The "Gradient Structure" of Korean Words[*]

Hailey Hyekyeong Ceong
University of Victoria
hkc@uvic.ca

Hay and Baayen (2005) propose a probabilistic analysis of Indo-European word structure in which they argue that morphological structure is non-discrete – i.e., that it is *gradient*. This paper argues for a similar "gradient structure" approach to Korean words, with a particular focus on Hannate ("Sino-Korean") words. Hannate words are usually considered loanwords, but most of them acquire their lexical category by combining with native suffixes, as with the adjective *namca-tapta* 'manly'. The word *namsengmi* 'masculine beauty' may be interpreted as a complex or a compound word, depending on the treatment of *mi*. This problem of determinability is similar to the ambiguity encountered in English neoclassical compounds (Bauer 1998) . The adoption of Hannate roots into Korean eliminates the ideographic and tonal information that fixes their meanings in Chinese. This paper helps shed light on the understanding of Hannate words in Korean and explores various ways that language contact and the borrowing of words have consequences for the expanded lexicon of the borrowing language, which includes native items, borrowed items, and the products of reanalysis and analogy by speakers over time.

1   Introduction

This paper examines the structure of Korean words containing at least one root of Chinese origin. Previous literature has employed the terms 'Sino-Korean' to such refer to Chinese loanwords in Korean; however, I believe the term "Hannate" is more appropriate for these items. The term 'Hannate' parallels the term 'Latinate' in its usage, and therefore helps to improve terminological consistency in Linguistics. Since many Hannate words were coined in Japanese or Korean by employing Chinese roots and following Chinese word formation rules, categorizing them as loanwords may not appropriate. Furthermore, Chinese

---

roots do not generally occur independently in Korean. Most Hannate words acquire their lexical category by combining with native suffixes, as with *chinhata* 'intimate' in (1). It is unclear whether *chinhata* belongs to Sino-Korean, in the sense of being a loanword.

(1)

Hannate root	Hannate words	Hankul	Chinese character
chin	chinkwu 'friend'	친구	親舊
	chincel-han 'kind'	친절한	親切
	chin-hata 'intimate'	친-하다	親

The presence of *han* 漢 in all words related to Chinese characters (Chinese roots) suggests that the term "Hannate" is better than Sino-Korean. Consider: *Gudai hanyu* 'classic Chinese', *hanzi* 'Chinese character' in Chinese, and *hanmwun* 'Chinese writing', *hanca* 'Chinese character', and *hancae* 'Hannate words' in Korean. By contrast, the prefix *Sino* is used to refer to China in general, and it thus may refer to any one of fifty-six ethnic groups who speak different mother tongues. Chinese roots in Korean and Japanese come exclusively from languages of Han Chinese.

The purpose of this paper is twofold. First, I would like to apply a new approach to the analysis of the structure of Hannate words by exploring the hypothesis that "morphological structure is intrinsically gradient and has probabilistic structure" (Hay & Baayen 2005). This approach resists deterministic, discrete non-probabilistic approaches to the internal structure of words, and views morphological structure as structure that emerges from the regularities that identify the forms and meanings of words. My examination of Hannate words supports the probabilistic view of the internal structure of words and claim that the morphological structure of Korean is indeed 'gradient'. A root is identified as Hannate if its Korean alphabetic form can be replaced by a Chinese character (*Hanca*) and this character is indicated in the Korean dictionary: For example, the entry for *chinkwu* 'friend' in Korean dictionaries is "친구 (親舊). 명. 오랫동안 가까이 사귀어 온 벗. [friend. noun. 'an intimate mate you have known for a long time]' (translation by the author)". Like Latinate roots in English, the Hannate roots which make up 60% of the Korean lexicon are problematic for morpheme-based analyses. How do we account for non-native morphemes that may or may not be analyzable to all speakers in the borrowing language? How do we account for non-native morphemes that may be analyzable in certain contexts, but are not necessarily analyzable in other contexts? The definition of the morpheme as "the smallest individually meaningful element in

the utterances of a language" (Hockett 1958, cited in Aronoff 1981) is problematic for analyzing the internal structure of Korean words.

The first part of this paper shows the paradigmatic relation between Korean words of similar form and meaning. In the latter part of the paper, I will develop a three-dimensional space for forming Korean words in the framework of Bauer (1998). Hannate words, as shown in (1) above, are neither completely loanwords nor entirely native words, neither compounds nor derivatives. Hannate words pose the same problem in Korean as neoclassical compounds do in English; Bauer claims that it is difficult to determine whether a neoclassical root is an affix or a compound element, and thus whether the complete word is a derivative or a compound. Some monosyllabic Hannate roots are words, but most of them are not. Hannate roots can be words or bound morphemes which may or may not contain the full meaning of a true morpheme in Korean. Usually a combined form of two or more roots is a word, but it is not the case that all constituent morphemes contribute meaning to the meaning of the whole word. Moreover, except for nouns, all Hannate roots or combined root forms are assigned their grammatical category by native suffixes. Therefore, when Hannate words are taken into account, we see that Korean words have a fuzzy boundary, just like English word do (Bauer 1998). The present study shows the benefits of adopting a three-dimensional space for forming Korean words rather than separating them categorically into simple or complex, native word or loanword.

2   Fuzzy Boundaries of Hannate Words

Analyzing the morphological structure of Hannate words is problematic in a morpheme-based approach. Identifying morphemes in Hannate words is challenging. The words in (2) can be analyzed as simple or complex words. The Hannate roots *in*, *nam*, and *seng* are not words in Korean, but each root can indicate the meaning, 'person', 'male', and 'voice' respectively if it is written in Chinese characters.

(2)   a. miin   'the beauty'
       b. minam   'handsome'
       c. miseng   'sweet voice'

While *mi* in (2) looks like a prefix or base, *mi* in (3) looks like suffix. *namseng* 'male' in (3b) is a word, but *kak*, *kaksen*, and s*enmi* are not words; only *sen* 'line' and *kaksenmi* are words. It is plausible to analyze *mi* as a suffix, but it is impossible to analyze *kak* is a prefix. There are two homophonous prefixes *kak-* 'each' and *kak-* 'square' which are irrelevant to the whole meaning of word, as in *kaksenmi* ' the beauty of leg line'.

(3)　a.　kaksenmi　　　'the beauty of leg line'
　　　b.　namsengmi　　'masculine beauty'

Even though *mi* has a dictionary entry and is categorized as a noun, the root *mi* itself does not appear freely as a word, but appears only in idiomatic phrases as in (4).

(4)　a.　hankwuk-uy mi　　Korea-GEN　beauty　　'the beauty of Korea'
　　　b.　cohwa-uy mi　　　harmony-GEN beauty　'the beauty of harmony'
　　　c.　yucong-uy mi　　　round off-GEN beauty　'crowning glory'
　　　d.　cayen-uy mi　　　　nature-GEN beauty　　'the beauty of nature'

The native adjective *alumtaun* 'beautiful' modifies a native word like *moksoli* 'voice' or a Hannate word *yein* 'woman' as in (5), whereas the Hannate root *mi* cannot be used in this context, as exemplified in (6). Koreans say *kunyeuy alumtaum* 'the beauty of you' but do not utter *kunyeuy mi*.

(5)　　　Hannate word　　　Korean phrase　　　　glossary
　　　a.　miin　　　　　　　alumtaun　yein　　　'beautiful woman'
　　　b.　miseng　　　　　　alumtaun moksoli　　'beautiful voice'

(6)　　　na-nun　kunye-uy　alumtaum/*mi-e　　panha-ss-ta.
　　　　　I-TOP　her-GEN　beauty- LOC　　　was charmed by.
　　　　　'I was attracted by her beauty.'

In that case, is *mi* an allomorph of the native Korean adjective *alumtaun*, or are the two words synonyms? The situation is further complicated by the fact that *mi* cannot be interpreted as 'beauty' in isolation, since there are many homophones of *mi* that can signify meanings such as 'rice', 'taste', 'un-', and 'eyebrow'.

3　Paradigmatic Lexical Relations of Hannate Words

Hay and Baayen (2005) exploit the paradigmatic lexical relationship between inflectional paradigms and morphological families in English to claim that morphological structure emerges from the regularities that identify the forms and meanings of words. I apply this approach to Hannate words containing *mi* 'beautiful' to show the similarity of form and meaning in words which share the same roots. The properties of *mi* were presented in section two. Korean speakers who have not learned the meaning of the Hannate root may generalize *mi* as 'beautiful' from the parallels within the word group, since *mi* is a free root in the particular context discussed in (4). The other combining roots *in* 'person', *nye* female', and *nam* 'male' also obtain meaning from words containing the same roots as in figure 1.

However, not all internal structures of Hannate words are transparent. Often there is no large paradigm to facilitate a speaker's learning of the meaning of a root, as for instance for the root *kak* 'leg' in *kaksenmi*. The Hannate roots *kak* 'leg' is not a free root in Korean, and *kak* does not exist in a paradigm with many other words of similar form and meaning. Therefore, the meaning of *kak* 'leg' and the internal structure of words that contain *kak* probably will not emerge from the paradigm.

sangin	miin	sengin
'merchant'	'the beauty'	'adult'
chwunye	minye	swuknye
'ugly woman'	'the beauty'	'lady'
chwunam	minam	namca
'ungly man'	'handsome'	'man'
sengakka	miseng	umseng
'vocalist'	'sweet voice'	'voice'
tamhwa	mitam	sangtam
'talk'	'praiseworthy anecdote'	'consultation'
namseng	namsengmi	namsenghwa
'male'	'masculine beauty'	'virilism'
*kaksen	kaksenmi	sen
*kak	'the beauty of leg line'	'line'
	hankwuk-uy mi	hankwuk
?mi	'the beauty of Korea'	'Korea'

Figure 1. Examples of paradigmatic lexical relations of Hannate words in Korean. The meaning of root *mi* 'beautiful' is consistently shared in its paradigm; colours show how each root connects to words.

Another example is the Hannate word *sakwa* 'apple'. The meaning of *kwa* is transparent in Korean since there is a word family constructed from the paradigm, containing forms such as *kwail* 'fruit' and *kwaswuwen* 'fruit farm'. However, the meaning of *sa* is not transparent since there is no word family that shows the meaning of *sa*. Hay (2001) discusses the relationship between the transparent meaning of prefixed words and the base word in the context of "relative frequency". The fact that *kak* 'leg' and *sa* 'sand' are not free roots in Korean implies that the meaning of combined forms are more transparent than the meaning of roots. In addition, if a combined word has shifted or proliferated in meaning and thus is not transparent, then the meaning of its component parts will not be easily defined by referring to the combined form. An example of this

is the word *kakpon* 'script', the meaning of which is not obviously derivable from *kak* 'leg' or *pon* 'original; model'. In addition to a root being obligatorily bound, its form and meaning may not be shared in a large paradigm and its meaning may not be consistently shared.

In the next section I will develop a three-dimensional space for forming Korean words in the framework of Bauer (1998).

4   Dimensional Space for Word Formation

Bauer (1998) proposes a three-dimensional space for English word formation: a simple-compound dimension, a native-foreign dimension and an abbreviated-non abbreviated dimension. Bauer observes that some word types in English do not fit easily into a specific category; for example, it is difficult to determine whether neoclassical compounds are derivatives or compounds. This study applies Bauer's three-dimensional space to Korean word formation.
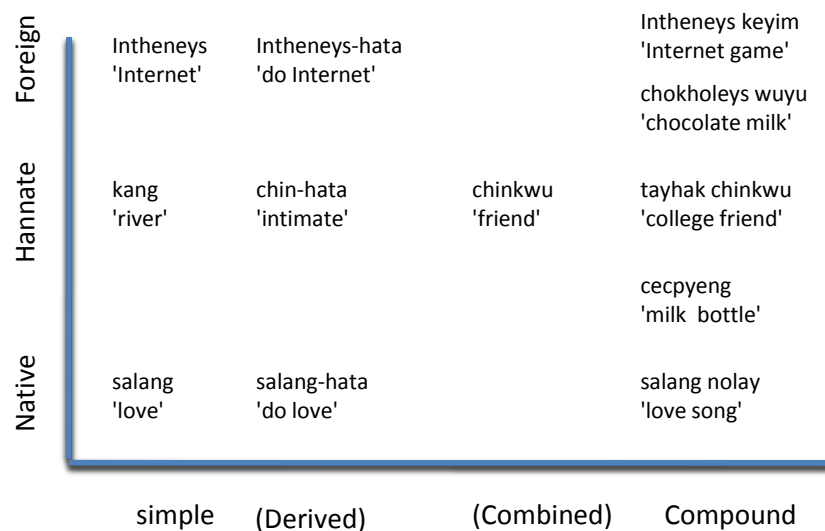

Figure 2. A two-dimensional matrix for Korean

I modify Bauer's proposed dimensional spaces slightly to suit Korean word formation. First, Korean needs three etymological spaces: Hannate, foreign loanword, and native. The native and loanword spaces sit at the edges of the dimension, and the Hannate word sits in the middle. The reason for this

modification is that foreign loanwords from modern Indo-European languages and native words are very different from Hannate words morphologically and phonologically (Cho 1999).

A second dimension along which Bauer distinguishes formation types is the traditional one of compound versus affixed versus simplex word: "Simplex words are un-analyzable. Derivatives are analyzable, but one of the elements involved is not a potential stem in English" (p.410). In figure 2, *salang* is a simple native word, *kang* is a simple Hannate word, and *intheneys* is a simple foreign word. The simple words become bases for the verbal suffix *–hata* 'do/be': *salang-hata, chin-hata* and *inteheneys-hata* are derived words. Words with Hannate bases but native Korean suffixes compromises between native and Hannate derivatives in the dimensional space; these words are neither loanwords nor native words. This paper adopts the term 'combining form' (Bauer 1983) for a root that has a Hannate constituent and consists of more than two roots in Korean. The majority of Hannate nouns and bases are formed by combining two or more roots: *chinkwu* 'friend'; *miin* 'the beauty'; *minam* 'handsome'. The combined forms are words; the combining forms are not words but bear particular affixal properties. Sohn (1999) categorizes the affixal combining form as a prefix or suffix, but the terms prefix and suffix are problematic for analyzing Hannate words in the same way that these terms are problematic for Neoclassical compounds such as *aerophobe, epitope*, and *lysosome* (Cannon 1992, cited in Baeskow 2004): two affixes alone can form a word. Compounding is a very productive word-formation process in Korean. Two simple words or two combined Hannate words coin compounds in Figure 2: *salang ssawum, tayhak chinkwu*, and *intheneys keyim* are example of native, Hannate , and Foreign compounds, respectively. There are compounds coined with native and Hannate words, and with foreign loanwords and Hannate words. For example, *wuyu* 'milk' is a Hannate word, *milkhu* 'milk' is a foreign loanword, and *cec* 'milk' is native. The compound word *chokholeys wuyu* 'chocolate milk' consists of the loanword and the Hannate word; the compound word *cecpyeng* consists of the native *cec* 'milk' and the Hannate word *pyeng* 'bottle'. These compounds sit at the interface between each dimensional space. We have seen in section two that the boundary between complex nouns and compounds in Korean is not straightforward. Hence we need graded structure from simple to compound along one axis when we draw a matrix graph.

Bauer suggests a third dimension, abbreviated versus non-abbreviated, because some neoclassical compounds fail to be taken into account in the previous two dimensions. This occurs when at least one of the elements is clipped: *heli-tele* 'TV or video camera mounted on a helicopter' (p. 411). However, Hannate words are mostly monosyllabic or disyllabic and Korean employs syllabic writing, so words in Korean are abbreviated using the initial syllable (or, rarely, the final syllable) rather than the initial letter. Therefore,

abbreviation is productive in compounds but not in simple and combining words. Combining and compounding are probably the most productive word creation spaces for forming a word with Hannate roots in Korean.

Up to now, I have demonstrated that Bauer's analysis of neoclassical compounds in English can be applied to understand the non-discrete structure of Hannate words in Korean. The point is that there are many intermediate stages on all three dimensions and that Hannate words occupy an area of dimensional space rather than a clear-cut category of Sino-Korean loanwords, as claimed by previous linguists. This section concludes with the table below that demonstrates five cells that arise from the intersection of the three dimensions. Native and foreign loanwords have been excluded from this table, as they are not the subjects of this paper.

Table 1.
Two-dimensional space for forming Hannate words

Type			Example		
Hannate	Simple	Full	kan 'liver'	kang 'river'	kum 'gold'
Hannate	Derived	Full	pyun-hata 'change'	chin-hata 'intimate'	
Hannate	Combined	Full	chinkwu 'friend'	miin 'the beauty'	namca 'man'
Hannate	Compound	Full	dayhak chinkwu ' college friend' kaceng kyoyuk 'home discipline'		
Hannate	Compound	Abbre-viated	noco (**no**tong **co**hap) 'labor union', hankwuk (tay**han**min**kwuk**) 'Korea'		

5   Conclusion

In this paper I have demonstrated that Hannate words in Korean should be analyzed as having gradient structure. Two approaches, Bauer (1998) and Hay & Baayen (2005), have helped to shed a lot of light on the issues that arise in understanding the internal structure of Hannate words. The three-dimensional space approach shows that we need to adopt a dimensional description of Hannate forms rather than assigning words to discrete categories. The paradigmatic lexical relation of Hannate words which I have developed in this paper is based on the approach proposed by Hay and Baayen, and shows how the meanings of Hannate forms emerge from a paradigm of similar forms and meanings.

References

Aronoff, M. (1981). *Word Formation in Generative Grammar*. Massachusetts: The MIT Press.

Baeskow, H. (2004). *Lexical properties of selected non-native morphemes of English*. Tübingen: Gunter Narr Verlag Tübingen.

Bauer, L. (1983). *English word-formation*. Cambridge: Cambridge University Press.

Bauer, L. (1998). Is there a class of neoclassical compounds, and if so is it productive? *Linguistics* 36(3) 403- 422.

Cho, Y. (1999). Language change and the phonological lexicon of Korean. *Historical linguistics 1999: selected papers from the 14th international conference on historical linguistics, Vancouver, 9-13 August 1999*.

Hay, J. (2001). Lexical frequency in morphology: is everything is relative? *Linguistics* 39. 1041-1070.

Hay, J. & Baayen. H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9. 7. 343-348.

Sohn, H. (1999). *The Korean language*. Cambridge: Cambridge University Press.