin-ul kekkwulo sinesta
M-Nom rubber shoes-Acc backwards wore
'Mary betrayed her lover.'
B: #Leslie-to kuke kekkwulo sinesse. (kuke = komwusin)
L-Also it backwards wore
'Literal only: Leslie also wore it backwards.'

- (28) A: Mary-ka namphyen-hanthey pakaci-ul kulkessta.
M-Nom husband-to gourd dipper scraped
‘Mary nagged her husband.’
- B: #Leslie-to caki nampyen-hanthe kuke kulkesse. (kuke = pakaci)
L-Also self husband-to it scraped
‘Literal only: Leslie also scraped it to her husband.’

According to Nunberg et al. (1994), one of the features of an idiom is that the idiomatic meaning is available only when all the components of the idiom co-occur. In examples like (26) - (28), the nominal components are replaced with the pronominal *kuke*, making their specific phonological forms unavailable for the other components. Such a violation of the co-occurrence restriction appears to be responsible for the absence of the idiom meanings in (26B) - (28B).

Relativization: The nominal components of the decomposable idioms tend not to be relativized. All of the following examples where the nominal is modified by a relative clause carry only the literal meanings, which are semantically all weird:

- (29) a. Mary-ka {[kekkwulo sin-un] komwusin] ttaymwuney, ku namca-nun
M-Nom backwards wear-Mod. rubber-shoes because-of the man-Top
maum-i sanghayssta.
mind-Nom hurt
‘Literal only: Because of the rubber shoes that Mary wore backwards, the man felt bad.’
- b. Sacang-i {[calu-n] mok] ttaymwuney, ku namca-nun mopsi
president-Nom cut-Mod. neck because-of the man-Top very
sulphessta.
sad
‘Literal only: Because of the neck that the president cut, the man felt so

sad.'

- c. Mary-ka [[ku sihem-eyse mek-un] miyekkwuk] ttaymwuney,
M-Nom the exam-in eat-Mod. seaweed-soup because-of
kacok motwu-ka sulphessta.
family all-Nom sad

'Literal only: Because of the seaweed soup that Mary ate during the exam, all her family were sad.'

Again the absence of relativizability of an idiomatic nominal seems to arise from its non-referential property and accommodation problem. As is well known, there is a close relationship between topicalization and relativization: the relativized nominal (the so-called antecedent) has a property of a topic while the relative clause has the property of the comment clause. It is thus natural to expect that we cannot relativize the expression that cannot be topicalized.

Cleft: It seems that the nominal component of the decomposable idioms cannot occur at the focus position of the cleft construction.

- (30) a. Mary-ka kkekkwulo sin-un kes-un komwusin-iessta.
M-Nom backwards wear-Mod. thing-Top rubber shoes-were
'Literal only: It was rubber shoes that Mary wore backwards.'

- b. Mary-ka namphyen-eykey kulk-un kes-un pakaci-yessta.
M-Nom husband-to scrape-Mod. thing-Top gourd dipper-was
'Literal only: It was a gourd dipper that Mary scraped to her husband.'

- c. Mary-ka ku simhem-eyse mek-un kes-un miyekkwu-iessta.
M-Nom the exam-in eat-Mod. thing-Top seaweed soup-was
'Literal only: It was seaweed soup that Mary ate during the exam.'

In these cleft constructions, two main components are the cleft clause headed by

kes and the focus phrase with the copula verb. Even though the nominal component of the decomposable idiom has a certain contribution to the whole idiomatic meaning, we cannot cleft the nominal component. This impossibility appears to be related to the semantic property of the nominal and the head of the cleft clause *kes*. The head *kes* in the cleft clause which is semantically linked to the missing element in the cleft clause canonically refers to a concrete object (see Kim and Sells 2007). This implies that the cleft clause would lead us only to a literal meaning. That is, when we process the cleft construction, *Mary-ka kekkwulo sin-un kes-un* in (30a), it is hard to obtain an idiomatic reading.

Afterthought: Even though Korean is one of the typical head final languages, arguments can occur after the matrix head verb in a certain context. This is often called the ‘afterthought’ construction because the rightward-extracted argument occurs after a sentence break intonation. No nominal component of the decomposable idioms can appear as an afterthought expression:

- (31) a. *Mary-ka kekkwulo sinesta, komwusin-ul.*
 M-Nom backwards wore rubber shoes-Acc
 ‘Literal only: Mary wore rubber shoes backwards.’
- b. *Mary-ka namphyen-hanthey kulkessta, pakaci-ul*
 M-Nom husband-to scraped gourd dipper-Acc
 ‘Literal only: Mary scraped a gourd dipper to her husband.’
- c. *Mary-ka ku sihem-eyse mekessta, miyekkwuk-ul.*
 M-Nom the exam-in ate seaweed soup-Acc
 ‘Literal only: Mary ate seaweed soup during the exam.’

Here again the co-occurrence restriction seems to be responsible for the absence of clefting. In the afterthought construction, the afterthought component is intervened by a sentence boundary, which is realized by an intonation break. So even though

the verbal and the nominal components look like occurring adjacent, they are actually intervened by a finite clause boundary. Speakers do not expect an idiom component follows the sentence intonation break.

To sum up, a component of the decomposable idiom can go through various syntactic processes. Even though many decomposable idioms do not allow some of the syntactic processes, the absence of these appears to be related to extra-grammatical factors such as language processing (cf. O'Grady 1998, Riehemann 2001, McGinnis 2002). One can attribute the absence of some syntactic processes to other grammatical factors. For example, we may assume those idioms should be analyzed like a syntactically fixed phrasal word by mapping the whole phonological values to the complete idiomatic meaning. However, this analysis has at least two burdens. First, even such less flexible idioms all can go through scrambling, external and internal modification, and optional case marking. We then need to account for how those syntactic processes still occur even in the syntactically less flexible idioms. Second, the categorical distinction between the questionable fixed idioms and the more flexible idioms may fall in a vicious circular logic, i.e., the former are fixed because they are syntactically less flexible, and they are less flexible because they are fixed idioms. (See Sag et al. 2002, Sailer 2000, Soehn 2006, Schenk 1995, Totone and Connine 1999). To avoid these burdens, we consider the less flexible idioms also belong to decomposable idioms. Even in the less flexible idioms, each component carries a meaning that contributes to the whole idiom meaning. For example, in the idiom like *komwusin-ul kkekkwulo sinta* 'betray one's lover', *komwusin* 'rubber shoes' means something like 'love' while the predicate *kkekkwulo sinta* 'wear backwards' means something like 'betray'.

An HPSG Analysis

Non-decomposable Idioms

The treatment of non-decomposable idioms, whose syntactic processes are quite rigid, is rather simple. For example, for the expression *kkwueta nohun poli-calwu* to have the idiomatic meaning ‘dumb like a wallflower’, the entire idiom is encoded in the lexicon as a single word:

$$(32) \left[\begin{array}{l} \text{FORM } \langle \text{kkwueta nohun poli-calwu} \rangle \\ \text{SYN} \mid \text{POS } \textit{noun} \\ \text{SEM} \left[\begin{array}{l} \text{IND } i \\ \text{RELS} \left\langle \left[\begin{array}{l} \text{PRED } \textit{dumb-like-a-wallflower_rel} \\ \text{INSTANCE } i \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

The lexical entry tells us that the string *kkwueta nohun poli-calwu* is a noun (POS) and semantically refers to an individual ‘*i*’ with the meaning of ‘dumb-like-a-wallflower’.

The idiom *nwun-ul kam-* also needs to be recorded as a word:⁴⁾

$$(33) \left[\begin{array}{l} \text{FORM } \langle \textit{nwun-ul kam-ta} \rangle \\ \text{SYN} \left[\begin{array}{l} \text{POS } \textit{verb} \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ } \langle \textit{NP}_i \rangle \\ \text{COMPS } \langle \rangle \end{array} \right] \end{array} \right] \\ \text{SEM} \left[\begin{array}{l} \text{IND } \textit{s1} \\ \text{RELS} \left\langle \left[\begin{array}{l} \text{PRED } \textit{die_rel} \\ \text{INSTANCE } i \end{array} \right] \right\rangle \end{array} \right] \end{array} \right]$$

4) To be more precise, the lexicon also encodes this idiom without the accusative case as an independent complex word but it cannot have a delimiter marker at all.

This idiom is a single verb selecting one subject (SUBJ) and no complement (COMPS) while referring to a situation 's1' whose meaning is 'die'. Notice that since the string *nwun-ul kam* is recorded as a verb, we allow it to be inflected with an honorific or with an inflectional ending:

- (34) a. *sensayng-nim-i ecey nwun-ul kam-usi-ess.*
 teacher-Hon.-Nom yesterday eyes-Acc close-Hon-Decl
 'Idiomatic: The teacher died yesterday.'
 'Literal: The teacher closed his eyes yesterday.'
- b. *ecey nwu-ka nwun-ul kam-ass-ta-ko?*
 yesterday who-Nom eyes-ACC close-Past-Decl-Quot
 'Who did you say was dead yesterday?'

Since these non-decomposable expressions are recorded as one lexical unit, we expect no syntactic operations to be interpolate into such a word-like idiom.

- (35) a. *nwun-un John-i ecey kam-ss-ta*
 eyes-Top John-Nom yesterday close-Past-Decl
 'Literal only: As for the eyes, John closed them yesterday.'
- b. *nwun-i ecey John-eyuyhay kam-ki-e ci-ss-ta*
 eyes-Nom yesterday John-y close-Pass-Comp become-Past-Decl
 'Literal only: The eyes were closed by John.'

These idioms even cannot have the (inflectional) short form negation though they can have the (syntactic) long form negation:

- (36) a. *John-i acik nwun-ul an kam-ass-ta.*
 John-Nom still eyes not close-Past-Decl
 'Literal only: John still didn't close the eyes.'

- b. John-i acik nwun-ul kam-ci ahn-ass-ta.
 John-Nom still eyes close-Comp Not-Past-Decl.
 ‘Idiomatic: John isn't dead yet.’

This contrast between the short form and long negation again tells us that the idiomatic expression *nwun-ul kam-* is listed as a lexical item whose internal structure no syntactic processes can refer to.

Decomposable Idioms

As we have seen so far, unlike non-decomposable idioms, decomposable idioms are more flexible in terms of occurring in syntactic configurations:

- (37) a. Mary-ka John-hanthe kkoli-lul chi-ess-ta
 Mary-Nom John-to tail-Acc wag-Past-Decl
 ‘Idiomatic: Mary seduced John.’
- b. Mary-ka John-hanthe kkoli-lul mence chi-ess-ta
 Mary-Nom John-to tail-Acc first wag-Past-Decl
 ‘Idiomatic: Mary seduced John first.’
- c. kkoli-nun Mary-ka John-hanthe mence chi-ess-ta
 tail-Top Mary-Nom John-to first wag-Past-Decl
 ‘Idiomatic: It was Mary who seduced John first.’
- d. Mary-ka John-hanthe kkoli-lul an chi-ess-ta
 Mary-Nom John-to tail-Acc not wag-Past-Decl
 ‘Idiomatic: Mary didn't seduce John.’

However, for such idioms to get an idiomatic interpretation, copredication is a necessary condition (cf. Sag 2007). That is, for an idiom like *kkoli-lul chi-* to get an idiomatic interpretation, the argument of *chi-* must have *kkoli* predicated of it. For

this, the grammar we assume here includes the following lexical entry for *miyekkwuk*:

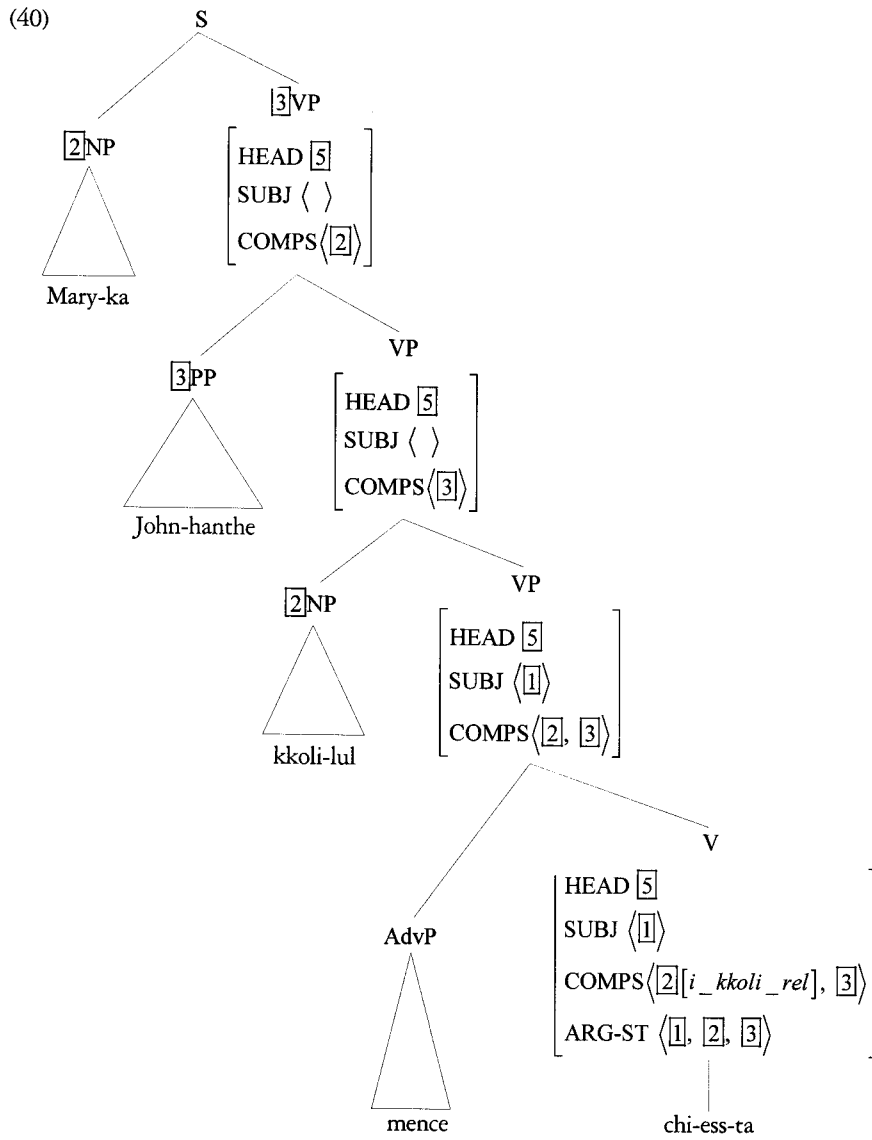
$$(38) \left[\begin{array}{l} \text{FORM} \langle \text{kkoli} \rangle \\ \text{SYN} \left[\begin{array}{l} \text{POS } \textit{noun} \\ \text{LID } \textit{i_kkoli_rel} \end{array} \right] \\ \text{SEM} | \text{RELS} \langle \quad \rangle \end{array} \right]$$

The feature LID (Lexical-Identifier), adopted from Sag (2007), is to individuate lexical items semantically. Assuming that words can have literal and idiomatic relations,⁵⁾ the noun *kkoli* will default to its literal interpretation (*l_kkoli_rel*) except when its LID value is resolved to the idiomatic relation (*i_kkoli_rel*) by the idiomatic verb *chi-ta*, whose lexical entry is given in the following:

$$(39) \left[\begin{array}{l} \text{FORM} \langle \text{chi-ta} \rangle \\ \text{SYN} \left[\begin{array}{l} \text{HEAD} | \text{POS } \textit{verb} \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ} \langle \text{NP}_i \rangle \\ \text{COMPS} \langle \text{NP}[\textit{i_kkoli_rel}], \text{NP}_j[\textit{dat}] \rangle \end{array} \right] \end{array} \right] \\ \text{SEM} | \text{RELS} \left\langle \begin{array}{l} \text{PRED } \textit{seduce_rel} \\ \text{ARG1 } \textit{i} \\ \text{ARG2 } \textit{j} \end{array} \right\rangle \end{array} \right]$$

The lexical entry means the idiomatic verb *chi-ta* selects two complements: one NP with the LID value *i_kkoli_rel* and a dative NP. This selection will then assign the following structure for (37b):

5) For example, the expression *kkoli* will have both *l_kkoli_rel* (literal meaning) and *i_kkoli_rel* (idiomatic meaning) as its subtypes. See Sag (2007).



The idiom verb *chi-ess-ta* will select three arguments, the first of which is realized as the SUBJ whereas the remaining two are realized as the COMPS.⁶⁾ In the tree structure, the verb first combines with the modifier *mence*, forming a head-modifier

structure. The resulting phrase will then combine with the complement *kkoli-lul* with the LID value, and the resulting expression then combines with the dative argument. This strictly satisfies the subcategorization requirement of the verb. The present analysis, treating each argument of the idiom as an independent syntactic unit, can resolve issues related to the strong locality of idioms.

This kind of analysis can also offer us a way of dealing with idioms that obligatory select an adverbial element, whose data we repeated here from (19):

- (41) Mary-ka komwusin-ul kekkwulo sinesta
 M-Nom rubber shoes-Acc backwards wore
 ‘Mary betrayed (her) lover.’

The presence of the adverb *kekkwulo* is obligatory to trigger the intended idiomatic reading. One way to resolve this problem is to take the specified adverb as a complement as given in the following lexical entry:⁷⁾

- (42)
$$\left[\begin{array}{l} \text{FORM} \langle \text{sin-} \rangle \\ \text{HEAD} | \text{POS } \textit{verb} \\ \text{VAL} \left[\begin{array}{l} \text{SUBJ} \langle \text{NP}_i \rangle \\ \text{COMPS} \langle \text{AdvP} [i_kekkwulo_rel], \text{NP} [i_komwusin_rel] \rangle \end{array} \right] \end{array} \right]$$

The lexical entry indicates that the idiomatic verb *sin-* selects two complements: *kekkwulo* and *komwusin*, each of which has its own idiomatic relation LID value. Since the analysis treats each component of such idioms as an independent syntactic element, we can expect it can be scrambled or contrastive-marked:

6) For the syntax of Korean, see (Kim 2004).

7) Verbs like *chesinba-* ‘behave’ or *kwul-* ‘behave’ also select an obligatory adverbial element.

- (43) a. Mary-ka komwusin-ul swusilo kekkwulo sinessta
M-Nom rubber shoes-Acc anytime backwards wore
'Mary always betrayed (her) lover.'
- b. pilok Mary-ka komwusin-un kekkwulo siness-ciman,...
though Mary-Nom rubber shoes-Cont.Top backwards wore-though,...
'Though Mary betrayed (her) lover, (she still loves him).'

In the previous section, we even observed that some of the decomposable idioms allow their component to be relativized:

- (44) Mary-ka kelchi-n yangtali ttaymwuney...
Mary-Nom lay-across-Mod. both-legs because-of
'Because of Mary's opportunism,...'

Even though the idiom verb and its complement in (44) are not in the same local domain, as the reader can verify, the present analysis correctly allows the proper composition of the idiomatic meaning. This effect is mainly achieved by assigning an idiomatic meaning to the verb *kelchi-* when it takes an NP complement whose LID value is *i_yantali_rel* and further allowing this lexical information to closely interact with other grammatical components (see Sag (2007)).

Conclusion

This paper has identified two types of Korean idioms: decomposable and non-decomposable idioms and then review their morpho-syntactic properties. It has then sketched an HPSG-style analysis reflecting how to process them efficiently.

For the syntactically fixed non-decomposable idioms, the idiomatic expression is listed in the lexicon as a lexical element which cannot be interpolated by any

syntactic operations. As for more flexible, decomposable idioms, restrictions are imposed only on the semantics of the idiomatic words. This loose syntactic restriction allows the idioms to undergo various syntactic processes such as scrambling, case alternation, topicalization, and relativization. The absence of such syntactic processes in certain decomposable idioms is suggested to concern the semantic properties of the idiomatic nominal component as well as extragrammatical factors such as language processing.

References

- Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.
- Fillmore, Charles, Paul Kay, and Mary O'Conner. 1988. Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language* 64, 3. 501-538.
- Jackendoff, Ray. 1997. Twistin' the Night Away. *Language* 67: 320-338.
- Kim, Jong-Bok and Peter Sells. 2007. Korean cleft constructions and information structure. Paper presented in the Workshop on Japanese/Korean semantics and pragmatics as JK 17 Pre-conference Workshop, UCLA.
- McCawley, James. 1983. What's with *with*? *Language* 59, 2. 271-287.
- Nunberg, Geoffrey, Ivan Sag, and Thomas Wasow. 1994. Idioms. *Language* 70, 3. 491-538.
- Katz, Jerrold. 1973. Compositionality, idiomaticity and lexical substitution. *A Festschrift for Morris Halle*, ed. by Stephen R. Anderson and Paul Kiparsky. 357-376.
- Kay, Paul, and Charles Fillmore. 1999. Grammatical constructions and linguistic generalizations: The What's X doing Y? construction. *Language* 75, 1. 1-33.
- Kim, Jong-Bok. 2004. *Korean Phrase Structure Grammar* (In Korean). Hankwuk Publishing.
- McGinnis, Martha. 2002. On the systematic aspects of idioms. *Linguistic Inquiry* 33:4,

665-672.

- O'Grady, William. 1998. The syntax of idioms. *Natural Language and Linguistic Theory* 16: 279-312.
- Park, Man-Ghyu. 2003. To establish the identity of the category of idioms: with emphasis on semantic analysis. *Korean Linguistics*, 41: 308-353.
- Riehemann, Susanne. 2001. *A Constructional Approach to Idioms and Word Formation*. Ph.D. Dissertation. Stanford University.
- Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger. 2002. Multiword expression: a pain in the neck for NLP. *CICLING 2002*. Mexico City, Mexico. 1-15.
- Sailer, Manfred. 2000. *Combinatorial Semantics and Idiomatic expressions in HPSG*. Ph.D. Dissertation, Eberhard-Karls-Universitt Tübingen.
- Sag, Ivan. 2007. Feature Geometry and Predictions of Locality. Ms. Stanford University.
- Soehn, J.-P. 2006. On Idiom Parts and their contexts. *Linguistik online* 27.2
- Schenk, Andre. 1995. The syntactic behavior of idioms. *Idioms*, ed. by Everaert, Martin et al. Lawrence Erlbaum Associates, Publishers: Hillsdale, New Jersey. 253-272.
- Titone, Debra and Cynthia Connine. 1999. On the compositional and noncompositional nature of idiomatic expression. *Journal of Pragmatics* 31: 1655-1674.
- van der Linden, Erik-Jan. 1992. Incremental processing and the hierarchical lexicon. *Computational Linguistics* 18. 219-238.

1 차원고접수 : 2007. 11. 19

최종게재승인 : 2007. 12. 17

요 약

한국어 속어의 유형과 처리 방법: 제약기반이론적 접근

정 찬
동서대학교

김 종 복
경희대학교

일반 속어들의 구성요소들은 통사적 현상에 적용을 받지 않을 뿐만 아니라, 의미적 합성성원리를 지키지 않는다. 그러나 이러한 속어들과는 달리 속어의 구성소들이 형태, 통사적 현상에 의해 영향을 받거나 의미적 합성 과정을 거치는 속어들도 존재한다. 본 논문은 한국어에서 이러한 두 가지 유형의 속어들을 파악하고, 이들의 속성을 살펴본다. 이러한 경험적 연구를 바탕으로, 이들 구문들을 처리할 수 있는 제약기반이론적 분석을 제공한다.

주제어 : 한국어 속어, 분리가능속어, 비분리 속어, 제약기반, 핵어중심구조문법