

# Myths in Korean Morphology and Their Computational Implications

Hee-Rahk Chae

Department of Linguistics and Cognitive Science

Hankuk University of Foreign Studies

Yongin, Gyeonggi, 449-791, Korea

hrchae@hufs.ac.kr

## Abstract

This paper examines some popular misanalyses in Korean morphology. For example, contrary to popular myth, the verbal *ha-* and the element *-(nu)n-* cannot be analyzed as a derivational affix and as a present tense marker, respectively. We will see that *ha-* is an independent word and that *-(nu)n-* is part of a portmanteau morph. In providing reasonable analyses of them, we will consider some computational implications of the misanalyses. It is really mysterious that such wrong analyses can become so popular in a scientific field of linguistics.

## 1 Introduction

This paper aims at examining some popular misanalyses in Korean morphology. Focusing on the verbal *ha-* and what is called the present tense marker *-(nu)n-*, we will see that, contrary to popular myth, they cannot be analyzed as a derivational affix and as a present tense marker, respectively. In providing reasonable analyses of them, we will consider some implications of the misanalyses, especially from a computational point of view.

Most Korean linguists assume that the *ha-* in *kongpwu-ha-* ('to study'), for example, is a derivational affix and, hence, *kongpwu-ha-* as a

whole is a verb.<sup>1</sup> However, as we can see shortly, *ha-* itself is an independent word and [*kongpwu ha-*] is a phrase. More Korean linguists assume that the element *-(nu)n-* is a present tense marker. However, the Korean tense system becomes far simpler, if we assume that the present tense marker is null (*-∅-*) rather than *-(nu)n-*.

## 2 The Morpho-syntactic Status of Some Dependent Elements

As an agglutinative language, Korean has rather complex structures of word-like expressions. Hence, it is not always easy to determine the

<sup>1</sup> Noticeable exceptions are Song (1967: 64-71), Suh (1991: 486, 1994: 578, 1996: 346), Chae (1996) and some others. They have shown, for example, that *ha-* in *kongpwu-ha-* cannot be a derivational affix and that [*kongpwu ha-*] and [*kongpwu-lul ha-*] are realizations of the same syntactic structure.

The Japanese counterpart of the Korean *ha-* is *suru*. The unit of verbal noun plus *suru* is also regarded as a word by most Japanese linguists. However, this is very dubious.

- a) *bengkyou-bakari/wa/... suru*  
study -only/Contr do
- b) <sup>??</sup>*bengkyou yoku/nagaku/... suru*  
well/long time
- c) [*bengkyou-to undou*]-*bakari/wa/... suru*  
-and exercise

Although it is not very natural for such independent words as *yoku* and *nagaku* to come between the two elements, as we can see in (b), delimiters like *-bakari* and *-wa* are allowed as in (a). In addition, the verbal noun before *suru* can be conjoined, as we can see in (c). These facts show that *bengkyou-suru* is not a word but a phrase (and, hence, such expressions should not be registered as head words in dictionaries).



before the predicate. In addition, such adverbs as *manhi* ‘many/much’ can be inserted between the noun and the predicate. These facts clearly indicate that the expressions in (a) are all phrases rather than compound words. Secondly, such verbal elements as *ha-*, *toy-* and *sikhi-* are assumed to be derivational affixes not only in most Korean grammar books and dictionaries but also in most research papers (cf. footnote 1).

- (4) a. phakoy-ha-ta  
destruction-do-Decl ‘to destroy’  
b. phakoy-toy-ta  
-become ‘to be destroyed’  
c. phakoy-sikhi-ta  
-let ... do ‘to (let ...) destroy’
- (5) a. phakoy-lul ha-ta  
-Acc  
b. phakoy-ka toy-ta  
-Nom  
c. phakoy-lul sikhi-ta  
-Acc

As they are analyzed as derivational affixes, all the expressions in (4) are regarded as verbs rather than verb phrases. However, they cannot be verbs as we can see from the data in (5), which show that accusative or nominative markers, which can only come at the end of object/subject phrases, can be inserted in between.

Among the numerous examples of misanalyses (cf. Chae 2010), the type in (4) is the least expected one, because a regular word is analyzed as a derivational affix. Regular words are completely independent from the preceding root/word and, hence, they do not belong to the dependent elements listed in (2). They are more independent units than clitics. Derivational affixes are the least independent from its root. There is another unexpected type of misanalysis. In this type part of a word which cannot be a separate morpheme is analyzed as one. Although there are not many examples of this type, it is also unusual in the sense that morphemes are not difficult to factor out, especially in an agglutinative language. In the remaining sections of this paper, we will focus on only one example from each of these two types of misanalyses: the “light verb” *ha-* and the assumed present tense marker *-(nu)n-*. We will not only elucidate their morpho-syntactic statuses but also consider computational implications of the misanalysis.

### 3 The Verbal *ha-*

In this section, we will firstly examine the morpho-syntactic status of the verbal *ha-*. Then, we will consider what kinds of implications the popular misanalysis has for automatic analyses.

#### 3.1 The Morpho-syntactic Status

The agglutinative nature of Korean makes it difficult to distinguish between word-internal elements like (derivational and inflectional) affixes and word-external elements like clitics. What makes the belief that the verbal *ha-* is a derivational affix be mysterious is that it is not even a clitic but a wholly independent word. Let us examine the following examples:

- (6) cyon-i kongpwu-ha-ko  
John-Nom study do-Progr  
iss-ta  
be-(Pres)-Decl  
‘John is studying.’
- (7) cyon-i kongpwu(-lul) cal/manhi/...  
-Acc well/much/...  
ha-ko iss-ta  
‘John is studying well/much/...’

Judging from the data in (7), which show that external elements can be inserted between *kongpwu* and *ha-*, it becomes clear that *ha-* is a word and [*kongpwu ha-*] is a phrase. That is, *kongpwu* and *ha-* are two independent words (Song 1967, Suh 1991, Chae 1996, Chae and Chong 2011, among others). Firstly, the accusative marker *-(l)ul* can be inserted between them. Secondly, such adverbs as *cal* and *manhi* can also be inserted between them freely. We do not need any more evidence to establish the morpho-syntactic status of *ha-* as an independent word.

Those who take the wordhood of *kongpwu-ha-* for granted argue that such expressions as [*kongpwu cal ha-*] are derived from the phrase [*kongpwu-lul ha-*], deleting the accusative marker *-lul* and adding the adverb *cal*. Under this kind of argumentation, it is assumed that [*kongpwu cal ha-*] has nothing to do with the “word” *kongpwu-ha-*. However, there are serious problems with such an approach. First of all, it is not understandable at all that *kongpwu-ha-* does not have any (formal) relationship with [*kongpwu-lul ha-*] or [*kongpwu cal ha-*]. These latter expressions have no special meanings different from that of the former expression,

except that they contain *-lul* and *cal*, respectively. Secondly, the argument is not falsifiable, which leads to a non-scientific research. It is not falsifiable because all units of [NP V] can be argued to be words rather than phrases:

- (8) a. pap-ul (cal) mek-ta  
rice-Acc well eat-Decl  
b. pap (cal) mek-ta  
'to eat boiled rice (well)'

- (9) a. hakkyo-ey (cacu) ka-ta  
school-to often go-Decl  
b. hakkyo (cacu) ka-ta  
'to go to school (often)'

If *kongpwu-ha-* is argued to be a word despite such expressions as [*kongpwu-lul ha-*] and [*kongpwu cal ha-*],<sup>4</sup> it can also be argued that [*pap mek-*] in (8b) and [*hakkyo ka-*] in (9b) are words rather than phrases. Under this kind of argumentation, we can say that [*pap cal mek-*] and [*hakkyo cacu ka-*] are derived from [*pap-ul mek-*] in (8a) and [*hakkyo-ey ka-*] in (9a), respectively, rather than from the "words" [*pap mek-*] and [*hakkyo ka-*]. However, even those who assume that *kongpwu-ha-* is a word will not accept that [*pap mek-*] and [*hakkyo ka-*] are words.

### 3.2 Computational Implications

If we cannot factor out a regular word *ha-* from expressions like *kongpwu-ha-*, we cannot provide a systematic analysis of the expressions containing it. In that case, *kongpwu-ha-* and [*kongpwu cal ha-*], for example, can only be analyzed with reference to two unrelated mechanisms. The former should be listed in the dictionary because it is assumed to be a word. The latter, on the other hand, should be treated in the syntactic component on the basis of the three lexical items *kongpwu*, *cal*, and *ha-* and relevant syntactic rules and/or principles.

The situation becomes more serious in automatic analyses than in manual analyses. First of all, it is impossible to capture any formal relationships between *kongpwu-ha-* and [*kongpwu cal ha-*], because they are outputs of two different components and they do not even share any lexical items. However, it is clear that

<sup>4</sup> One might argue that the verbal *ha-* cannot be regarded as an independent word because it does not have its own meaning. However, semantic facts do not necessarily go together with morpho-syntactic facts. That is, the meaning of a unit cannot tell whether it is a word or not.

the only difference between them is due to the (non-)existence of the adverb *cal*, which is impossible to capture under the popular approach. Secondly, it is very difficult, though may not be impossible, to capture the semantic relationship between the two expressions. Thirdly, all the lexical entries involved have to be registered twice, leading to a significant amount of redundancy (Chae 2010). Although *kongpwu-ha-* is registered in the dictionary, *kongpwu* and *ha-* have to be registered as well. Notice that these words appear in the phrase [*kongpwu cal ha-*], in which the adverb *cal* is in between the two words. Lastly, the system will produce two different analyses of *kongpwu-ha-*: as a lexical item and as a syntactic construct. As we have *kongpwu* and *ha-* as separate lexical items, there is no reasonable way of preventing the combination of them to produce [*kongpwu ha-*], which is the same as the lexical item *kongpwu-ha-*.

We have seen problems with only one example. From a computational point of view, the sheer number of *ha-* expressions in Korean makes the popular misanalysis more difficult to maintain. It may be the case that expressions containing *ha-* would be more than half of the whole verbal expressions in representative Korean corpora.

## 4 The Verbal Element *-(nu)n-*

In this section, we will examine the behavior of the verbal element *-(nu)n-*. Although it is usually assumed to be a present tense marker, the assumption is based on superficial observations. A more careful observation will lead to the conclusion that the present tense marker, more accurately, the non-past tense marker is null (*-ø-*) rather than *-(nu)n-*. Of course, there are some previous works which argue for this position like Kang (1988), Suh (1994) and others. However, the argument has not been taken seriously in Korean linguistics, just like that for the wordhood of *ha-* in *kongpwu-ha-* (cf. footnote 1).

### 4.1 The Morpho-syntactic Status

The popular belief that *-(nu)n-* is a present tense marker is based on such data as the following.<sup>5</sup>

<sup>5</sup> The verbal marker *-(nu)n* has two variants: *-nun* after a verb ending in a consonant and *-n* after a verb ending in a vowel.

- (10) a. cyon-i cip-ey ka-n-ta  
John-Nom house-to go-Pres-Decl
- b. cyon-i cip-ey ka-ass-ta  
-Past  
'John goes/went home.'
- (11) a. cyon-i pap-ul mek-nun-ta  
John-Nom rice-Acc eat-Pres-Decl
- b. cyon-i pap-ul mek-ess-ta  
-Past  
'John eats/ate boiled rice.'

When we compare the two sentences in (10) and in (11), it seems to be very obvious that *-(nu)n-* is in a paradigmatic relation with the past tense marker *-ass/ess*.

However, if we observe the behavior of the element *-(nu)n-* more carefully, we will see that there are many problems with the popular belief. First of all, *-(nu)n-* is not actually in a paradigmatic relation with the past tense marker.

- (12) a. ka(\*-n)-keyss-ta  
Go -Modality-Decl  
ka(-ass)-keyss-ta
- b. mek(\*-nun)-keyss-ta  
eat  
mek(-ess)-keyss-ta

The past tense marker can occur before the irrealis modality marker *-keyss*, but the assumed present tense marker cannot.

Secondly, the distribution of *-(nu)n-* is very limited:

- (13) a. ka(\*-n)-(nu)nya, mek(\*-nun)-(nu)nya  
-Interrogative
- b. ka(\*-n)-kela, mek(\*-nun)-ela  
-Directive
- c. ka(\*-n)-ca, mek(\*-nun)-ca  
-Propositive
- (14) a. ka(\*-n/<sup>ok</sup>-ass)-a/e,  
mek(\*-nun/<sup>ok</sup>-ess)-e
- b. ka(\*-n/<sup>ok</sup>-ass)-ney,  
mek(\*-nun/<sup>ok</sup>-ess)-ney
- c. ka(\*-n/<sup>ok</sup>-ass)-o,  
mek(\*-nun/<sup>ok</sup>-ess)-uo/o
- d. ka(\*-n/<sup>ok</sup>-ass)-a/e-yo,  
mek(\*-nun/<sup>ok</sup>-ess)-e-yo
- e. ka(\*-n/<sup>ok</sup>-ass)-pnita/supnita,  
mek(\*-nun/<sup>ok</sup>-ess)-supnita

Korean verbal endings have different forms according to sentence type and speech level. There are at least four different sentence types: declaratives, interrogatives, directives and propositives. There are six different speech levels, from the least formal to the most formal. Among the twenty four possible combinations of the two grammatical categories, only one combination requires the element *-n-* or *-nun-*: that of the declarative sentence<sup>6</sup> and the (least formal) plain level sentence, as we can see in (10a) and (11a). The element does not appear in the other combinations. As we can see in (13), it cannot combine with the interrogative, directive or propositive ending, even when the speech level concerned is the plain level. In addition, as we can see in (14), it cannot combine with any of the other speech level endings.

We can easily solve these problems if we assume that the non-past tense marker is *-∅-*. Under this assumption, the variants of *-(nu)n-*, i.e. *-n-* and *-nun-*, are just parts of the (present) declarative endings of verbs in the plain speech level. That is, we can assume that *-nta* and *-nunnta* are “portmanteau” morphs,<sup>7</sup> i.e. those morphs which can be analyzed into more than one morpheme (Crystal 1980, Spencer 1991).<sup>8</sup> The

<sup>6</sup> What seems to be “exclamative endings,” among others, also contain *-nun-*.

- a) cip-ey ka-nunkwuna/nunkwun.  
house-to go-Ending  
'(He/She) does go home!'
- b) cal mek-nunkwuna/nunkwun.  
well eat-Ending  
'How well (he/she) eats!'
- c) san-i khu/cak-kwuna/kwun.  
mountain-Nom be big/small-Ending  
'How big/small the mountain is!'

Compared with the endings after adjectives (or descriptive verbs) in (c), those after verbs have the extra element *-nun-* in (a-b). However, there is enough evidence to show that Korean does not have a separate sentence type of exclamative. What seems to be exclamative sentences have the formal properties of declarative sentences. Hence, the sentences above belong to declaratives in Korean (Lee 2005: 170-171).

<sup>7</sup> We are in line with Yongkyoon No's assumption in “... the selection from allomorphs *-nunnta/nta/ta ...*” (Chae and No 1998: 91). He regards *-nunnta*, *-nta* and *-ta* as allomorphs of one and same morpheme.

<sup>8</sup> Portmanteau morphs are defined/described in the literature as follows: “A term used in morphological analysis referring to cases where a single morph can be analysed into more than one morpheme, ...” (Crystal 1980: 276); “... the term portmanteau, which in this context means type of fusion of two morphemes into one. ... we have four morphemes all realized by a single portmanteau morph ... In a portmanteau morph, then,

two portmanteau morphs indicate the present tense of the plain level declarative sentence. The former is used when the stem of the verb concerned ends in a vowel, and the latter when it ends in a consonant. The point here is that they are indivisible morphs which contain not only the information about the sentence type and the sentence level but also the information about the tense of the verb concerned.

Under the  $-\emptyset$ -tense marker approach,  $-(nu)n-$  is inseparable from the predicative ending  $-ta$  and, hence, cannot take the position of tense markers. In addition, the non-past and the past tense markers take the same position:

- (15) a.  $ka-\emptyset-nta$ ,  $mek-\emptyset-nunta$  (cf. (10-11))  
 b.  $ka-\emptyset-keyss-ta$ ,  $mek-\emptyset-keyss-ta$   
 (cf. (12))  
 c.  $ka-\emptyset-(nu)nya$ ,  $mek-\emptyset-(nu)nya$   
 (cf. (13a))  
 d.  $ka-\emptyset/ass-e$ ,  $mek-\emptyset/ess-e$  (cf. (14a))

As we can see from this reanalysis of the data in (10-14), we can account for the ungrammatical data in (12-14) very naturally. In (12), the inseparable  $-(nu)n-$  and  $-ta$  are separated from each other. In (13) and (14),  $-(nu)n-$  stands alone without its inseparable “partner”  $-ta$ .

Before leaving this section, we need to introduce a constraint, with reference to the following data:

- (16) a. \* $ka-ass-nunta$ , \* $mek-ess-nunta$   
 b. \* $ka-\emptyset-keyss-nunta$ ,  
 \* $mek-\emptyset-keyss-nunta$

In (a), although the past tense marker takes the same position as that of the non-past tense marker (cf. (15a)), the expressions concerned are ungrammatical. They are ungrammatical just because the portmanteau morph  $-nunta$  occurs with the past tense marker. In (b), although the morph  $-nunta$  occurs with the non-past marker, the expressions are ungrammatical as well. We need to postulate that the morph has to be immediately preceded by the non-past tense marker. Notice that this constraint accounts for both types of data in (16).

---

several categories are realized by one surface formative, an instance of a one-many correspondence between form and function” (Spencer 1991: 50-51).

## 4.2 Computational Implications

As we have seen with reference to the data in (13) and (14), among dozens of possible combinations of speech level and sentence type, only one combination of the plain level and the declarative sentence requires  $-n-$  or  $-nun-$ . All the other combinations cannot have the element. Then, it would be very difficult to account for the distribution of  $-(nu)n-$  computationally, if we assume that it is a present tense marker. Please notice that, as is shown in (14), the past tense marker  $-ass/ess$  can occur in the position where the element  $-(nu)n-$  is not allowed to occur.

When we deal with computational systems, we have to consider the understanding process and the productions process separately, just as the two areas of speech recognition and speech synthesis show. From an understanding point of view, the traditional approach fails to interpret many present tense forms. For example,  $ka-a$  and  $mek-e$  are correct present tense forms, although they do not have  $-(nu)n-$  (cf. (14)). From a production point of view, the approach produces a lot of ill-formed expressions: including all the ill-formed ones in (12-14). It would not be easy to filter out these expressions.

## 5 Conclusion

In this paper, we have surveyed some popular misanalyses in Korean morphology, focusing on two unexpected ones: the verbal  $ha-$  as a derivational affix and the verbal element  $-(nu)n-$  as a present tense marker. We have shown that careful observations reveal that  $ha-$  is an independent verb and that  $-nun-$  and  $-n-$  are parts of portmanteau morphs rather than independent morphemes themselves. It is really mysterious that such wrong analyses can become so popular in a scientific field of linguistics.

## Acknowledgments

We are thankful to the anonymous reviewers, whose valuable comments have been very helpful in improving the quality of this paper. This work was supported by a 2013 research grant from Hankuk University of Foreign Studies.

## References

- Chae, Hee-Rahk. 1996. Properties of  $ha-$  and Light Predicate Constructions [written in Korean]. *Language Research*, 32(3), 409-476.

- Chae, Hee-Rahk. 2007. Clitics and a Classification of Parts of Speech in Korean [written in Korean]. *Korean Journal of Linguistics*, 32(4), 803-826.
- Chae, Hee-Rahk. 2010. Basic Units of Lexicons and Ontologies: Words, Senses and Concepts. *Proceedings of the 24th Pacific Asia Conference on Language, Information and Computation*, 35-44, Tohoku University.
- Chae, Hee-Rahk and Wuk-Jae Chong. 2011. A Procedure for the Identification of Word Units in Korean. *Harvard Studies in Korean Linguistics XIV*, 67-76, Harvard-Yenching Institute.
- Chae, Hee-Rahk and Yongkyoon No. 1998. A Survey of Morphological Issues in Korean: Focusing on Syntactically Relevant Phenomena. *Korean Linguistics* 9, International Circle of Korean Linguistics.
- Crystal, David. 1980. *A First Dictionary of Linguistics and Phonetics*. Blackwell.
- Kang, Beom-mo. 1988. *Functional Inheritance, Anaphora, and Semantic Interpretation*. Ph.D. dissertation, Brown University.
- Lee, Iksop. 2005. *A Korean Grammar* [written in Korean]. Seoul National University Press.
- Roh, Chang-Hwa. 2013. *A Study of Verb Sequences in Korean: Focusing on [V-e V] Expressions* [written in Korean]. Ph.D. dissertation, Hankuk University of Foreign Studies.
- Sepencer, Andrew. 1991. *Morphological Theory*. Blackwell.
- Song, Seok-Choong. 1967. *Some Transformational Rules in Korean*. Ph.D. dissertation, Indiana University.
- Suh, Cheong-Soo. 1991. On the Korean Verbs Ha and TOY [written in Korean]. *Language Research* 27(3), 481-505.
- Suh, Cheong-Soo. 1994. *Korean Grammar* [written in Korean]. The Deep-Rooted Tree Publishing.
- Suh, Chong-Soo. 1996. *Contemporary Korean Grammar* [written in Korean]. Hanyang University Press.
- Zwicky, Arnold. M. 1985. Clitics and Particles. *Language*, 61(2), 283-305.
- Zwicky, Arnold M. and Geoffrey K. Pullum. 1983. Cliticization vs. Inflection: English *n't*. *Language* 59(3), 502-513.