Gradient vowel cooccurrence restrictions in monomorphemic native Korean roots*

Sung-Hoon Hong
(Hankuk University of Foreign Studies)

Hong, Sung-Hoon. 2010. Gradient vowel cooccurrence restrictions in monomorphemic native Korean roots. *Studies in Phonetics, Phonology and Morphology* 16.2. 279-295. This paper presents evidence for gradient vowel cooccurrence restrictions holding in the monomorphemic roots of the native Korean words. Noting that vowel cooccurrence often emerges as a result of vowel harmony and vowel harmony was once active in Korean, I examine root-internal vowels to see if they conform to the general patterns of the vowel harmony. For this examination, I extract multisyllable words registered as monomorphemic and native in *Pyo-jun-gug-eo-dai-sa-jeon* (The National Institute of the Korean Language, 2001). I then measure the degree to which vowels occur together utilizing the metric of O/E, the 'observed' frequency divided by the 'expected' frequency (Pierrehumbert 1993). The O/E values obtained between two vowels in the first three syllables of the multisyllable words are compared to the possible and impossible patterns of the vowel harmony. The results show that although not absolute, vowel cooccurrence restrictions similar to those enforced by the vowel harmony hold in a gradient manner within a monomorphemic native root in Korean. **(Hankuk University of Foreign Studies)**

Keywords: O/E, observed frequency, expected frequency, phonotactics, gradient, vowel cooccurrence, vowel harmony, ideophones

1. Introduction

This paper examines vowel cooccurrence restrictions that hold in the monomorphemic roots of the native Korean words. Restrictions on vowel ooccurrence often emerge as a result of vowel harmony. Vowel harmony was once active in the history of Korean, enforcing fixed patterns of vowel cooccurrence both within a root and across the root-suffix boundary (S-N. Lee 1947, K-M. Lee 1972, C-W. Kim 1978). In present-day Korean, however, vowel harmony is no longer productive, and its vestage is observed only in ideophones and in the suffixation of /ə/ in verbs and adjectives (Kim-Renaud 1976, Y-S. Kim 1984, B-G. Lee 1985).

Based on phonotactic knowledge being gradient (Pierrehumbert 1993, Berkley 1994, Frisch 1996, Frisch et al. 2004, Kawahara 2007, Coetzee & Pater 2008), this paper presents evidence that the patterns of the vowel harmony are still manifested in present-day Korean even within a non-ideophonic root. This finding is significant because most of the evidence

* This work was supported by Hankuk University of Foreign Studies Research Fund of 2009. A part of this paper was presented at the LSK & KASELL 2010 International Winter Conference on Linguistics, held at Konkuk University from Jan 5 to Jan 6, 2010. I appreciate anonymous reviewers for their helpful comments.

for gradient phonotactics in the literature has come from dissimilatory phenomena related to the Obligatory Contour Principle (McCarthy 1986, Yip 1988).[1]

The data in this study are the vowel pairs extracted from multisyllabic words registered as monomorphemic and native in *Pyo-jun-gug-eo-dai-sa-jeon* (*Grand Dictionary of Standard Korean*, The National Institute of the Korean Language (NIKL), 2001). The gradient vowel phonotactics will be presented by using Pierrehumbert's (1993) O/E, the ratio of the observed number of occurring vowel pairs (O) to the number that would be expected if vowels combined at random (E). We will see that some of the vowel pairs are severly underrepresented, and these underrrepresented pairs correspond to those which are not predicted to occur according to the vowel harmony.

This paper is organized as follows. In section 2, we will review the basic patterns of the vowel harmony, focusing on the harmony that occurs in ideophones. In section 3, I will introduce some of the previous cases where gradient phonotactics were claimed to be important. Then, we will see how the values of the expected frequencies (E) and O/E are measured. In section 4, I will show how the test words were gathered, and then present the O/E values for the vowel pairs in monomorphemic native roots in Korean. It will be shown that some of the vowel pairs have O/E values lower than 1, and these underrepresented pairs correspond to the vowel sequences that are not predicted to occur by the harmony. Finally, section 5 will serve as a conclusion and summary of the paper.

2. Basic patterns of vowel harmony

This section reviews the basic patterns of vowel harmony in Korean, focusing on the harmony patterns that occur in ideophones. Ideophones in present-day Korean are still subject to severe restrictions in the types of vowels that are allowed to occur together. Assuming eight underlying vowels in Korean, we can divide the vowels into two harmonic groups, /i, ɨ, u, e, ə/ vs. /æ, a, o/, depending on whether the vowels cooccur with one another. As for the harmonic feature distinguishing these two groups, I will adopt [Advanced Tongue Root (ATR)] (/i, ɨ, u, e, ə/ are [+ATR], and /æ, a, o/ are [-ATR]), following Y-S. Kim (1984), J-S. Lee (1992), and Y-S. Lee (1993).

1 Only Kawahara's (2007) study is based on Japanese rap rhymes, not the OCP-related phenomena. He showed that rhymability is gradient and can be measured by the O/E values.

(1) Harmonic groups

i	ɨ	u	[+ATR]
e	ə	o	
æ	a		[-ATR]

These vowels, unless they are neutral, cooccur only with another vowel from the same group. Based on the distribution of the vowels, we can identify the following patterns of the vowel harmony:[2]

(2) Basic harmony patterns

a. [-ATR] vowels cooccur only with another [-ATR] vowel:
allok 'colorful' k'æcak 'halfheartedly'

b. Nonhigh [+ATR] vowels cooccur with another nonhigh [+ATR] vowel:
k'ecək 'halfheartedly' t'eŋkəŋ 'at one stroke'

c. Initial high vowels cooccur with a [+ATR] vowel:
cilkəŋ 'chewing' pit'ul 'crookedly'
k'ɨt'cək 'nodding' pukɨl 'hubble-bubble'
chulləŋ 'splashing' siŋkɨl 'with a gentle smile'

d. Non-initial high vowels are neutral; they cooccur with a [+ATR] vowel or a [-ATR] vowel:
pusɨl, posɨl 'drizzling'
t'ekul, t'ækul 'rolling'
k'ecilək, k'æcilak 'halfheartedly'

What attracts our attention is the behavior of high vowels. If a high vowel is placed in the initial position, it acts just like another [+ATR] vowel (2c). But if it is placed in the medial position, it is neutral (transparent, more specifically) to the harmony (2d).

The types of vowel pairs predicted by the harmony patterns are as follows: First, the predicted vowel pairs in the initial two syllables of a root (referred to henceforce as V_1V_2) are as listed in (3), and the pairs that are not predicted to occur are as given in (4).

2 The harmonic feature is morphemic, and depending upon the specification of [ATR], the so-called 'dark' vs. 'light' ideophones are distinguished. The dark ideophones, which are made up of [-ATR] vowels with one or more optional neutral vowels, are 'heavier', 'more intensive' and 'bigger' in motion than their light counterparts, which consist of [+ATR] vowels. The vowel harmony results in interesting patterns of vowel alternation, which have been the subject of many previous studies (Kim-Renaud 1976, McCarty 1983, Y.-S. Kim 1984, Sohn 1987, J.-S. Lee 1992, Y.-S. Lee 1993). This paper focuses on vowel cooccurrence rather than alternation, and thus the alternation patterns will not be discussed here.

(3) Vowel pairs that are predicted to occur in V_1V_2:

a.	[-ATR]$_1$ – [-ATR]$_2$	cf. (2a), allok
b.	[-high, +ATR]$_1$– [-high, +ATR]$_2$	cf. (2b), k’ecək
c.	[+high, +ATR]$_1$ – [-high, +ATR]$_2$	cf. (2c), cilkəŋ
d.	[+high, +ATR]$_1$ – [+high, +ATR]$_2$	cf. (2d), pusɨl
	[-high, +ATR]$_1$ – [+high, +ATR]$_2$	cf. (2d), t’ekul
	[-ATR]$_1$– [+high, +ATR]$_2$	cf. (2d), posɨl, t’ækul

(4) Vowel pairs that are NOT predicted to occur in V_1V_2:

a. *[+high, +ATR]$_1$ – [-ATR]$_2$
b. *[-high, +ATR]$_1$ – [-ATR]$_2$
c. *[-ATR]$_1$ – [-high, +ATR]$_2$

Note that an initial [+ATR] vowel, whether it is high or nonhigh, is not followed by a [+ATR] vowel.[3] Also, if the initial vowel is [-ATR], the next vowel cannot be a nonhigh [+ATR] vowel.

These patterns of the harmony are also observed between vowels in the first and third syllables of a root (V_1V_3). This is because the vowel in the V_2 position is either identical to V_1 in the [ATR] specification or transparent to the harmony.

(5) Vowel pairs that are predicted to occur in V_1V_3:

a.	[-ATR]$_1$ – [-ATR]$_3$	eg. cʰa̱lkʰada̱k ‘with a click’
b.	[-high, +ATR]$_1$ – [-high, +ATR]$_3$	eg. cʰə̱lkʰədə̱k ‘with a click’
c.	[+high, +ATR]$_1$– [-high, +ATR]$_3$	eg. ci̱lpʰədə̱k ‘sloppily’
d.	[+high, +ATR]$_1$ – [+high, +ATR]$_3$	eg. pu̱cici̱k ‘with a rip’
	[-high, +ATR]$_1$ – [+high, +ATR]$_3$	eg. ə̱lisu̱ŋ ‘foolishly’
	[-ATR]$_1$– [+high, +ATR]$_3$	eg. kya̱ut’u̱ŋ ‘tilting’

The patterns of vowel cooccurrence in the V_1V_2 and V_1V_3 positions are presented in the following table, where shaded cells represent the impossible vowel pairs that are not predicted to occur.

3 One of the reviewers pointed out examples like *t’ukt’ak* ‘hammering’, *p’it’ak* ‘inclined’, and *hutatak* ‘sudden move’, where a [+high, +ATR] vowel is followed by a [-ATR] vowel, counter to (4a). There exist a small number of counterexamples like these, but the point of this section is to present the main patterns of vowel harmony, and so I will not address them further here.

(6) Predicted vowel pairs: V_1V_2, V_1V_3

		[+hi, +ATR]			[-hi, +ATR]		[-ATR]		
	$V_1\backslash V_2$ $V_1\backslash V_3$	i	ɨ	u	e	ə	æ	a	o
+high +ATR	i	√	√	√	√	√	(4a)		
	ɨ	√	√	√	√	√			
	u	√	√	√	√	√			
-high +ATR	e	√	√	√	√	√	(4b)		
	ə	√	√	√	√	√			
-ATR	æ	√	√	√	(4c)		√	√	√
	a	√	√	√			√	√	√
	o	√	√	√			√	√	√

Unlike a high vowel in the initial position where it triggers the harmony of [+ATR], high vowels in the medial V_2 position are neutral to the harmony. Thus, the patterns of vowel occurrence for the second and third syllables of a word (V_2V_3) are different from those for V_1V_2. Specifically, a [-ATR] vowel may appear after a medial [+high, +ATR] vowel because this medial high vowel is neutral and it may be preceded by another [-ATR] vowel. In other words, (4a), in addition to the pairs listed in (3), can be a possible vowel pair in the V_2V_3 position. The pairs of vowels that occur in V_2V_3 are presented below in (7), followed by the tabulation (8):

(7) Vowel pairs that are predicted to occur in V_2V_3:

a.	$[\text{-ATR}]_2$ – $[\text{-ATR}]_3$	eg. c^{h}alk^{h}adak 'with a click'
b,	$[\text{-high, +ATR}]_2$ – $[\text{-high, +ATR}]_3$	eg. c^{h}əlk^{h}ədək 'with a click'
c.	$[\text{+high, +ATR}]_2$ – $[\text{-high, +ATR}]_3$	eg. k'eciək 'halfheartedly'
d.	$[\text{+high, +ATR}]_2$ – $[\text{+high, +ATR}]_3$	eg. pucicik 'with a rip'
	$[\text{-high, +ATR}]_2$ – $[\text{+high, +ATR}]_3$	eg. hebəlc'uk 'wide open'
	$[\text{-ATR}]_2$ – $[\text{+high, +ATR}]_3$	eg. t'ækt'ækul 'rolling'
e.	$[\text{+high, +ATR}]_2$ – $[\text{-ATR}]_3$	eg. k'æcilak 'halfheartedly'

(8) Predicted vowel pairs: V_2V_3

		[+hi, +ATR]			[-hi, +ATR]		[-ATR]		
	V_2\\V_3	i	ɨ	u	e	ə	æ	a	o
+high +ATR	i	√	√	√	√	√	√	√	√
	ɨ	√	√	√	√	√	√	√	√
	u	√	√	√	√	√	√	√	√
-high +ATR	e	√	√	√	√	√	(4b)		
	ə	√	√	√	√	√			
-ATR	æ	√	√	√	(4c)		√	√	√
	a	√	√	√			√	√	√
	o	√	√	√			√	√	√

These patterns of the vowel harmony are confirmed by actual vowel occurrences in 806 three-syllable ideophones extracted from the list of ideophones provided by NIKL (2000). If we divide the vowels into three groups, [+high, +ATR], [-high, +ATR], [-ATR], there 27 possible types of vowel cooccurrence patterns for the three-syllable ideophones. As we see below, the patterns that appear in these three-syllable ideophones all fall into the predicted types except for the three examples of High-[-ATR]-[-ATR], *hutatak*/*hutaktak* 'quickly', *utaŋt^{h}aŋ* 'with a thump'.

(9) Actual vowel cooccurrence in three-syllable ideophones[4]

Predicted	V_1	V_2	V_3	Occurrences	Examples
√	[+ATR]	[+ATR]	[+ATR]	91	c^{h}əlk^{h}ədək
	[+ATR]	[+ATR]	[-ATR]		
√	[+ATR]	[+ATR]	Hi	9	hebəlc'uk
	[+ATR]	[-ATR]	[+ATR]		
	[+ATR]	[-ATR]	[-ATR]		
	[+ATR]	[-ATR]	Hi		
√	[+ATR]	Hi	[+ATR]	82	k'eciłək
	[+ATR]	Hi	[-ATR]		
√	[+ATR]	Hi	Hi	18	əlisuŋ
	[-ATR]	[+ATR]	[+ATR]		
	[-ATR]	[+ATR]	[-ATR]		
	[-ATR]	[+ATR]	Hi		
	[-ATR]	[-ATR]	[+ATR]		

4 In this table, [+ATR] represents only [-high, +ATR] vowels; [+high, +ATR] is referred to here simply as 'Hi'. To avoid unnecessary confusion about whether a form is an ideophone or not, I considered here only the forms whose reduplicated variants were also provided.

√	[-ATR]	[-ATR]	[-ATR]	145	c^h^alk^h^adak
√	[-ATR]	[-ATR]	Hi	11	t'ækt'ækul
	[-ATR]	Hi	[+ATR]		
√	[-ATR]	Hi	[-ATR]	165	k'æcilak
√	[-ATR]	Hi	Hi	88	kyaut'uŋ
√	Hi	[+ATR]	[+ATR]	21	cilp^h^ədək
	Hi	[+ATR]	[-ATR]		
√	Hi	[+ATR]	Hi	3	k^h^uŋtək^h^uŋ
	Hi	[-ATR]	[+ATR]		
	Hi	[-ATR]	[-ATR]	3	
	Hi	[-ATR]	Hi		
√	Hi	Hi	[+ATR]	91	mik'ɨtəŋ
	Hi	Hi	[-ATR]		
√	Hi	Hi	Hi	79	pucicik

So far we have seen the basic patterns of vowel cooccurrence predicted to hold in the first three syllables of a word. In the next sections we will examine the actual phonotactic patterns that appear in monomorphemic roots in native Korean words.

3. Gradient phonotactics and O/E

Recently, evidence has been mounting to show that our phonotactic knowledge is gradient. One good example of gradient phonotactics is the Obligatory Contour Principle (OCP; McCarthy 1986, Yip 1988), which prohibits adjacent identical elements. Previous studies show that the OCP holds in the following examples in a gradient manner to exclude adjacent consonants sharing an identical place of articulation.

(10) a. Among consonants in triliteral verbal roots in Arabic (Pierrehumbert 1993, Frisch et al. 2004)
b. Between onset and coda consonants in monomorphemic monosyllable words in English (Berkley 1994)
c. Between adjacent onset consonants in monomorphemic Yamato Japanese (Kawahara et al. 2006)
d. Between adjacent onset consonants in (V)CVCV and (V) CVCVC roots in Muna (Coetzee & Pater 2008)

What is important here is that although the OCP disallows neighboring homorganic consonants, such prohibition is not absolute but rather gradient. In Arabic, for example, the OCP is more likely to hold between coronal consonants if they share the feature [sonorant] and they are strictly adjacent to each other.

As for the formula used to measure phonotactic gradience, Pierrehumbert (1993) proposed O/E, the ratio of observed frequency to expected frequency. The observed frequency of a pair is simply the number of occurrences of the pair in the test data. The expected frequency of a pair is the value that would be expected if the two elements of the pair are combined at random. The expected frequency is calculated by the formula $E(x, y) = P(x)*P(y)*N$ (where N is the total number of pairs).

The analogy provided by Kawahara (2007) would be helpful to understand the concept of the expected frequency. Suppose that there are 100 cars in our sample, of which 10 are red and 60 are owned by a male. It is expected that the number of cars that are both red and owned by a male is 6 (= 10/100 * 60/100 * 100). This is the expected frequency of the red-male pair. If the actual number of red cars that are owned by a male is greater than 6, the O/E value is greater than 1, which we refer to as an ‘overrepresented’ case. If the actual number of male-owned red cars is smaller than 6, the O/E value is less than 1, and we call cases like this ‘underrepresented’. In this way, “O/E values provide a measure of ... the frequency of two elements to cooccur” (Kawahara 2007:118).

How the O/E values are obtained in actual examples and how they are relevant to phonotactics are well illustrated by the distribution of onset and coda consonants in monomorphemic monosyllable English words, studied by Berkeley (1994). She examined the distribution of the consonant pairs in monosyllables separated by exactly one segment (as in *pip* and *king*) and presented the results as follows:

(11)

C1\C2	labial	coronal	dorsal	Total
labial	26	256	42	324
coronal	204	428	160	792
dorsal	22	110	10	142
Total	252	794	212	1258

These are the observed frequencies of onset and coda consonants in the same monosyllable words. The expected frequencies are calculated from these observed frequencies. The expected frequency of the labial-labial sequence, for example, is obtained by (*the probability of onset being labial*) * (*the probability of coda being labial*) * (*the total number of the pairs*), that is (324/1258) * (252/1258) * (1258) = 64.9. Similarly, the expected frequency of the coronal-labial pair is (792/1258) * (252/1258) * (1258) = 158.7. The expected frequencies of all the consonant pairs computed this way are given below:

(12)

C1\C2	labial	coronal	dorsal
labial	64.9	204.5	54.6
coronal	158.7	499.9	133.5
dorsal	28.4	89.6	23.9

The O/E values are acquired by dividing the observed frequencies by the expected frequencies. The O/E values of the labial-labial and coronal-coronal sequences, for example, are 26/64.9 = 0.40 and 204/158.7 = 1.29, respectively. The entire O/E values are as follow:

(13)

C1\C2	labial	coronal	dorsal
labial	0.40	1.25	0.77
coronal	1.29	0.86	1.20
dorsal	0.77	1.23	0.42

Consonant pairs whose O/E is greater than 1 are overrepresented, and those whose O/E is less than 1 are underrepresented. From the above table of the O/E values, we can see that the labial-labial and dorsal-dorsal sequences are the two most severely underrepresented consonant pairs.[5]

4. Gradient vowel cooccurrence in Korean

This section presents evidence for gradient vowel cooccurrence restrictions that hold in monomorphemic native roots in Korean. The patterns of vowel cooccurrence will be examined in the first three syllables of a root. In particular, we will consider vowel cooccurrence as they appear in the first and second syllables of a root (V_1V_2), the first and third syllables of a root (V_1V_3), and the second and third syllables of a root (V_2V_3).

The test words for this study were gathered from *Pyo-jun-gug-eo-dai-sa-jeon* (NIKL 2001), where the headwords contain information about morpheme boundaries and word origins.[6] Using this information, I extracted 52,487 monomorphemic native roots from the headword list, excluding Sino-Korean, foreign borrowings and morphologically complex words.[7]

These roots were trimmed to 48,098 roots with at least two syllables to examine the patterns of vowel cooccurrence in the initial two syllables of a root (V_1V_2). As we saw in section 2, the three groups of vowels, [+high, +ATR], [-high, +ATR] and [-ATR], were relevant for the harmony, and thus, we counted the observed frequencies of the vowel pairs based on these three vowel groups. We then calculated the expected frequencies and the O/E values. The results are presented in the following tables, where the observed and expected frequencies and the O/E values are given in this order in each cell.

5 Berkley (1994) further examined the distribution of coronal-coronal sequences in relation to the OCP, and pointed out that the OCP holds more rigidly if the coronal consonants are of the same specification for [sonorant], just as in Arabic (McCarthy 1988, Pierrehumbert 1993).

6 In the headword list of *Pyo-jun-gug-eo-dai-sa-jeon*, morpheme boundaries are marked by hyphens (except for the verb/adjective ending *–da*, for which no boundary symbol was provided); and word origins are presented by specifying Chinese characters if the word is Sino-Korean, or by providing the word form in the source language if it is another type of foreign borrowings.

7 A Perl script written by the author was used to extract the test words.

(14) Vowel cooccurrence: V_1V_2

$V_1 \backslash V_2$	[+hi, +ATR] (i ɨ u)	[-hi, +ATR] (e ə)	[-ATR] (æ a o)	Total
[+high, +ATR] (i ɨ u)	O: 8093 E: 7484.79 O/E: 1.08	3290 2314.78 1.42	3446 5029.43 0.69	14829
[-high, +ATR] (e ə)	4980 4923.24 1.01	2652 1522.58 1.74	2122 3308.18 0.64	9754
[-ATR] (æ a o)	11204 11868.97 0.94	1566 3670.64 0.43	10745 7975.39 1.35	23515
Total	24277	7508	16313	48098

(N=48098, χ^2=4428.359, p < .001)

Shaded cells indicate the vowel pairs that are not predicted to occur according to the vowel harmony. What is important is that the O/E values of these impossible vowel pairs are significantly lower than 1 (μ=0.63, σ=0.18, t=-9.299, p<.001, df=20). These O/E values are contrasted with those of the possible vowel pairs that are predicted to occur, whose O/E values are significantly higher than 1 (μ=1.21, σ=0.37, t=3.598, p=.001, df=42). The difference of the O/E values indicates that the unpredicted vowel pairs are severely underrepresented, in contrast to the predicted pairs which are overrepresented in general.

To know about the vowel cooccurrence patterns exhibited in other two-syllable positions, V_2V_3 and V_1V_3, we further reduced the data set to 29,622 words with more than two syllables. The observed and expected frequencies, and the O/E values were obtained for the pairs of vowels that were categorized into three groups, [+high, +ATR], [-high, +ATR] and [-ATR]. Relevant results are given below:

(15) Vowel cooccurrence: V_1V_3

$V_1\backslash V_3$	[+hi, +ATR] (i ɨ u)	[-hi, +ATR] (e ə)	[-ATR] (æ a o)	Total
[+high, +ATR] (i ɨ u)	O: 4976 E: 4436.03 O/E: 1.12	1687 1463.11 1.15	2496 3259.85 0.77	9159
[-high, +ATR] (e ə)	2634 2800.91 0.94	1605 923.81 1.74	1544 2058.27 0.75	5783
[-ATR] (æ a o)	6737 7110.05 0.95	1440 2345.07 0.61	6503 5224.87 1.24	14680
Total	14347	4732	10543	29622

(N=29622, χ^2=1400.332, p < .001)

As in the case of V_1V_2, the average O/E value of the three vowel pairs in the shaded cells is less than 1 (μ=0.74, σ=0.17, t=-7.024, p<.001, df=20), but that of the other vowel pairs is higher than 1 (μ=1.16, σ=0.45, t=2.284, p<.05, df=42). Note that this is exactly as predicted by the vowel harmony that we discussed in section 2, although the result suggests that the effect of the harmony seems to weaken on nonadjacent vowels.[8]

Vowel cooccurrence patterns in V_2V_3 are different from those of V_1V_2 and V_1V_3. As we noted earlier, a high vowel is neutral to the harmony, and thus a [-ATR] vowel is allowed to occur in V_3 even after a [+high, +ATR] vowel if the first vowel is also [-ATR]. The observed/expected frequencies and the O/E values for the vowel pairs in the V_2V_3 position were tabulated as follows:

8 McCarthy (1986) and Pierrehumbert (1993) also report that phonotactic constraints on Arabic trilateral roots weaken over nonadjacent consonants.

(16) Vowel cooccurrence: V_2V_3

$V_2 \backslash V_3$	[+hi, +ATR] (i ɨ u)	[-hi, +ATR] (e ə)	[-ATR] (æ a o)	Total
[+high, +ATR] (i ɨ u)	O: 6173 E: 7204.98 O/E: 0.86	3056 2376.38 1.29	5647 5394.64 1.07	14876
[-high, +ATR] (e ə)	2559 2211.48 1.16	991 729.40 1.36	1016 1625.12 0.63	4566
[-ATR] (æ a o)	5615 4930.54 1.14	685 1626.22 0.42	3880 3623.24 1.07	10180
Total	14347	4732	10543	29622

(N=29622, χ^2=1601.241, $p < .001$)

Of notice is the occurrence of the sequence [+high, +ATR] followed by [-ATR]. Unlike the previous two cases, this sequence is not underrepresented in the V_2V_3 position, which is exactly as predicted by the harmony. The vowel pairs that are not predicted to occur by the harmony are still significantly below 1 in their average O/E value (μ=0.71, σ=0.20, t=-5.087, p<.001, df=11).

To summarize, vowel pairs in V_1V_2 or V_1V_3 positions are underrepresented exactly as predicted by the vowel harmony, although the degree of underrepresentation in the latter is less severe than the first. Three types of vowel pairs, [+high, +ATR] or [-high, +ATR] followed by [-ATR], and [-ATR] followed by [-high, +ATR], are the underrepresented pairs with low O/E values. These are the sequences that are not predicted to occur according to the vowel harmony. Vowel pairs in V_2V_3 position show a different pattern: [+high, +ATR] followed by [-ATR] is not underrepresented, which also conforms to the general patterns of the vowel harmony.

Now we will compare the patterns of root-internal vowel cooccurrence with the actual occurrence of vowel harmony in ideophones. The actual frequencies of ideophones were obtained from the ideophone list compiled by NIKL (2000). Among the listed words, I considered only those whose reduplicated forms were also provided and those whose base has at least two syllables. There were 3,191 ideophones that met these conditions. I examined the vowel cooccurrence patterns that appear in the initial two syllables of these words. The observed and expected frequences, and the O/E values are as follows:

(17) Observed and expected frequencies, O/E in ideophones: V_1V_2

V_1\V_2	i	ɨ	u	e	ə	æ	a	o	Total
i	31 34.16 0.91	102 72.01 1.42	72 40.89 1.76	0 1.89 0	53 49.59 1.07	0 2.05 0	4 48.85 0.08	0 12.56 0	262
ɨ	27 21.38 1.26	54 45.07 1.20	16 25.59 0.63	6 1.18 5.08	57 31.04 1.84	1 1.28 0.78	3 30.58 0.10	0 7.86 0	164
u	80 65.18 1.23	119 137.42 0.87	116 78.03 1.49	6 3.60 1.66	169 94.64 1.79	0 3.92 0	10 93.23 0.11	0 23.97 0	500
e	4 7.30 0.55	12 15.39 0.78	10 8.74 1.14	2 0.40 4.95	28 10.60 2.64	0 0.44 0	0 10.44 0	0 2.69 0	56
ə	67 82.39 0.81	151 173.70 0.87	111 98.63 1.26	9 4.56 4.95	294 119.63 2.46	0 4.95 0	0 117.84 0	0 30.30 0	632
æ	22 27.12 0.81	74 57.17 1.29	41 32.46 1.26	0 1.50 0	1 39.37 0.03	2 1.63 1.23	47 38.78 1.21	21 9.97 2.11	208
a	113 119.42 0.95	264 251.75 1.05	96 142.95 0.67	0 6.60 0	2 173.38 0.01	15 7.18 2.09	374 170.80 2.19	52 43.92 1.18	916
o	72 59.06 1.22	101 124.50 0.81	36 70.70 0.51	0 3.27 0	0 85.74 0	7 3.55 1.97	157 84.47 1.86	80 21.72 3.68	453
Total	416	877	498	23	604	25	595	153	3191

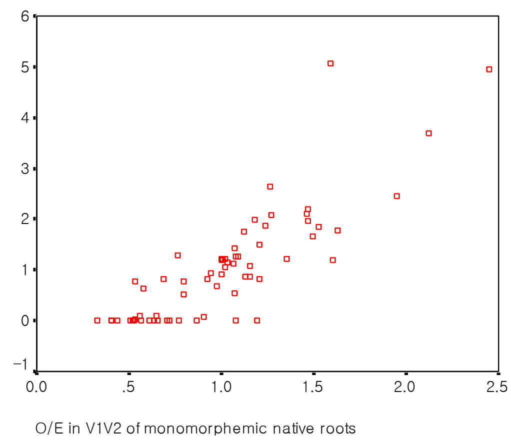
What we can instantly see from this table is that the actual harmony patterns which occur in ideophones are similar to the patterns predicted by the vowel harmony we discussed earlier in section 2. The O/E values of the vowel pairs that are not predicted to occur are close to zero (μ=0.05). These unpredicted vowel pairs are in fact so underrepresented that we can take the occurrence of the vowel harmony as categorical. The O/E values of the predicted vowel pairs are significantly higher than 1 (μ=1.53, σ =1.02, t=3.364, $p < .01$, df=42).

Now thses O/E values of ideophones are compared to the O/E values of the vowel pairs in the first two syllables of the monomorphemic native roots[9]. This is to verify whether the patterns of vowel coocurrence in roots

9 The O/E values of all the vowel pairs that appear in the V_1V_2 position of monomorphemic native roots are given in Appendix 1.

coincide with the harmony patterns that are actually found in ideophones. The scatter plot below represents the correlation between the O/E values of root-internal vowel cooccurrence and the vowel pairs that actually occur in ideophones.

(18) Vowel cooccurrence and actual vowel harmony in V_1V_2 position



It turned out that the O/E values of the vowel pairs actually occurring in ideophones are significantly correlated with the O/E values of the vowel pairs that occur in the first two syllables of monomorphemic native roots (r=.841, p<.001). This indicates strongly that vowel cooccurrence in roots reflects the actual harmony patterns.[10]

5. Conclusion

In this paper, we examined vowel cooccurrence restrictions that hold in monomorphemic native roots in Korean. For this purpose, multisyllable words listed as monomorphemic and native in *Pyo-jun-gug-eo-dai-sa-jeon* were extracted. Since vowel cooccurrence restrictions are often the result of vowel harmony, root-internal cooccurrence patterns were examined to see how much they reflect the general patterns of the harmony. Assuming gradient phonotactics that can be measured by O/E, we obtained the O/E values of the vowel pairs that can appear in the first three syllables of monomorphemic roots, V_1V_2, V_1V_3 and V_2V_3, and compared the O/E

---

10 Vowel cooccurrence in the V_1V_2 position of a root was also compared to the actual occurrences of vowel pairs in ideophones provided by Park (1993). We considered the vowels occurring in the first two syllables of 2,364 tokens of ideophones he examined. It was found that root-internal vowel cooccurrence in the V_1V_2 position also correlates significantly with vowel occurrence in ideophones that Park examined (r=.732, p<.001).

values to the patterns predicted by the vowel harmony. It was found that the O/E values of the vowel pairs which were not predicted to occur were substantially lower than 1 in monomorphemic native roots. It was further pointed out that the underrepresentation of these unpredicted pairs was statistically significant, and that this underrepresentation is uniformly hold throughout the unpredicted pairs, although their occurrences differ depending on the position where the vowel pairs occur. As a last proof, we compared root-internal vowel cooccurrence to the actual patterns of the vowel harmony that occur in ideophones, and confirmed that the patterns of root-internal vowel cooccurrence significantly correlate with those of the actually found harmony patterns.

Appendix A. Observed & expected frequencies, and O/E in V_1V_2

$V_1\backslash V_2$	i	ɨ	u	e	ə	æ	a	o
i	1103 1106.11 1.00	882 825.25 1.07	1068 951.72 1.12	213 178.26 1.19	825 713.38 1.16	286 396.77 0.72	1021 1132.00 0.90	314 408.53 0.77
ɨ	528 483.15 1.09	577 360.47 1.60	239 415.71 0.57	124 77.86 1.59	475 311.60 1.52	138 173.31 0.80	321 494.47 0.65	93 178.44 0.52
u	1285 1282.33 1.00	1082 956.72 1.13	1329 1103.35 1.20	308 206.65 1.49	1345 827.03 1.63	291 459.98 0.63	734 1312.34 0.56	248 473.61 0.52
e	357 334.43 1.07	132 249.51 0.53	297 287.75 1.03	132 53.89 2.45	273 215.69 1.27	61 119.96 0.51	368 342.26 1.08	107 123.52 0.87
ə	1431 1554.40 0.92	1337 1159.71 1.15	1426 1337.44 1.07	296 250.50 1.18	1951 1002.50 1.95	365 557.57 0.65	969 1590.78 0.61	252 574.10 0.44
æ	957 794.92 1.20	453 593.07 0.76	735 683.97 1.07	52 128.11 0.41	272 512.68 0.53	385 285.14 1.35	821 813.52 1.01	430 293.59 1.46
a	2155 2286.38 0.94	1734 1705.83 1.02	1913 1967.26 0.97	209 368.46 0.57	483 1474.59 0.33	1039 820.14 1.27	3432 2339.90 1.47	842 844.44 1.00
o	1498 1472.29 1.02	752 1098.45 0.68	1007 1266.8 0.79	167 237.27 0.70	383 949.55 0.40	776 528.12 1.47	1866 1506.75 1.24	1154 543.77 2.12

Appendix B. Observed frequencies in V_1V_3 and V_2V_3

$V_1 \backslash V_3$	i	ɨ	u	e	ə	æ	a	o
i	1244	280	328	76	512	263	531	136
ɨ	417	265	118	53	224	113	250	67
u	1249	331	744	119	703	377	587	172
e	317	55	89	69	135	60	187	40
ə	1385	393	395	127	1274	349	717	191
æ	682	193	235	19	239	403	497	131
a	2066	611	655	176	591	573	2305	453
o	1630	313	352	119	296	451	994	696

$V_2 \backslash V_3$	i	ɨ	u	e	ə	æ	a	o
i	1050	320	550	118	652	457	1043	407
ɨ	733	937	330	236	979	466	1104	272
u	1369	311	573	208	863	597	1077	224
e	545	26	52	12	33	19	94	28
ə	1200	300	436	80	866	254	448	173
æ	1226	38	157	11	94	64	208	104
a	2112	406	695	59	370	485	1661	387
o	7550	103	123	34	117	247	433	291

REFERENCES

BERKLEY, DEBORAH MILAM. 1994. Variability in Obligatory Contour Principle effects. *CLS* 30, 1-12.

COETZEE, ANDRIES AND JOE PATER. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26, 289-337.

FRISCH, STEFAN A. 1996. *Similarity and Frequency in Phonology*. PhD Dissertation. Northwestern University.

FRISCH, STEFAN A., JANET B. PIERREHUMBERT, AND MICHAEL B. BROE. 2004. Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* 22, 179-228.

KAWAHARA, SHIGETO. 2007. Half rhymes in Japanese rap lyrics and knowledge of similarity. *Journal of East Asian Linguistics* 16, 113-144.

KAWAHARA, SHIGETO, HAJIME ONO, AND KIYOSHI SUDO. 2006. Consonant Cooccurrence Restrictions in Yamato Japanese. *Japanese/Korean Linguistics* 14, 27-38.

KIM, CHIN-WU. 1978. "Diagonal" harmony?: Some implications for historical phonology. In Fisiak, J. (ed.) *Recent Developments in*

Historical Phonology, 221-235. The Hague: Mouton.
KIM, YOUNG-SEOK. 1984. *Aspects of Korean Morphology*. PhD Dissertation. University of Texas, Austin.
KIM-RENAUD, YOUNG-KEY. 1976. Semantic features in phonology: evidence from vowel harmony in Korean. *CLS* 12, 397-412.
LEE, BYUNG-GUN. 1985. Some tangible properties of vowel harmony. *In-mun-non-chong* 15, 4-34. Institute of humanities, Seoul National University.
LEE, JIN-SEONG. 1992. *Phonology and Sound Symbolism of Korean Ideophones*. PhD Dissertation. Indiana University, Bloomington.
LEE, KI-MOON. 1972. *Gug-eo-eum-un-sa Yeon-gu* [A study of Korean historical phonology]. Han-gug-mun-wha-yeon-gu-so, Seoul National University.
LEE, SEUNG-NYEONG. 1947. Mo-eum-cho-wha Yeon-gu [A study of vowel harmony]. *Chin-tan-hak-po* 16, 1-109.
LEE, YONG-SUNG. 1993. *Topics in the Vowel Phonology of Korean*. PhD Dissertation. Indiana University, Bloomington.
MCCARTHY, JOHN J. 1986. OCP effects: gemination and antigemination. *Linguistic Inquiry* 17, 207-63.
NATIONAL INSTITUTE OF THE KOREAN LANGUAGE (ed.) 2000. Pyo-jun-gug-eo-dai-sa-jeon Pyeon-can-yong Ui-seong-ui-tae-eo-mok-lok [The list of ideophones compiled for the publication of the Grand Dictionary of Standard Korean]. Available at http://www.korean.go.kr.
_______________. (ed.) 2001. *Pyo-jun-gug-eo-dai-sa-jeon* [*The Grand Dictionary of Standard Korean*]. CD-Rom version. Doo-san-dong-A, Seoul.
PARK, CHANG-WON. 1993. Hyeon-tae Gug-eo Ui-seong Ui-tae-eo-ui Hyeong-tae-wa Eum-un [Morphology and phonology of ideophones in modern Korean]. *Sae-gug-eo-saeng-whal* 3.2, 16-53.
PIERREHUMBERT, JANET. 1993. Dissimilarity in the Arabic verbal roots. *NELS* 23, 367-381.
SOHN, HYANG-SOOK. 1987. *Underspecification in Korean Phonology*. Ph. D dissertation. University of Illinois, Urbana Champaign.
YIP, MOIRA. 1988. The Obligatory Contour Principle and phonological rules: a loss of identity. *Linguistic Inquiry*:65-100.

Sung-Hoon Hong
Department of English Linguistics
Hankuk University of Foreign Studies
270 Imun-dong, Dongdaemun-gu, Seoul
Korea 130-791
e-mail: hongshoon@hufs.ac.kr

received: June 29, 2010
accepted: August 7, 2010