Korean Verbal Base Is Determined Based on Usage, Not Informativeness

Taewoo Kim
Seoul National University

1 Introduction

This paper elucidates the role of token frequency in forming a verbal paradigm in Korean. Since Mańczak (1958), token frequency has been viewed as a pivotal factor in paradigm leveling that regularizes allomorphs of a paradigm. There are, however, always the exceptions to the usual tendency of language change, and recently, Albright (2002) argues that the informativeness can successfully explain the examples that token frequency cannot. Albright & Kang (2008) apply the informativeness-based approach to the Korean verbal paradigm, and argue that the informativeness exerts more influence than token frequency in forming a verbal paradigm.

This work reexamines the paradigm leveling data in various dialects and introduces new sets of data that are hard to be dealt within the informativeness-based approach. Based on them, it is argued that a dominant single base of the paradigm, if it exists, is locally determined by each form's token frequency as a main factor. The discussion starts with the brief introduction to the morphophonology of Korean verbs.

2 Background

2.1 Verbal Suffixes

Korean verbal stems cannot stand alone in sentences but always occur with inflectional suffixes. The number of inflectional suffixes that can immediately follow the stems is large, but they are traditionally divided into three phonological classes named after the beginning sounds. *A-initial suffixes* are those starting with /a/ and /ə/ according to the last vowel of the preceding stems, which reflects the tongue root vowel harmony of the past. *ɨ-initial suffixes* are those that begin with /ɨ/ that disappear after a vowel or /l/ final stem, and *C-initial suffixes* are those that start with various consonants. These three classes of inflectional suffixes are shown in (1).

(1) Three classes of suffixes
   A-initial suffix: -a/ə, -(a/ə)ss-, -(a/ə)sə, -(a/ə)la, …
   ɨ-initial suffix: -ɨsi, -ɨl, -ɨn, -ɨm, -ɨlə, -ɨlsulok, -ɨmyə, -ɨmyən, -ɨni, -ɨnik'a, -ɨo …
   C-initial suffix: -kəna, -ke, -kes'-, -ko, -kun, -na, -nya, -nye, -nɨn, -ta, -taka, -tən,
                     -lako, -sɨpnita, …

There are a few phonological processes that bring about stem allomorphy when the verbal suffixes combine with a stem as shown in (2). Although these processes are regular post-lexical phonological processes of Korean, they sometimes incur surface ambiguities. For example, because of the final /ɨ/-deletion rule and the degemination rule, when combining with an A-initial suffix, the surface vowel of /ɨ/ and /ə/ final stems are always /ə/ as shown in (2a) and (2cii).

(2) a. Final /ɨ/-deletion
           kip'ɨ + ə   →   kip'_ -ə
   b. Glide Formation
       i) k'o + a   →   k'o-a ~ k'wa
       ii) ki + ə   →   ki-ə ~ kjə
   c. Degemination
       i) ka + a   →   ka
       ii) sə + ə   →   sə

2.2   Regular and Irregular Conjugation

Korean allows seven consonants at syllable coda position. While some of them do not alternate when combining with verbal suffixes, several classes of verbs alternate as shown in (3).[1] In (3a), /p/ of 'kup-' alternates with /w/ in front of vowel initial suffixes, while /p/ of 'cap-' does not alternate. /s/ and /t/ in (3b, c) do the same. Following the Korean linguists who work in *Kukeohak* (國語學, National Language studies) tradition, we refer to the latter classes with the non-alternating stem-final consonants as *regular conjugation classes* and the former with the alternating stem-final consonants as *irregular conjugation classes*. It is not predictable, partially because of the loss of the vowel length, whether the verbs ended with /p, s, t/ belong to the regular classes or the irregular classes when we only hear the C-initial suffix form.

(3)			A-initial	ɨ-initial	C-initial
a.	kup-	'to roast'	kuw-ə	kuw-ɨmyən	kup-ko
cf.)	cap-	'to catch'	cap-a	cap-ɨmyən	cap-ko
b.	kɨs-	'to draw'	kɨ_-ə	kɨ-ɨmyən	kɨs-ko
cf.)	pəs-	'to take off'	pəs-ə	pəs-ɨmyən	pəs-ko
c.	mut-	'to ask'	mul-ə	mul-ɨmyən	mut-ko
cf.)	mut-	'to bury'	mut-ə	mut-ɨmyən	mut-ko

There are other classes of irregular verbs that do not end with consonants. We refer to these cases as regular conjugation classes if the verb does not have stem allomorphy, and irregular conjugation classes if it does. (4) illustrates the conjugation of the three separate classes of stems that end with /lɨ/. The stem of (4a) alternates between /hɨlɨ-/ and /hɨll(ɨ)-/, and that of (4b)

---

[1] Korean has several other classes of irregular verbs, such as 'h'-irregular verbs, 'u'-irregular verbs, 'ɨ-irregular verbs and so on. Because they are not directly related to the discussion here, this paper will not deal with them in more detail.

alternates between /ilɨ-/ and /ilɨlɨ-/, so they belong to the irregular conjugation classes. (4c) is the only regular verb that ends with /lɨ/.

(4)			A-initial	ɨ-initial	C-initial
a.	hɨlɨ-	'to flow'	hɨll-ə	hɨlɨ-myən	hɨlɨ-ko
b.	ilɨ-	'to arrive'	ilɨl-ə	ilɨ-myən	ilɨ-ko
c.	t'alɨ-	'to follow'	t'al-ə	t'alɨ-myən	t'alɨ-ko

2.3 Surface Ambiguity

The irregular stems cause severe surface ambiguities in stem-final position before C-initial suffixes on the one hand, and before V-initial suffixes (A-initial and ɨ-initial suffixes) on the other as illustrated in (5). The locus of surface ambiguity is highlighted with shadow.

(5)	A-initial	ɨ-initial	C-initial
seu- 'to build'	sew-ə (←seu- ə)	se.u-myən	seu-ko
kup- 'to roast'	kuw-ə	ku.u-myən (←kuw-ɨmyən)	kup-ko
cap- 'to catch'	cap-a	cap-ɨmyən	cap-ko

The irregular class of verbs ending with /p/ have potential ambiguity with the regular /u/ stems in front of V-initial suffixes, whereas they are confused with the regular /p/ stems before C-initial suffixes. This is the case for other classes of irregular verbs, too. For example, /t/ irregular stems give ambiguity with /l/ final regular stems before V-initial suffixes, and with /t/ final regular stems before C-initial suffixes, as shown in (6).

(6) d.	A-initial	ɨ-initial	C-initial
ul- 'to cry'	uɾ-ə (←seu- ə)	ul-myən	ul-ko
mut- 'to ask'	muɾ-ə	muɾi-myən	mut-ko
mut- 'to bury'	mut-ə	mut-ɨmyən	mut-ko

The surface ambiguity not only occurs between regular conjugation classes and irregular conjugation classes but it also occurs between regular classes. Because of the syllable final neutralization of stops, all coronal consonants are pronounced as /t/ in that position as in (7). This is also true between /p/ regular stems and /pʰ/ regular stems and between /k/ regular stems and /kʰ/ regular stems.

(7)	A-initial	ɨ-initial	C-initial
tat- 'to close'	tat-a	tat-ɨmyən	tat-ko
nac- 'to be low'	nac-a	nac-ɨmyən	nat-ko
c'ocʰ- 'to chase'	c'ocʰ-a	c'ocʰ -ɨmyən	c'ot-ko
s'is- 'to wash'	s'is-ə	s'is-ɨmyən	s'it-ko

If we take the colloquial variations into consideration, the surface ambiguity becomes far more serious. For example, a velar consonant that appears at the syllable initial position usually

assimilates a syllable final coronal and labial consonant in the colloquial Korean. When this assimilation occurs, all consonant final regular stems are disambiguated with /k/ regular stems before velar initial suffixes.

Linguists have treated these ambiguities by positing distinct underlying representations. For instance, the underlying phoneme for stem final sound of /p/ irregular verbs is regarded as /w/, whereas the one for /p/ regular verbs is viewed as /p/ (Kim 1970). This is the same in the constrained-based theories such as Optimality Theory (Prince & Smolensky 1993). This kind of treatment assumes the exposure to the surface forms that occur before all three classes of suffixes, and the equal status of these three classes of suffixes over the speaker's mind in forming the paradigm. As shown in the next section, however, this assumption cannot be maintained if we try to account for the asymmetrical relation between the suffixed forms in paradigm leveling. Given the direction of paradigm leveling, it is likely that some suffixed forms may have greater influence over others within each paradigm.

3 Paradigm Leveling and the Verbal Base

3.1 The Direction of Paradigm Leveling in Dialects

Leveling of verb stems does not occur very frequently but examining the dialect data, you can see it is not that rare. Since Korean verb stems are not free forms, the leveling of the Korean verbal paradigm is accompanied by reanalysis. Some examples are given in (8). The stem form has been modified as an /h/ final stem in (8b), while in (8c), the boundary between a stem and a suffix has been reanalyzed such that the position of /i/ of the i-initial suffixes has been moved from the right to the left of the stem-suffix boundary which is indicated by the dash. It can be assumed that there are also covert reanalyses in (8a) and (8d).

(8)		A-initial	i-initial	C-initial	
a.	top - 'to help'	tow-a	tow-imyən	to:p-ko	(Standard)
	➜	tow-a	tou-myən	**tou-ko**	(Leveled, Lee 1975)
b.	k'o - 'to twist'	k'o-a	k'o:-ni	k'o:-ci	(Standard Form)
	➜	k'o-a	k'o:-ni	**k'o:-c^{h}i**	(Dialect, Kim 2001)
c.	kap^h-'to repay'	kap^h-a	kap^h-imyən	kap-k'o	(Standard Form)
	➜	kap^h-a	kap^{h}i-myən	**kap^{h}i-ko**	(Leveled Form, Baek 2013)
d.	sit - 'to carry'	sil-ə	sil-imyən	si:t-ko	(Standard Form)
	➜	sil-ə	sil-imyən	**si(:)l-ko**	(Leveled, Choi 1993)

As pointed out in the previous literature, these reanalyses would not happen unless the speakers take A-initial suffixed forms as a locus of reanalysis (Kwak 1994, Kim 2001, Kang 2006), and project the rest of the paradigm based on it (Albright 2002). When a stem ends with a consonant as in the examples above, however, the allomorphs in front of A-initial suffixes and i-initial suffixes are converged so that it is hard to determine which one is the locus of reanalysis. If stems that do not end with a consonant are taken into account, the primacy of A-initial suffixes is revealed as in (9) below. In (9a), the i-initial suffix form and the C-initial suffix form are changed to the exact same form as the A-initial suffix form. Thus, we can assume that reanalyses based solely on the A-initial suffix form are involved in the leveling of verbal paradigms.

(9)   A-initial   ɨ-initial   C-initial

			A-initial	ɨ-initial	C-initial	
a.	ha- 'to do'		hay-ə (hɛ)	ha-myən	ha-ko	(Standard Form)
		➢	hɛ	**hɛ-myən**	**hɛ-ko**	(Leveled Form)
b.	sə- 'to stand'		sə	sə-myən	sə-ko	(Standard Form)
		➢	sə	**sɨ-myən**	**sɨ-ko**	(Leveled Form)

In sum, in all the examples above, and the examples in Kim (2001) and Kang (2006), A-initial suffix forms do not change. Although ɨ-initial suffix forms also do not change, since it happens only when A-initial suffix forms and ɨ-initial suffix forms are identical, the dominant status is given only to A-initial suffix form. C-initial suffix forms are subject to change when paradigm leveling occurs, towards the stem form which conforms to that before A-initial suffix forms.

3.2 Single Base Hypothesis and the Informativeness-based Approach

To explain the dominance of an A-initial suffix form, it is necessary to escape the premise that stem forms in front of three classes of suffixes have the same influence within each paradigm. Kang (2006) argues that, applying Albright's (2002) Single Base Hypothesis that explains the direction of paradigm leveling in other languages such as Latin, paradigm leveling has occurred because there is a dominant suffixed form in verbal paradigm, a base, and other suffixed forms are projected from it. In other words, the speakers do not form a paradigm inductively from hearing all suffixed forms but forms a paradigm deductively or abductively from one surface form entrenched in mind.

Albright & Kang (2008) subsequently assert that the informativeness plays a pivotal role in determining the base. The informativeness here can be restated as predictability of deriving other surface forms in a paradigm from a certain surface form in the same paradigm. The most informative form, therefore, has to "preserve the most contrasts and permit accurate and productive generation of as many word forms as possible" (Albright 2002:129). All other forms are projected from the most informative base by probabilistic morphological rules, and paradigm leveling occurs, in my understanding, when these rules do not work. Morphological mapping rules are built based on phonological contexts, and the reliability that is calculated using the type frequency of the rule, the number of competitors, etc. is given to each rule.[2]

The informativeness-based approach succeeded mainly in explaining the directionality and the selectionality of the paradigm leveling in /s/ final nominals in Latin (Albright 2002). In this change, because the nominative form had been replaced by the oblique form, it is contrary to the usual direction of paradigm leveling. In addition, this change is restricted to masculine and feminine polysyllables words. Monosyllables and neuter nominals are not affected in general. In fact, the directionality problem can be solved by the other approaches such as Manczak's (1958) fourth law of analogy, but the selectionality problem can be solved only when we consult the informativeness-based approach.

Albright & Kang (2008) applied the informativeness-based approach to Korean in the same way as Albright (2002) did to Latin. They calculated the informativeness of surface forms with three classes of suffixes. As a result, they verify that A-initial suffix forms are indeed the most informative forms. Because it is consistent with the direction of paradigm leveling illustrated

---

[2] One can refer to Albright (2002) for a more detailed process of forming the morphological rules and calculating its reliability.

above, they conclude that the base selection of Korean is also explainable with the Single Base Hypothesis and the informativeness-based theory. At the same time, however, they are deeply concerned with the role of token frequency. This is because A-initial suffix forms appear with the largest frequency.[3] Since token frequency of each predicate is not reflected in calculating the reliability of rules, if the most frequent form and the most informative form are the same form, the problem arises as to which factor is decisive in selecting a base. To cope with these difficulties, Albright & Kang (2008) appeal to the fact that although C-suffixed forms are also found with high frequency, there is no case that the paradigm is leveled towards a possible stem form in front of C-suffixed forms, and that it is hard to predict the direction of paradigm leveling solely by token frequency in numerous languages.

The approach discussed above complements the limitation of previous approaches that utilizes an underlying representation and rules (or constraints) in that it can explain both the synchronic and diachronic problems at the same time. Nevertheless, it is still questionable as to the role of token frequency in language learning and diachronic change. In this regard, the next section presents examples that contradict the directionality of paradigm leveling in Korean, and argues that the informativeness-based explanation cannot give a satisfying answer to the difference in the direction of paradigm leveling. Rather, it is shown that the token frequency alone is sufficient in predicting this asymmetry in directionality.

4   Exceptions in the Direction of Paradigm Leveling

The above section reviewed the previous studies that claim that A-initial suffix forms play dominant roles in forming the paradigm and in paradigm leveling in various dialects of Korean. They also present the results of computer modeling that support their argument that the A-suffixed form is indeed the most informative form among the paradigm. Yet this is not an absolute generalization. This section introduces new sets of data that are opposed to the observation given in Kang (2006) and others. It is expected that these counterexamples to the paradigm leveling are of critical importance in clarifying the verbal base controversy.

Two kinds of data are presented: i) variation in /lɨ/-irregular adjectives in various kinds of data, that is, the SNS survey, the colloquial corpus and the dialectal implication and ii) the diachronic restructuring of /-hʌ-/ suffixed adjectives and verbs. In these lines of data, the formation of the paradigm is centered on the ɨ-initial suffix form, which makes these examples interesting.

4.1   SNS Data

Han (2014) surveys the usage of verbs that belong to the /lɨ/-irregular conjugation class from Tweeter, a social network service (SNS), for twenty-four hours. In various dialects in Korean, /lɨ/-irregular class of verbs are regularized on the basis of the A-initial suffix form as in (10).

(10)			A-initial	ɨ-initial	C-initial	
	hɨlɨ-	'to flow'	hɨll-ə	hɨlɨ-myən	hɨlɨ-ko	(Standard Form)
	➔		hɨll-ə	**hɨlli-myən**	**hɨlli-ko**	(Leveled From)

---

[3] This is partly because that /-a(ə)/ of A-initial suffixes has the most basic and unmarked semantics. It is referred to as 'infinitive' by Martin (1992), and thus can be used as a declarative, interrogative, imperative, hortative and so on. It therefore shows up with high frequency in casual speech and children speech (Kang 2006). Another reason that the A-initial suffix form has high frequency is that the past tense suffix /-a(ə)s'-/ is an A-initial suffix.

Han (2014) focuses on whether the leveled stem form occurs before ɨ-initial and C-initial suffix forms, and finds out that there is a discrepancy between predicates. Even though Han (2014) does not point out, the frequency of the leveled forms differs significantly according to whether the verbal stem is an adjective or a verb.[4] As shown in table (1), the adjectival stems (shaded rows) are surprisingly less leveled than verbs, whereas the verbs are more leveled than expected.

Table 1: The frequency of 9 most frequently used /lɨ/ final predicates in SNS (drawn from Han 2014)

	Total Frequency	Leveled Form Frequency	O/E Ratio	POS
molɨ-	4154	162	1.39	VERB
talɨ-	1987	7	0.12	ADJ
pulɨ-	1776	75	1.51	VERB
olɨ-	1050	4	0.13	VERB
ppalɨ-	669	6	0.32	ADJ
nulɨ-	325	33	3.64	VERB
kilɨ-	323	13	1.44	VERB
olpalɨ-	259	0	0	ADJ
palɨ-	214	0	0	ADJ

4.2   Colloquial Corpus

The colloquial corpus established by Seo (2015) shows the same tendency in the leveling of /lɨ/ irregular verbs. Table 2 show the same list of verbs that are surveyed in Han (2014). In the following table again, it is observed that the adjectives are more resistant to the leveling than the verbs. There are a few words that deviate from this trend, but the overall tendency that the verbs are more liable to the leveling remains unchanged. This observation that verbs are more prone to the leveling survives if we expand the list to the all /lɨ/-irregular predicates in the corpus. Seo (2015) contains twelve /lɨ/ irregular verbs and six /lɨ/ irregular adjectives. The leveled forms of verbs occur 17.4 percent on average, while the leveled forms of adjectives occur only 7.8 percent on average.

Table 2: The frequency of the 9 /lɨ/ final predicates from Seo (2015)

	Total Frequency	Leveled Form Frequency	O/E Ratio	POS
molɨ-	2360	57	0.64	VERB
talɨ-	945	28	0.78	ADJ
pulɨ-	285	27	2.51	VERB
olɨ-	71	5	1.87	VERB
ppalɨ-	160	4	0.66	ADJ
nulɨ-	125	29	6.17	VERB
kilɨ-	37	2	1.43	VERB
olpalɨ-	36	0	0	ADJ
palɨ-	24	0	0	ADJ

---

[4] /t'alɨ-/, included in Han's (2014) original list, is omitted in this list because it does not belong to the /lɨ/- irregular conjugation class. See (4c) for the conjugation class of /t'alɨ-/.

The following two homonyms in Table 3 clearly show the contrast between a verb and an adjective in terms of paradigm leveling. In the comparison of the verb and the adjective /palɨ-/, the leveled form of the verb /palɨ-/ shows up quite often, but no leveled adjective appears at all.

Table 3: the frequency asymmetry by POS between homonyms

	POS	Leveled	Unleveled	Total
palɨ-	Verb ('to apply')	22	56	78
	ADJ ('to be right')	0	24	24
olpalɨ-	ADJ ('to be right')	0	36	36

4.3   Implicational Relations in Dialects

Some dialects of Kangwon Province, the east-northern part of South Korea, show interesting implicational relations in the paradigm leveling of /lɨ/ irregular predicates. Four verbs and adjectives that are contained in AKS (1985) are surveyed. As a result, in overall fashion if not all, in the cities where the adjectives have undergone the paradigm leveling, the verbs also have gone through it, but not vice versa. In other words, there is a city where only verbs underwent the paradigm leveling, but there is no city where only adjectives are regularized. For example, we can find that both adjectives and verbs underwent the leveling in Wonju and Cheolwon, but only verbs are leveled in Kosung. There are many cities in which both verbs and adjectives are not leveled. The regional distribution is schematized in the Figure 1.[5]

Figure 1: The schematized map of Kangwon Province
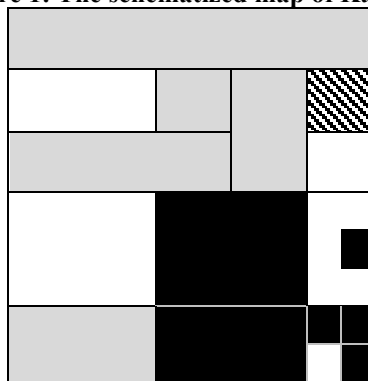
The regional distribution of paradigm leveling is worthy of attention. The cities where no paradigm leveling has happened cluster around the east-southern part of Kangwon Province, while the cities with paradigm leveling are located in the west-northern part of it. The city in which only verbs underwent the paradigm leveling is located in east-northern area. Considering this regional distribution, it is likely that the verbs are leveled by a certain intra-paradigmatic force from the east-norther shore, but the adjectives have probably been regularized by an inter-paradigmatic analogical force.

---

[5] In Figure 1, each rectangle represents a city. The grey rectangles represent the cities where both the adjectives and verbs underwent the paradigm leveling. The black rectangles represent the cities where no adjective nor verb underwent the paradigm leveling. The diagonal-lined rectangle is the city where only verbs went through the paradigm leveling.

4.4 Diachronic Restructuring

The verbs and the adjectives differ in diachronic paradigm leveling, too. Middle Korean /-hʌ-/ is a suffix that derived from the lexical verb /hʌ-/. Because /hʌ-/ conjugates irregularly, /-hʌ-/ suffixed predicates also irregularly conjugate. When it is combined with A-initial suffixes, the epenthetic [y] is added. The irregularity of /hʌ-/ is shown in (11). It still forms an irregular class as illustrated in (9a) earlier.

(11)   A-initial   ɨ-initial   C-initial
        hʌ-   'to do'   hʌy-a   hʌ-myən   hʌ-ko   (Middle Korean)

There are several verbs and adjectives that are derived from /-hʌ-/ in Middle Korean. They have been leveled and in Present Day Korean they form a regular conjugation class. Again in this diachronic restructuring, according to Chung (2015), the direction in paradigm leveling differs between adjectives and verbs as illustrated in (12) and (13).

(12)   A-initial   ɨ-initial   C-initial
        tikhʌ- 'to guard'   **tikhʌy-a**   tikhʌ-myən   tikhʌ-ko
                >   **tikʰɨy-ə**   tikʰɨy-myən   tikʰɨy-ko
                (>   cikʰi-ə   cikʰi-myən   cikʰi-ko   )

(13)   A-initial   ɨ-initial   C-initial
        kʌthʌ- 'to be same'   kʌthʌy-a   **kʌthʌ-myən**   kʌthʌ-ko
                >   katʰ-a   **katʰ-ɨmyən**   katʰ-ko
        tyohʌ- 'to be good'   tyohʌy-a   tyohʌ-myən   tyohʌ- ko
                >   coh-a   coh-ɨmyən   co-kʰo

At each set of lines of a word, the upper line represents the conjugation before the paradigm leveling, and the lower line shows the conjugation after the paradigm leveling. In (12), where the paradigm leveling of the verb /tikhʌ-/ is depicted, we can see that the paradigm has restructured on the basis of the A-suffixed form. In contrast to the verb in (12), the adjectives in (13) have restructured based on the ɨ-initial suffix form. The new stem form that arises because of the reinterpretation of the stem-suffix boundary in front of ɨ-initial suffixes has spread to other suffixed forms. As a result, the allomorphy in front of A-initial suffixes has been eliminated.

4.5 The Base of Adjectives

The examples presented in this section show that the adjectives do not behave differently from the verbs. In the SNS data, as well as the data from colloquial corpus and dialectal data, the adjectives resist regularization. In the diachronic data, the adjectives are leveled in accordance with the form in front of the ɨ-initial suffix form. All these anomalies can be accounted for when we consider that the ɨ-initial suffix form exerts pressure on the paradigm, that is, the ɨ-suffixed form is the base for adjectives.

One might be skeptical about the data demonstrated in this section other than 4.4 because they still show the adjectives have been regularized in the same fashion, but at a different degree, as the verbs have. It is true that there is no example of leveling before the ɨ-suffixed form. As discussed in the following section, the base is likely to be determined based on the token frequency of each conjugated form, so that the base for /lɨ/ irregular stems can be either the ɨ-initial suffix form or the A-initial suffix form. In this paper, however, it is assumed that all

examples in this section indicate that the base for the above adjectives is the ɨ-initial suffix form. The reason that the stem before the ɨ-initial suffix form is not extended to the A-initial suffix form is ascribed to an independent reason. This is discussed in more detail in section 5.

5   Paradigm Leveling and Token Frequency

In the above section, I presented the examples that are contrary to the usual trend of paradigm leveling discussed in previous literature and argued that these deviations occur due to the fact that the base of the adjectives was in fact, unlike verbs, the ɨ-initial suffix form. Under the informativeness-based approach, it is difficult to explain this discrepancy that the surface forms that share the same phonological context have different bases. One might try to find a reason for this disparity from the differences in the suffix list that verbs and adjectives combine, but suffixes can combine equally with verbs and adjectives, except that adjectives cannot combine with the realis marker '-*ni*-' and the hortative marker '-ca.' This may affect the informativeness of each suffixed form, but the effects are minimal. Moreover, this is already taken into consideration in calculating the informativeness of each suffixed form because verbs and adjectives are included in the learning data set in Albright & Kang (2008).

The reason that the informativeness-based approach cannot account for the base difference between verbs and adjectives is that informativeness is global in its application. This means that the most informative morphological form, the A-initial suffix form in the case of Korean, has to act as a base in every individual predicate. It is therefore impossible on the premise of the informativeness-based theory that the base of the phonologically same, but distinct forms have the A-initial suffix form as the base on the one word, and the ɨ-initial suffix form on the other.

It is worth pointing out that Korean is somewhat in a different situation in applying the informativeness-based theory than Latin, a language in which the theory has established a major success. First of all, A-initial suffix form is the most frequent form, as well as the most informative form in Korean. In the Latin paradigm leveling, there are some classes of nominal paradigms where the informativeness conflicts with token frequency. For example, the token frequency of nominals that end with /s/ in the oblique cases is higher than those in the nominative cases, but the paradigm leveling only occurs with polysyllabic, non-neuter nouns (Albright 2002). The reason that monosyllabic and neuter nouns have not undergone the leveling is that their informativeness is low. In other words, in the case of Latin, in some classes of nominals, the most frequent form is not the most informative form, and informativeness overrides token frequency so that the leveling does not occur. Unfortunately in Korean, token frequency and informativeness both position the A-initial suffix form to be the dominant base of the paradigm. Thus, even if the A-initial suffix form is proven to be the most informative, it is still necessary to find evidence that token frequency does not play a crucial role in base selection in Korean. We have already seen that Albright & Kang (2008) struggle with the same issue.

In addition, the informativeness of the A-initial suffix form is not significantly higher than other suffixed forms. In the case of Latin in Albright (2002), the nominative form that has been leveled is 11~13% less informative than others. In Korean, on the other hand, according to Albright & Kang (2008), the C-initial suffix form which is considered to be the least informative is only about 3% less informative than that of the A-initial suffix form which is the most informative. Albright & Kang (2008) assume that this difference would be greater given all the number of suffixes in use, but the difference between the two is more likely to be reduced because the type frequency of C-suffixes is greater than that of A-suffixes.

It is thus inevitable that if there is evidence exists of token frequency being employed in choosing the base, the importance of informativeness in a verbal paradigm is weakened. The data shown in section 4 can be considered credible evidence in showing the correlation between token frequency and base selection. This paper argues that the discrepancy in the direction of leveling between verbs and adjectives in Korean is attributed to the difference in the collocational relationships with suffixes. Verbs and adjectives may combine with almost the same list of endings, but differ in the actual frequency combined with each token because there is a difference in the function they serve in the sentence. Although both verbs and adjectives can be used in attributive and predicative forms, verbs are mainly used to predicate, whereas the main function of adjectives is to modify nouns. In predicative use, A-initial suffix forms are used in high frequency, while attributive forms are mostly ɨ-initial suffix forms. This is confirmed in the frequency comparison between verbs and adjectives. In other words, the token frequency is consistent with the diverged direction of paradigm leveling in the data. The counterexamples introduced in section 4 can be successfully explained by the token frequency.

Table 4 shows the most collocated suffix with each verb surveyed by Han (2014). We can see that adjectives are most frequently combined with the ɨ-initial suffix. One noticeable fact is that in the case of adjectives, the portion combining with one ɨ-initial suffix is remarkably high.[6] One exception is /palɨ-/ which is most often combined with the C-initial suffix /-ke/. Although this deviates from other adjectives, the suffix /-ke/ is an adverbial ending that modifies the verbs so it still performs a sort of attributive function. Furthermore, if the base of /palɨ-/ is the C-suffixed form, the resistance to the leveling is still understandable in that the stem form in front of the C-initial suffix is different from that of the A-initial suffix.

Table 4: Each verb's collocational relationship

	Most Co-occurred Suffix	Suffix Type	POS
molɨ-	-nda (13.1%)	C-initial	VERB
talɨ-	-n (39.1%)	ɨ-initial	ADJ
pulɨ-	-nun (13.1%)	C-initial	VERB
olɨ-	-n (16.6%)	ɨ-initial	VERB
ppalɨ-	-n (39.2%)	ɨ-initial	ADJ
nulɨ-	-ə (16.9%)	A-initial	VERB
kilɨ-	-ə (16.4%)	A-initial	VERB
olpalɨ-	-n (80.7%)	ɨ-initial	ADJ
palɨ-	-ke (44.9%)	C-initial	ADJ

The same tendency is replicated in the written and the colloquial corpus. As shown in Table 5, in written Korean, the verbs combine most frequently with A-initial suffixes, whereas the adjectives combine most frequently with ɨ-initial suffixes. In spoken Korean, the attributive endings are most frequently used in adjectives as in Table 6.

---

[6] Since Han's (2014) results show only the most collocated suffix and does not present the overall figure, it is not clear in the table whether the verbs are most frequently combined with A-initial suffixes. Comparing to the ratio of ɨ-initial suffixes with the adjectives, one can question the percentage of the A-initial suffixes with the verbs to be too low to be the base. However, this should be viewed as a characteristic of SNS writing. Because it is not real-time interactional chatting, /-a(e)/ endings that frequently occur in spontaneous casual speech are likely to be underrepresented.

Table 5: Frequency difference in written Korean [Sejong Corpus Dec 2011] (Kim 2015: 241)

	-n, -l (ɨ-initial)	-e/a-, -ess/ass- (A-initial)	TOTAL
VERB	233,993 (13.6%)	478,283 (27.8%)	1,717,085
ADJECTIVE	132,277 (34.2%)	38,541 (9.9%)	385,954

Table 6: Suffix frequency after adjectives in spoken Korean [Sejong Corpus Dec 2011]

	-n, -l (ɨ-initial)	-e/a-, -ess/ass- (A-initial)	TOTAL
ADJECTIVE	9,268 (33.1%)	3,317 (11.8%)	28,021

It is notable that the difference in token frequency is consistent with the direction of paradigm leveling. Adjectives and verbs are not different in their informativeness, but there is a clear difference in the suffixes that they combine with frequently. The verbs that combine often with A-initial suffixes have undergone the paradigm leveling based on the A-initial suffix, but the adjectives that combine mostly with ɨ-initial suffixes have not.

The more direct relationship between token frequency and the direction of paradigm leveling is revealed in the diachronic restructuring of 'hʌ' irregular verbs examined in Chung (2015) and summarized in section 4.3. The Table 7 shows the frequency of the suffixes that co-occur with each of the /hʌ/ adjectives.[7] This table again verifies that adjectives mostly combined with ɨ-initial suffixes while the verb 'tikhʌ-' conjugated mostly with A-initial suffixes.

Table 7: The frequency combining with suffixes (Slightly modified from Chung 2015)

	A-initial	ɨ-initial	C-initial
kʌthʌ-	630 (12.0%)	3096 (59.2%)	1171 (22.4%)
tyohʌ-	38 (1.8%)	1611 (75.1%)	447 (20.9%)
olhʌ-	5 (0.6%)	489 (63.0%)	242 (31.2%)
manhʌ-	37 (20.3%)	86 (47.3%)	42 (23.1%)
tikhʌ-	13 (65%)	4 (20%)	3 (15%)

To recapitulate, the frequency survey confirms that ɨ-initial suffixes are the most frequent suffixes to combine with adjectives. Given that A-initial suffixes combine most frequently with verbs, it is token frequency that explains the base selection for both verbs and adjectives most satisfactorily.

It is worth noting that some of the frequencies of the base that I present above are not very high. For example, the percentage of 33.1% in Table 6 indicating the rate at which adjectives combine with attributive endings, a number that slightly increases if all the endings are considered, is never high enough to absolutely check the rest. Thus, we can assume that the speakers usually think of a couple of suffixed forms that are most active in mind, rather than a whole set of suffixes, and form the rest of the paradigm based on that. The strength of activeness in mind is related to the semantics of each predicate as well as the token frequency. In other words, the unmarked function of adjectives is to modify, and so they come to the speaker's mind with the attributive /-ɨn, -ɨl/, while the verbs are recalled in the speaker's mind with /-a(ə)/. In this way, the dominance within the paradigm is determined relying both on the token frequency and the semantics of each predicate. I would hypothesize that the semantic relationship takes precedence over token frequency, but this claim requires further follow-up research.

---

[7] This table is drawn from Chung (2015:318). The last line, the frequency for 'tikhʌ-' is added by the author.

One more concern in the adjectival paradigm is that, as pointed out in section 4.5, the ɨ-initial suffix form does not extend to the A-initial suffix form. Or, to put it differently, it is the A-initial suffix form that serves as a base of those adjectives and the ɨ-initial suffix form happens to be a suppressor by virtue of its high frequency. However, it is doubtful whether such a question is tenable in the case of /lɨ/ final verbs in Korean.

The problem lies in the fact that even if the stem form of the ɨ-initial suffix form has extended to the A-initial suffix form, it cannot be captured on the surface. This is because the morpho-phonological rule that /lɨ/ + /-a(ə)/ → /-llə/ is very robust. Because there is a regular phonological process /ɨ/ deletion before /-a(ə)/, one might think it should be /-lə/ when the ɨ-initial suffix form extends to A-initial suffixes, but this rule does not seem to work when the preceding consonant is /l/. Therefore, the surface forms of the paradigm remain constant after the extension. The speaker, rather, recovers /l(h)/ final regular stems when they are met with the A-suffixed form that ends with /-la(ə)/ as in (14). Thus, it is inadequate to say that the ɨ-initial suffix form has not extended to the A-initial suffix form.

(14)		A-initial	ɨ-initial	C-initial
t'alɨ-	'to follow'	t'al-a	t'alɨ-myən	**t'alɨ-ko**
	>	t'al-a	t'alɨ-myən	**t'al-kʰo**

(Kim 2001:100)

It is still problematic that in spite of the fact that the ɨ-suffixed form is the base, the leveling that verbs have undergone has occurred in adjectives. I conjecture that the paradigm leveling in adjectives occurs because of the inter-paradigmatic analogy. Because there is a large number of verbs that have already been leveled, the speakers apply the same process to the adjective classes. The possibility that this change has occurred is hinted at by the regional distribution discussed in section 4.4.

6 Conclusion

This work investigates the role of token frequency in the selection of the base in verbal paradigms. Two sets of data that contradict the findings in the previous literature are presented. The first set of data, drawn from the SNS, the colloquial corpus and the dialects, tell us that the /lɨ/ final adjectives are less prone to be leveled than /lɨ/ final verbs. The second set of data, drawn from the historical corpus, shows us that there is a case that the paradigm has been leveled towards the stem that occurs before the ɨ-initial suffixes.

To explain the asymmetry that is shown between the verbs and the adjectives, it is supposed that the base of those adjectives is the ɨ-suffixed form. Since the phonological structures of both are not discriminated, it is difficult to explain this asymmetry by the difference of informativeness. This paper, rather, argues that the token frequency of the ɨ-suffixed form plays a decisive role in determining the base. It is confirmed that those adjectives that have been leveled towards the stem form in front of the ɨ-initial suffixes have a substantially higher frequency with the ɨ-initial suffixes through the frequency survey of the SNS, the colloquial corpus and the historical corpus.

References

AKS [The Academy of Korean Studies]. 1985-1995. Hankwukpangencalyocip ('Korean Dialectal Data') vol. 1-9. Sengnam, Korea: The AKS.

Albright, A. & Y, Kang. 2008. Predicting innovative alternations in Korean verb paradigms. In *Proceedings of the 18ᵗʰ International Congress of Linguistics*, 893-913.

Albright, A. 2002. The Identification of Bases in Morphological Paradigms. Doctoral dissertation, UCLA.

Baek, E. 2013. Yongen ekan mal 'u' sapip hyensang [The phenomenon of 'ɨ' insertion at verbal stem-final], *Hangul* 299:49-75.

Choi, M. 1993. Ekanuy caykwucohwawa kyocheyhyenguy tanilhwa panghyang [Stem restructuring and the direction of paradigm leveling]. *Sengkoknonchong* 24:1599-1642.

Chung, K. 2015. *Hankwuke yongen hwalyong cheykyeyuy thongsicek pyenhwa* [Historical Change of the Korean Conjugation System]. Doctoral dissertation, Korea University.

Han, S. 2014. *Pwulkyuchik yongenuy hwalyonghyeng yenkwu* [A study on Conjugated Forms of Irregular Verbs in the Korean Language]. Doctoral Dissertation, Pusan National University.

Kang, Y. 2006. Neutralization and variation in Korean verbal paradigms. In *Harvard Studies in Korean Linguistics* XI, ed. by Kuno, S. et al., 183-196.

Kim, C. 1970. Swowi Pyenkyekywongenuy pipyenkyekseongey kwanhaye [On the non-irregularity of so-called irregular predicates], *Hankwukeoneomwunhak* 8-9:1-11.

Kim, H. 2001. hwalyonghyenguy caypwunsekey uyhan yongen ekan caykwucohwa [Verb stem restructuring on the basis of reanalyzing conjugated forms], *Kwukeohak* 37:85-113.

Kim, T. 2015. Hwalyongekankwa emiuy ehwi phyosang [The mental representation of inflected forms]. *Hyengthaylon* 17(2):233-263.

Kwak, C. 1994. *Kyeyhap nayeyseouy tanilhwaey uyhan ekan caykwucohwa* [Stem restructuring by the paradigm leveling], in *Kukeohak Yenkwu*, ed. by Hankukeo Yeonkwuhoy, 549-586. Seoul: Tayhaksa.

Lee, P. 1975. Phonological and morphological studies in Kyonggi subdialect. M.A. Thesis, Seoul National University.

Mańczak, W. 1958. Tendances generales des changements analogiques. *Lingua* 7:298-325.

Martin, S. 1992. *A Reference Grammar of Korean: A Complete Guide to the Grammar and History of the Korean Language*. Rutland, Vt.: C.E. Tuttle Co.

Prince, A. & P. Smolensky. 1993 [2004]. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell Publishing.

Seo, S. 2015. *Hankwuke kwue pinto sacen* [Colloquial Korean Frequency Dictionary] Vol. 1, 2. Seoul: Hankwuk Mwunhwasa.