Neutral vowels in sound-symbolic vowel harmony in Korean*

Darrell Larsen
Jeffrey Heinz
University of Delaware

The set of neutral vowels in the vowel-harmony pattern found in Modern Korean sound-symbolic words is traditionally considered to consist of two members: [i] and [ɨ]. Some authors have tentatively put forward the hypothesis that [u] and [y] are also neutral. A corpus study on the phonotactics of Korean sound-symbolic vowel harmony provides quantitative support both for the traditional under-standing of [i] and [ɨ] as neutral vowels and for the tentative hypothesis that [u] and [y] are also neutral. Additionally, the corpus study reveals that vowels are neutral to different degrees. The vowel [a] is shown to be partially neutral, but unlike the other neutral vowels which are transparent, [a] is partially opaque. The distribution of diphthongs is also studied.

1 Introduction

Korean exhibits two types of vowel harmony for distinct classes of words (Kim-Renaud 1976, Park 1990, Sohn 1999, *inter alia*): verbs and sound-symbolic words. The latter class includes a set of neutral vowels which have a dual functionality, aligning with one vowel-harmony type (the so-called DARK vowels) when occurring in the initial syllable and behaving as transparent vowels in non-initial syllables (Kim-Renaud 1976, Lee 1984, Sohn 1986, Park 1990, Lee 1992, Cho 1994, *inter alia*). This article makes use of a corpus of sound-symbolic stems in Korean to examine the strength of vowel harmony in sound-symbolic words from a phonotactic perspective, with special emphasis on the class of neutral vowels. The corpus study confirms the main results of earlier traditional analyses, but also shows the pattern is much more nuanced than earlier thought. In this way, this paper's findings are consistent with the general conclusions of other corpus studies of vowel harmony (Hayes & Londe 2006, Goldsmith

& Riggle 2012) that the patterns themselves are more intricate than originally assumed.

In particular, the neutrality of the traditional class of neutral vowels, [i] and [ɨ], is verified, and we show that the remaining high vowels in Korean, [u] and [y], which have traditionally been assigned to the class of dark vowels, are at least partially neutral. We also examine the vowel [a] and show that it too appears to be partially neutral. Unlike the other neutral vowels, however, [a] is better classified as opaque than transparent. Second, we examine the behaviour of diphthongs in sound-symbolic harmony and show that they pattern with their nuclei when occurring in initial syllables but not in other syllables.

This article is organised as follows. §2 discusses the vowel-harmony system in Korean, gives a brief historical overview and provides an introduction to the data in the corpus study: the sound-symbolic words of Korean. It also compares the Korean vowel-harmony pattern with the well-known patterns of Finnish and Hungarian. The corpus is described in §3 and the analysis and results pertaining to monophthongs in §4. §5 examines the behaviour of diphthongs in sound-symbolic stems. §6 provides a general discussion of the findings. Finally, we conclude in §7.

2 Overview of vowel harmony

2.1 Vowel harmony in Korean

A shared trait of Altaic languages such as Korean is the presence of vowel harmony (Sohn 1999). In Middle Korean, the classes of vowels – /o a ɔ/ belonging to the class of LIGHT vowels and /u ɨ i ə/ belonging to the class of DARK vowels – did not co-occur (e.g. Park 1990, Sohn 1999).[1] The harmony occurred within words and extended to particles (i.e. postpositions) and predicate suffixes (Sohn 1999). In Modern Korean, the prevalence of vowel harmony has been greatly reduced, with the breakdown beginning around the middle of the fifteenth century (Park 1990). This is largely due to the large influx of loanwords, mainly from Chinese, which did not obey the harmony process, as well as to a vowel shift causing the light /ɔ/ to become /o/, /a/ or the dark vowel /ɨ/ (Park 1990, Sohn 1999).

Despite these changes, Modern Korean continues to exhibit vowel harmony. There are two distinct forms of vowel harmony today: verbal suffix harmony and sound-symbolic harmony. This paper only provides a corpus study of sound-symbolic harmony, but, for completeness, both forms are reviewed here.

In verbal suffix harmony, verb stems end in either [-ʌ] or [-a],[2] and there are two past tense allomorphs [-ʌs'] and [-as']. The choice of [ʌ] or

---

[1] Park (1990) uses [ʌ] instead of our [ə], and Sohn (1999) uses [ʌ] instead of our [ɔ]. Sohn (1999: 48) calls [i] a neutral vowel in Late Middle Korean, while Park (1990: 470) states that it was dark in initial syllables and neutral elsewhere.

[2] The form [-ʌ] or [-a] is often (though not always) required to attach to the verb root before other suffixes can attach; however, it may appear alone on

[a] in these forms depends on the final vowel in the root. Verb roots whose final vowel is [o] or [a] take suffixes beginning with [a]. Verb roots with any other final vowel take suffixes beginning with [ʌ]. Examples are given in (1).[3]

(1) cap-a 'grab'   kipʰ-ʌ 'be deep'
    coh-a 'like, be good'   cuk-ʌ 'die'
    ka-a 'go'   se-ʌ 'count, be strong'
    po-a 'see'   mʌk-ʌ 'eat'

One interesting aspect of Korean vowel harmony is that the harmonic groupings do not fall into traditional natural classes. The vowel system of Korean (ignoring diphthongs) is given in (2), which shows the division between the two verbal suffix harmony classes.[4]

(2) *Verbal suffix harmony classes in Korean*

	front	front round	central	back (round)	
high	i	y	ɨ	u	dark
mid	e	ø	ʌ	o	light
low	æ		a		

In part because vowels do not divide into natural classes for vowel harmony in Korean, the terms DARK and LIGHT are often used, and we will adopt these terms in this paper.[5] In the case of verbal suffix harmony, the terms carry little meaning; however, as will be explained below, they

---

    the verb for various uses as well. The term stem is, therefore, used for simplicity only.

[3] Unless placed in / /, our transcription throughout is neither phonetic nor phonemic; rather, it is obtained via a mostly 1-to-1 transliteration from Korean spelling to IPA symbols (2-to-1 in the case of [y] and [ø], and 1-to-0 in the case of the null onset symbol). This results in a largely phonemic transcription, with some exceptions in the case of sound-symbolic forms. Many sound-symbolic words come in sets of two or more, some involving ablaut or the altering of consonants in some way, and in these words the Korean spelling and therefore our IPA transcription closely reflect the actual pronunciations rather than the abstract, phonemic form (but we will discuss a potential confound in §4.2.4). Since we are concerned with the phonotactics of vowel harmony, this is the exact level of detail we need.

[4] Although we include the front round vowels /y/ and /ø/ in our chart and discussion, in line with much of the previous research on Korean vowel harmony, modern-day younger South Korean speakers of the standard spoken (i.e. Seoul) dialect pronounce these as the diphthongs [ɥi]/[wi] and [we] respectively. However, not all researchers include these vowels. Lee (1992), for example, does not include these two vowels in her research on sound-symbolic words, because of their modern pronunciation. An anonymous reviewer also points out that the diphthongs [ɥi]/[wi] often simplify to [u] in word-initial position or [i] in non-initial position. Finally, we note that the diphthong /we/ also exists as a separate phoneme from /ø/, at least historically.

[5] The terms 'dark' and 'bright', or 'yin' and 'yang', are also commonly used by various authors to refer to the same two vowel-harmony classes.

describe the connotation the vowels add when used in sound-symbolic words.

The second type of vowel harmony is called sound-symbolic harmony, and occurs in sound-symbolic words. It is this type that is the subject of this paper. Sound-symbolic words include onomatopoeic words, in addition to words intended to vocally express some sensory perception (taste, sound, colour, action, etc.). Some examples are provided below. As will be observed, direct translation into English is often awkward, if possible at all.

(3) culʌŋ ~ colaŋ   'in clusters (of grapes)'
    pʌlkʌh ~ palkah   'red'
    pukɨl ~ pokɨl   'simmer(ingly), boil(ingly)'
    piŋkɨl ~ pæŋkɨl ~ pʌŋkɨl ~ paŋkɨl   'with a (broad) smile'

As is exemplified in (3), sound-symbolic words often come in phonetically similar sets. These sets are frequently composed of a pair of words in which one member contains light vowels while the other contains dark vowels. In sound-symbolic harmony, however, the sets of vowels which belong to each class differ from the sets in verbal suffix harmony. The example below shows the distribution of vowels into the light and dark classes. The curved lines demonstrate the correlations that exist between dark and light vowels, which show up when vowels alternate in sound-symbolic word sets.

(4) *Sound-symbolic vowel-harmony correspondences in Korean*
    (based on Kim-Renaud 1976)

	front	front round	central	back (round)	
high	i	y	ɨ	u	dark
mid	e	ø	ʌ	o	light
low	æ		a		
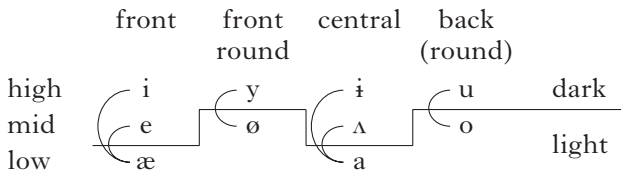
The light and dark classes of vowels in sound-symbolic words have slightly different connotations. Sohn (1999) provides the following list of nuances associated with each class.

(5) light: brightness, lightness, sharpness, quickness, smallness, thinness
    dark: darkness, heaviness, dullness, slowness, deepness, thickness

The examples in (6) show how the connotations differ in accordance with the harmonic class of vowels contained in them.[6]

---

[6] The last example in (3) contains two dark–light pairs: *piŋkɨl–pæŋkɨl* and *pʌŋkɨl–paŋkɨl*. The four forms have slightly different nuances.
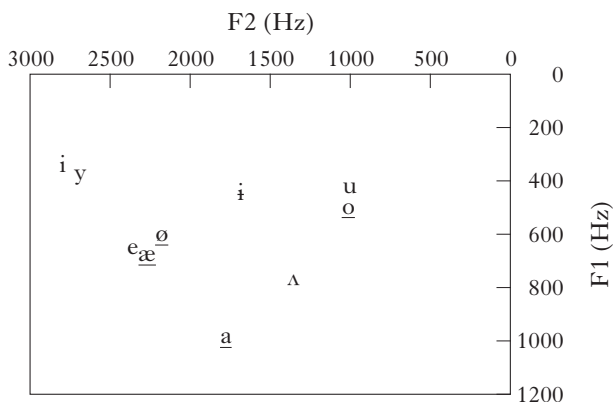
F2 (Hz)

Figure 1
Average formant frequencies for female speakers of Standard Korean
(data from Yang 1992). The light vowels are underlined.

(6) a. pʰuŋtʌŋ 'splash' (sound of large object hitting water)
    pʰoŋtaŋ 'splash' (sound of small object hitting water)
 b. pʌnc'ʌk 'sparkling, twinkling' (e.g. of a flash of light)
    panc'ak 'sparkling, twinkling' (e.g. of a star)

There are several thousand sound-symbolic morphemes in Korean, most of which have no direct English translations. These morphemes are generally used as adverbs. Additionally, they can often combine with certain verbs or auxiliaries, such as *hata* 'do' (often translated 'be' when combined with sound-symbolic words), or *kʌlita* 'keep doing/being'. The reader is referred to Lee (1992), Martin (1992) and Sohn (1999) for additional details of sound-symbolic words in Korean not covered here.

As in the case of verbal suffix harmony, it is not clear which features divide the vowels into the dark and light classes. Proposals typically treat [low] (e.g. McCarthy 1983, Sohn 1986) or [ATR] (e.g. Lee 1992, Cho 1994) as the distinguishing feature. Figure 1 shows the F1 and F2 values for Korean monophthongs, demonstrating that these phonetic dimensions do not distinguish the two classes. The figure summarises data from female speakers of Standard Korean presented in Yang (1992).

Returning to the classification of vowels, the high vowels [i] and [ɨ], given as dark in (6), actually behave as both dark and transparent (e.g. Kim-Renaud 1976, Lee 1984, Sohn 1986, Lee 1992, Cho 1994). In the initial syllable they behave as dark vowels, whereas in non-initial syllables they behave as transparent vowels (Lee 1984, Sohn 1986, Cho 1994). This dual functionality of neutral vowels is not unusual cross-linguistically, as neutral vowels in other languages often have a 'default' status. However, §2.2 below highlights important differences

between this dual functionality of neutral vowels in Korean and neutral vowels in the well-studied patterns in Finnish and Hungarian.

Some authors (e.g. Lee 1984, Lee 1992, Cho 1994) also claim that [u] behaves like a neutral vowel in some forms, but not to the same degree as the non-round high vowels. Lee (1984: n. 8) further suggests that [y] sometimes behaves as a neutral vowel in Modern Korean. We will later show that the corpus supports the hypothesis that [u] and [y] are at least partially neutral vowels; however, when conducting the corpus study below, we will group only the vowels [i] and [ɨ] together into the class of neutral vowels, while considering [y] and [u] individually.

In terms of vowel alternations, the dual functionality of neutral vowels means that in initial syllables neutral vowels undergo harmony alternations but not in non-initial position; thus, *k'icʌk-k'icʌk* 'scribbling' alternates with *k'æcak-k'æcak* 'scribbling' (of a different type). In non-initial position, neutral vowels do not undergo or trigger harmony themselves, but instead allow vowel harmony to spread over them. For example, *pusilʌk* 'rustling' becomes *posilak* 'rustling' (of a different type).

Sound-symbolic morphemes can frequently be reduplicated (the class of colours being a notable exception), leading to a slight difference in connotation (see Sohn 1999). Examples of common reduplicated types are shown below in (7), in order of number of syllables. Sound-symbolic reduplicants regularly involve the reduplication of monosyllabic (a), disyllabic (b, c) or trisyllabic (d–g) stems. In the 'four'-syllable reduplicant in (h), the individual parts (i.e. reduplicated one-syllable forms) exist in the language as well; thus this is best considered a compound. In fact, each of the ten 'four'-syllable reduplicants found in our corpus appears to be formed from two reduplicated one-syllable forms.

(7) a. sal-sal   'gently, softly, slowly'
    b. culʌŋ-culʌŋ   'in clusters' (e.g. grapes)
    c. als'oŋ-tals'oŋ   'jumbled, obscure'
    d. ucik'ɨn-ucik'ɨn 'with a snap, crackling' (e.g. a twig)
    e. allak'uŋ-tallak'uŋ 'messy'
    f. c'alɨlik-c'alɨlik   sound of sucking water with difficulty (e.g. with
                            a straw)
    g. pulɨliŋ-pulɨliŋ 'with a burr (of a combustion engine)'
    h. cʰikcʰikpʰokpʰok 'chugga chugga' (e.g. train)

It is not uncommon for the first consonant of the initial morpheme of this type to be unpronounced but to appear on the reduplicated portion, exemplified in (c) and (e) (see Sohn 1999 for further details).

Although many instances of reduplication in Korean are sound-symbolic, not all reduplication is sound-symbolic. We provide a few examples of non-sound-symbolic reduplication below.

(8) hana-hana   'one by one, one at a time'
    kʌlɨm-kʌlɨm   'step by step'
    kʌmɨs-kʌmɨs   'dotted with black'

The remainder of this paper examines the phonotactic pattern of sound-symbolic vowel harmony within individual forms only. We will not examine vowel harmony as a pattern of alternation as it applies to morphemes to create semantically nuanced pairs alternating between light and dark vowels, as in (3) and (6). This differs from much of the literature on Korean vowel harmony, which focuses on these light–dark alternations, including the effects of neutral vowels on the spread of such alternations. Furthermore, we will only be concerned with sound-symbolic words which are found in reduplicants. Reasons for this will be discussed in §3.

2.2 Comparison with other languages

Certain aspects of the sound-symbolic vowel-harmony pattern discussed above resemble the well-known patterns of Finnish and Hungarian. Like Korean, these languages also exhibit three vowel classes, two which harmonise and one which acts as neutral and transparent. Closer inspection of the three patterns, however, reveals differences between the behaviour of the neutral vowels in each case.

First, consider Finnish (Karlsson 1999, Ringen & Heinämäki 1999, Goldsmith & Riggle 2012). Like Korean, Finnish divides vowels into three classes: the neutral (N) vowels include [i] and [e]; the front (F) vowels include [y] (<ü>), [ø] (<ö>) and [æ] (<ä>); and the set of back (B) vowels contains [u], [o] and [ɑ]. Vowel harmony in Finnish both occurs within a morpheme and spreads onto suffixes, where the initial vowel of a root determines whether the subsequent vowels, including suffix vowels, will be front or back, as in (9) (from Karlsson 1999).

(9) BB   *talo-ssa*   'in the house'
    FF   *kylä-ssä*   'in the village'

When the initial suffix is a neutral vowel, it does not determine whether front or back harmony should be applied; instead, if no back vowels follow, then front harmony is applied, but if a back vowel follows, back harmony is applied, as in (10). Moreover, Finnish neutral vowels are transparent. As a result, back vowels trigger back harmony on suffixes even if a neutral vowel intervenes (also shown in (10)).

(10) NBN   *Kekkose-lta*   'from Kekkonen'
     NN   *velje-ltä*   'from the brother'

While Finnish may initially seem to be the same as Korean, there is a crucial difference. In Korean, initial neutral vowels trigger dark harmony to the same extent as dark vowels, and light vowels do not follow them

more than they follow initial dark vowels (see Table I in §4 below). Only in non-initial syllables do neutral vowels behave neutrally. In Finnish, on the other hand, neutral vowels can be analysed as neutral regardless of their position. In initial syllables, Finnish neutral vowels do not behave like front vowels. Whereas initial neutral vowels allow back vowels to follow them, initial front vowels generally do not.

Hungarian also exhibits a well-known harmony process (see e.g. Vago 1974, 1976, Hayes & Londe 2006). Hungarian divides vowels into front (front rounded), back (back) and neutral (front unrounded) classes, as shown in the chart below, which is based on Hayes & Londe (2006).

(11) *Hungarian vowel distribution*

	front unrounded		front rounded		back	
high	[i] *i*	[iː] *í*	[y] *ü*	[yː] *ű*	[u] *u*	[uː] *ú*
mid		[eː] *é*	[ø] *ö*	[øː] *ő*	[o] *o*	[oː] *ó*
low	[ɛ] *e*				[ɔ] *a*	[aː] *á*

Unlike Finnish, where the initial vowel controls harmony, the final vowel of a stem controls the harmony on suffixes in Hungarian. This is demonstrated by the following examples with disharmonic stems, where a final back vowel chooses a back stem, and a final front rounded vowel chooses a front (rounded or unrounded) stem (from Hayes & Londe 2006).

(12) FB   *glükóz-nak*   'glucose-DAT'
     BF   *sofőr-nek*   'chauffeur-DAT'

Stems with all neutral vowels usually take suffixes with front vowels, although there are some ('a few dozen'; 2006: 62) which take back suffixes. Additionally, there are some suffixes containing three allomorphs, such as the allative case marker *-hoz* (B), *-hez* (N) and *-höz* (F), in which cases neutral-final stems take the neutral suffix (Rounds 2001). When a final neutral vowel is preceded by a back vowel, however, the pattern is more complicated (at least for suffixes lacking a neutral allomorph, which Hayes & Londe do not discuss).

Hungarian vowel harmony more closely resembles the pattern in Korean than Finnish. Taking the stem-initial position in Korean and the stem-final position in Hungarian to be the 'strong' position for vowel harmony, one could argue that neutral vowels in both languages have a default harmony in the strong position (unless harmony has spread). Hungarian neutral vowels in the strong position appear to differ from Korean neutral vowels in the strong position, in that Hungarian neutral vowels can be transparent. However, this apparent difference may be illusory. Harmony spreads rightward in both languages, and stem-initial vowels by definition cannot be preceded by another vowel. Therefore, in Korean the situation which could reveal whether Korean neutral vowels in the strong position could act as transparent never arises.

As discussed above, while neutral vowels in Finnish and Hungarian resemble those in Korean, there are important differences. On the other hand, as we show below, there is one important similarity between the Korean harmony pattern on the one hand and the Finnish and Hungarian patterns on the other. In each case, careful inspection shows that the patterns are more intricate, variable and gradient than originally anticipated. This was one of the findings both of Hayes & Londe (2006) for Hungarian and of Goldsmith & Riggle (2012) for Finnish. We return to this point in §6.

3 About the corpus

The corpus used in this study was based on data from a corpus made to aid the National Institute of the Korean Language in the continued development of the *Great Standard Korean dictionary* (our translation).[7] The corpus consists of approximately 29,000 sound-symbolic words, although the distributor warns that a small number of the words may not truly be sound-symbolic, and that not all sound-symbolic words are found in the corpus.[8] Many of the words are variants, often combined with different verbs, built on a single underlying sound-symbolic stem. For example, the initial two words of the file, *kakul-kakul* 'winding' and *kakul-kakul-hata* 'to be winding', are an adverbial form and a basic verb form respectively, based on the same root. In the present study, only one token of each sound-symbolic stem was needed, and these were extracted as described in the following paragraph. Moreover, although the corpus includes a number of semantically distinct homophonous forms distinguished by the inclusion of a number after them, they were treated as non-distinct for the purpose of this study. Thus, although *katak-katak* was listed twice, only one base form, *katak*, was kept.

In this study, we chose to look only at reduplicant sound-symbolic forms. As noted above, reduplication is heavily used in Korean sound-symbolic words, and reduplicated forms are generally sound-symbolic. All full reduplicants which had a base form of two and three syllables in length were extracted.[9] Only the (unreduplicated) base form was kept in

---

[7] The data is available at http://www.korean.go.kr/ and http://www.hangeul.pe.kr/symbol/words.htm. The .hwp file with the data at www.korean.go.kr can be difficult to find. The path from the homepage is as follows:
> 자료실 > 기타공개자료 > 제목 = 표준국어대사전 편찬용 의성의태어목록 (작성일 = 2000-09-18). Searches may have to be done manually rather than by using the search function.

[8] We checked the meanings of approximately every fiftieth reduplicant, in its reduplicated form, in the combined corpora employed here (including both stems with monophthongs and diphthongs), using the Korean search engine Naver's dictionary. 100% of the 85 checked forms were listed in this dictionary with a sound-symbolic meaning. Therefore, we find little reason to question the accuracy of the corpus.

[9] 'Full' reduplicants are those in which no change occurred from the base form to the reduplicated form. Forms like those exemplified in (7c) and (7e) were not extracted for the study.

each case. Reduplicants with one-syllable bases were not kept, because they give no insight into vowel harmony. Likewise, reduplicants with 'four'-syllable bases were not kept, because they appear to be compounds of reduplicants with one-syllable bases (see §2.1, especially (7)). Moreover, the data on this latter type of reduplicant was too sparse to be used, as only ten such forms were found in the corpus.

We split the corpus study into two parts. The first part focuses only on those forms containing the vowels regularly included in the light–dark harmony distinction shown in (4); that is, it focuses only on the monophthongs. Forms containing diphthongs such as [wæ], [ja], etc. are not considered, as they have not been discussed in previous literature, and, as we will show, they do not pattern in an entirely predictable or comparable way. In total, 3972 forms were found to match the criteria described above. 2992 of these forms are disyllabic, and 980 are trisyllabic. The second part of our study examines the diphthongs, consisting of 263 forms, including 184 disyllabic and 79 trisyllabic forms. We will use the abbreviations D, L and N for dark, light and neutral. The appendices summarise the corpus and provide counts of the vowel-pattern sequences extracted from the corpus.

A few notes about the limitations of the corpus are in order. First, not all of the reduplicants in the corpus are common, and Korean speakers are not likely to be familiar with every one of them. Second, the corpus does not contain all of the sound-symbolic reduplicants in Korean. In fact, not all reduplicants found in the literature cited in this article are found in the corpus. We did not add any of these missing forms to the corpus.

Third, the reduplicants in the corpus are not limited to forms from one dialect. In a random check of 85 forms in the corpus (cf. note 8), 28 (around 33%), were labelled as North Korean forms. Since North Korean sound-symbolic forms are subject to the same harmony constraints, the forms remain representative of the vowel-harmony pattern we are studying. An anonymous reviewer points out that round and unround vowels in standard North Korean are merging or have merged; specifically, /o/ and /ʌ/ are merging as /o/, while /ɨ/ and /u/ are merging as /u/. If the merger were complete, then the /ʌ o/ merger in particular would be a potential confound, since it appears to transform the dark /ʌ/ into light /o/. While the status of the merger is reported to be complete in the Pyeongyang region (Kwak 2003), it is not complete in all areas of North Korea (Kwak 2003, Ko 2010).

Further complicating the issue, Kwak (2003) states that the pronunciation of the merged /ʌ o/ pair is (using his notation) [ɔ], [ɵ̈] and in some areas [ʌ]. Although the phonetic forms would appear to correspond to phonemic /ʌ/ rather than /o/, this is not the case. In South Korean, the letter ㅗ corresponds to the phoneme /o/, and is pronounced as [o] (ignoring the presence of [o]-raising, which we discuss later). The letter ㅓ corresponds to the phoneme /ʌ/, and is pronounced as [ʌ]. In North Korean dialects which have undergone the /ʌ o/ merger, on the other hand, Kwak labels ㅗ (/o/) as being pronounced synchronically as [ɔ]. In

contrast, he describes how ㅓ (/ʌ/) had come to be pronounced as [o] before merging with /o/. This raises the question of the merged vowel's status as a dark or light vowel.

We conclude that the effects of the /ʌ o/ merger on the North Korean on the sound-symbolic forms in the corpus are far from certain. For this reason, and because North Korean forms are otherwise representative of sound-symbolic harmony, we chose to keep the North Korean forms in the present analysis (though we consider the potential confound in the analysis in §4). Finally, another reason we chose not to add or delete any forms from the corpus for the present study is so that others can replicate it straightforwardly. A more comprehensive study in the future would have to include other forms extracted from larger corpora of natural data, and would distinguish between dialectal forms.

4 Results

4.1 Strength of vowel harmony in Modern Korean

The first aim of this study is to verify the extent to which vowel harmony is present in sound-symbolic reduplicants in Korean. If vowel harmony is prevalent in these forms, we should expect to find few stems containing both light and dark vowels.

Table I provides the counts for the number of stems containing sequences of an initial vowel A followed by another vowel B in Korean sound-symbolic reduplicants.[10] Cells containing counts of disharmonic forms are shaded. Additionally, we have used a darker shade for the cells when non-initial [u] and [y] occur after light vowels, since treatment of these vowels as dark would classify the combinations in question as disharmonic. When a stem contained three syllables and the second and third syllables contained distinct vowels, it was counted once for each vowel. When a stem contained three syllables and the second and third syllables contained the same vowel, it was only counted once. Although we could have counted each vowel separately, in this way we avoid unnecessarily inflating the number of *stems* containing harmonic or disharmonic sequences. Numbers in parentheses indicate the counts after the removal of suspected non-sound-symbolic forms (see below for further discussion).

Based on the assumptions (to be validated later) that neutral vowels are dark in initial position and that [i ɨ u y] are all neutral vowels, we find out of a total of 3972 forms only 133 which contain an initial vowel of one type and a non-initial vowel of the opposite harmony type, comprising 3·3 % of the total.[11] This strongly supports earlier claims that vowel harmony is robust in sound-symbolic stems.

---

[10] We use 'initial' to refer to any vowel in the first syllable, regardless of whether a consonant precedes it or not.

[11] Here and below, we use the counts prior to the removal of suspected non-sound-symbolic forms, i.e. we do not use the counts in parentheses.

A \ B	L[a] 890	L[o] 217	L[æ] 33	L[ø] 0	D[ʌ] 968	D[e] 33	N[y] 10	N[u] 729	N[i] 551	N[ɨ] 1052
L[a] 945	454	66	13		10 (9)	3 (2)	3	123 (122)	124	274
L[o] 597	257	115	9		3	3		59 (56)	97	124
L[æ] 281	76	21	8		9			74 (73)	28	105
L[ø] 27	4	12						9	1	2
D[ʌ] 766	31 (27)				394	11	5	150	88	182
D[e] 85	10				43	2		18	9	16
D[y] 36	1				13			21	1	4
D[u] 632	28		2 (0)		271	8	1	150	116	137
D[i] 377	22 (21)	3			122		1	101	49	132
D[ɨ] 226	7		1		103	6		24	38	76

Table I

Number of stems with AB sequences, where A is a monophthong in the initial syllable and B is a monophthong in a non-initial syllable. Initial vowels are indicated in the left column, and non-initial vowels are indicated in the top row. The numbers in the left column and top row indicate the total number of stems containing each vowel in initial and non-initial position respectively. Cells containing disharmonic forms are shaded. Darker shading is used for cells in which non-initial [u] and [y] occur after light vowels. Numbers in parentheses indicate the counts after the removal of suspected non-sound-symbolic forms.

While Table I shows a strong correlation between vowels and the three harmony classes, there are a number of apparent disharmonic forms included above, especially involving non-initial [a]. We return to this in §4.3.

We saw in §3 (see also note 8) that the corpus may contain a small number of non-sound-symbolic forms. If this is the case, we might expect them to occur in the disharmonic forms. We therefore checked all of the disharmonic forms, using the dictionary provided by the Korean search

engine Naver.[12] Of those with non-initial [o] or [y], all were found in the Naver dictionary, as shown in (13) (NK denotes a North Korean form).

(13) a. pio-pio   sound of a black kite (bird)
       sikʰoŋ-sikʰoŋ   sound of a motor as it starts
       cicoŋ-cicoŋ (NK)   sound of a skylark
    b. tataky-tataky   'in clusters'
       aky-aky   said of vigorous chewing
       akycak-akycak (NK)   said of vigorous chewing

Of those with non-initial [æ], only one was found in the dictionary (14).[13]

(14) ɨŋæ-ɨŋæ   'crying' (of a baby)

Only one form with with non-initial [ʌ] was not found in the dictionary, and it did not appear to occur as a sound-symbolic reduplicant in natural data. Two were listed as mistakes of forms where [ʌ] should be [a]; nonetheless, they occur in natural data (15a, b).

(15) a. mancicʌk-mancicʌk   'fingering, fumbling'
       cf. mancicak-mancicak
    b. pæcʰʌk-pæcʰʌk   said of staggering with the body leaning
       cf. pæcʰak-pæcʰak,   to one side
       from pæcʰicak-pæcʰicak
    c. alɨlʌŋ-alɨlʌŋ   said of repeated howling or wailing of
                                   a small, fierce animal
    d. cakɨntʌk-cakɨntʌk   said of persistent irritating and
                                   annoying behaviour
    e. hacʌn-hacʌn   feeling of being surrounded by
                                   emptiness

Of the forms containing non-initial [e], half (three) were not found in a dictionary. Two of those were found to occur as sound-symbolic forms in

---

[12] An anonymous reviewer notes that the content of the *Naver* dictionary is provided by the *National Institute of the Korean Language* (http://stdweb2.korean.go.kr/main.jsp), which also produced the corpus. The reviewer considers this a potential concern, stating 'it is not surprising that the great majority of corpus data are attested in the dictionary'. Our purpose in verifying the data in the corpus was stated at the beginning of §3 (see also note 8), and we do not find it problematic that our verification method relies on a publication by the same organisation that produced the corpus. On the contrary, it is to be preferred, as we should expect that all valid entries in the corpus would appear in the dictionary (though they may be absent from other publishers' dictionaries), while the invalid entries would have been removed by the lexicographers prior to publication.

[13] The forms not found (as sound-symbolic words) were *uæ-uæ* and *uŋkæ-uŋkæ*. Note that, while not in the corpus or the Naver dictionary, *uŋæ-uŋæ* was found in internet searches as an alternative to *ɨŋæ-ɨŋæ* 'crying'. Thus the sequence u … æ is minimally attested, although not in the corpus.

natural data ((16a) was found in one context, repeated on several websites; (16b) was common), while the remaining one was not. Of the three in the dictionary, two were listed as North Korean forms.

(16) a. k'adek-k'adek   said of nodding the head back and forth
            cf. k'adak-k'adak
      b. colle-colle   'frivolously'
            cf. collæ-collæ
      c. tʰalle-tʰalle (NK)   said of a petite body moving in a light manner
      d. salle-salle (NK)   said of waving part of the body using small
                                    motions
      e. tolle-tolle   said of looking around in all directions

Last, almost all forms with non-initial [a] (94/99) were found in the Naver dictionary; however, another concern with this set of morphemes is the presence of North Korean forms which may have undergone the /ʌ o/ merger mentioned in §3. The corpus contains 31 stems with initial [ʌ] and non-initial [a], 27 of which are North Korean, while the remaining four are unattested in the dictionary. If the merger is complete in these forms, such that /ʌ/ is treated as /o/, then these forms are not disharmonic at all, because /o/, like /a/, is light.

   Additional discussion of stems with non-initial [a] is provided in §4.3. Forms with non-initial [u] and [y] are discussed in the following section. Despite the few potential errors described above, it is important to observe that almost all of the stems containing disharmonic forms in the corpus were found in the dictionary.

   Before closing this subsection, we wish to briefly compare our results with those of Hong (2010). Hong (2010) carried out two corpus studies, one on Korean roots in general and one on sound-symbolic words, with the purpose of determining the degree to which Korean roots as a whole conform to the general vowel-harmony pattern found in sound-symbolic forms. In the study on sound-symbolic words, Hong appears to use the same corpus as us; however, his counts do not match ours. While Hong extracted a total of 3191 sound-symbolic morphemes with two or more syllables which occurred in reduplicated forms, we found 3972 such forms. Part of the explanation is due to the fact that Hong does not consider the vowels [y] and [ø] (cf. note 4), but we can only speculate about the reasons for the remaining difference. Overall, Hong's analysis agrees with the one here: the forms are overwhelmingly harmonic. The largest difference is that Hong found only 21 disharmonic forms when comparing the first and second vowels of sound-symbolic morphemes, whereas we found 116 disharmonic forms in this condition (see Appendix A) and even more when we include the third syllable (which Hong does not). An additional difference between our studies is the statistical methodology, as Hong employs Pierrehumbert's (1993) O/E measure, while we use Fisher's exact test below.

	$D_{e,\Lambda}$	L
#$D_{e,\Lambda}$ …	450	41
#y …	13	1
#u …	279	30
#$N_{i,ɨ}$ …	231	33
#L …	28	1021

Table II

Stem frequency with which various harmony types
in initial position precede dark or light vowels.

4.2 Neutral vowels

As mentioned earlier, the neutral vowels [i] and [ɨ] are claimed to act as dark vowels when in the initial syllable and as transparent vowels elsewhere (Kim-Renaud 1976, Lee 1984, Sohn 1986, Park 1990, Lee 1992, Cho 1994). Furthermore, Lee (1984) and Cho (1994) claim that the vowel [u] also behaves like a neutral vowel to some extent, and Lee extends this claim to [y]. In this section, we will use 'neutral' to refer to [i] and [ɨ] only. Likewise, we use the term 'dark' to the exclusion of the (non-initial) vowels [u] and [y], referring to them simply as [u] and [y].

This section first demonstrates statistically that neutral vowels behave as dark vowels in initial syllables. It then examines vowels in non-initial syllables, showing that non-initial dark, light and neutral vowels have distinct distributions, and that the vowels [u] and [y] pattern more closely with neutral vowels than with dark or light vowels. Finally, we verify that N vowels are transparent, and show that the vowels [u] and [y] are also transparent.

4.2.1 *Neutral vowels in initial position.* We first demonstrate that [u], [y] and the neutral vowels behave as dark vowels, and are distinct from light vowels in initial syllables. Table II shows the number of stems with initial D, L, N, [u] and [y] vowels followed by the uncontroversial D (i.e. [e] and [ʌ]; henceforth $D_{e,\Lambda}$) and L vowels. Any time a stem contained an initial sound followed by a dark or light vowel in the second and/or third syllables, it was counted; however, if a stem contained an initial vowel followed by two dark or light vowels, it was only counted once (since we are counting total stems only).

To determine whether the differences in the frequencies of occurrence of dark and light vowels in the second and third syllables as a function of the initial vowel are significant, we conducted Fisher's exact tests on pairs of rows in Table II with the R statistical software package (R Development Core Team 2012). The results of the tests are presented in Table III.

	#L …	#N$_{i,ɨ}$ …	#u …	#y …
#D$_{e,ʌ}$ …	$< 2 \cdot 2 \times 10^{-16}$	0·07	0·53	1
#y …	$< 2 \cdot 2 \times 10^{-16}$	1	1	
#u …	$< 2 \cdot 2 \times 10^{-16}$	0·35		
#N$_{i,ɨ}$ …	$< 2 \cdot 2 \times 10^{-16}$			

Table III

P-values from Fisher's exact test on pairs of rows from Table II. Cells where no statistical difference was found ($p > 0 \cdot 05$) are shaded in this and subsequent tables.

All initial non-light vowels were statistically different from initial light vowels. Furthermore, no pairwise comparison of neutral vowels, dark vowels, [u] and [y] showed a significant difference. These results show that, in the aggregate, initial vowels pattern into two distinct classes. This is consistent with the traditional view that vowels in initial syllables in Korean sound-symbolic words behave as either dark or light.[14]

4.2.2 *Neutral vowels in non-initial position.* Next, we examine whether non-initial neutral vowels can co-occur with dark and light vowels, and more importantly, whether [u] and [y] pattern with dark vowels or with neutral vowels. Table IV provides the number of forms wherein D, [y], [u], N and L vowels are preceded by D, initial [y], initial [u], initial N and L vowels.

Again, to determine whether any of the differences between non-initial D, [y], [u], N and L vowels are significant, we conducted Fisher's exact test on pairs of rows in Table IV. We present the results in Table V.

Not surprisingly, non-initial light vowels were significantly different from all other vowels. Likewise, non-initial dark vowels were distinct from all other vowels, including the vowels [u] and [y], with a high degree of certainty. Neutral vowels were significantly different from both light and dark vowels, supporting the traditional view that non-initial neutral vowels behave like neither dark nor light vowels non-initially.

---

[14] If, due to the North Korean /ʌ o/ merger, the #ʌ … a sequences are not disharmonic, this would affect the counts in Table II as follows. The count for the sequence #D$_{e,ʌ}$ … L would change from 41 to 10, while #L … L would change from 1021 to 1052 (the corpus contains 17 disyllabic ʌ … a and 14 trisyllabic ʌ … a … a sequences). The results of the Fisher exact test in Table III would change as follows: {#D$_{e,ʌ}$ …, #N$_{i,ɨ}$ …} = $3 \cdot 52 \times 10^{-8}$, {#D$_{e,ʌ}$ …, #u …} = $7 \cdot 32 \times 10^{-6}$ and {#D$_{e,ʌ}$ …, #y …} = $-0 \cdot 28$. As a result, initial D$_{e,ʌ}$ vowels would become significantly different from initial N vowels and initial [u], while initial D$_{e,ʌ}$ vowels and initial [y] would remain the same.

	$D_{e,\Lambda} \cup \#N_{i,ɨ,u,y}$	L
... $D_{e,\Lambda}$	973	28
... y	7	3
... u	464	266
... $N_{i,ɨ}$	843	752
... L	105	1050

Table IV
Stem frequency with which various harmony types follow dark or light vowels.

	... L	... $N_{i,ɨ}$	... u	... y
... $D_{e,\Lambda}$	$< 2 \cdot 2 \times 10^{-16}$	$< 2 \cdot 2 \times 10^{-16}$	$< 2 \cdot 2 \times 10^{-16}$	$< 2 \cdot 8 \times 10^{-3}$
... y	$5 \cdot 98 \times 10^{-6}$	$0 \cdot 35$	$0 \cdot 75$	
... u	$< 2 \cdot 2 \times 10^{-16}$	$1 \cdot 4 \times 10^{-6}$		
... $N_{i,ɨ}$	$< 2 \cdot 2 \times 10^{-16}$			

Table V
P-values from Fisher's exact test on pairs of rows from Table IV.

Interestingly, while the results demonstrate with a very high degree of certainty that [u] and [y] are neither light nor dark, only [y] is not statistically different from the neutral vowels. [u], on the other hand, is statistically different from the neutral vowels, but not from [y].[15]

Rather than interpret the above results to mean there is a fourth class of vowels, we take this to indicate varying degrees of neutrality. We believe the statistical difference between [u] and neutral vowels is due to the fact that non-initial [u] is preceded by light vowels proportionally less than non-initial neutral vowels are. While this could be interpreted to mean that [u] is not neutral, there are two reasons not to think so. First, the fact that fewer light vowels precede [u] than precede neutral vowels probably has a historical basis. As noted in §2, vowel harmony in Middle Korean was more prevalent and rigid than in Modern Korean. The dark vowels, including [u], only occurred with neutral and dark vowels; the shift of [u]

---

[15] These results are stable even considering the potential /ʌ o/ merger. In Table IV, the count for the sequence $D_{e,\Lambda} \cup \#N_{i,ɨ,u,y}$ ... L would change from 105 to 74, while L ... L would change from 1050 to 1081. The results of the Fisher exact tests maintain the significant/non-significant distinctions shown in Table V.

towards a neutral status appears to be a modern development (see e.g. Lee 1984, Cho 1994). We would, therefore, expect there to be fewer light vowels preceding the 'newly minted' neutral vowel [u] than those preceding the traditional neutral vowels. Second, like non-initial neutral vowels, [u] vowels are preceded by relatively large numbers of both dark and light vowels, which supports the hypothesis that they are neutral. At the very least we can conclude that non-initial [u] vowels are much more like non-initial neutral vowels than non-initial dark or light vowels, though some may wish to call [u] a 'partially' neutral vowel.

One additional consideration points to the 'partial' neutrality of [u]. In a discussion of vowel alternations in harmony processes, Cho (1994: 116ff) (see also Kim-Renaud 1976: 400 and Lee 1992: 247ff) claims that, depending on the morpheme, [u] either must, may or may not undergo the alternation. The following examples are adapted from Cho (1994).

(17) ʌlluk   (*)alluk   allok   'mottled'
      pʌtuŋ   patuŋ   patoŋ   'struggling'
      cumullʌk   comullak   (*)comollak   'winding'

Online searches for the ungrammatical forms above reveal that they are actually attested in natural data and in the Naver dictionary (*alluk* being North Korean), which is why we have placed Cho's judgements in parentheses. Nonetheless, the fact that Cho notes a distinction suggests that [u] is not entirely neutral.

4.2.3 *Vowel harmony across medial vowels.*   Finally, we demonstrate that [u], [y] and the neutral vowels are transparent, allowing both light and dark vowel harmony to spread over them.

Table VI shows the number of forms in which D, L, N and [u] vowels occur between the four possible binary combinations of dark and light vowels. We have split neutral vowels into [i] and [ɨ], to demonstrate that each is transparent.

Again, we applied Fisher's exact test to pairs of rows in Table VI, and present the results below in Table VII.

While every pairing with L or D is significantly different, no pairing of [i], [ɨ], [u] and [y] is significantly different.[16] Although the neutral vowels do occasionally appear between disharmonic vowels (in fewer than 5% of instances for each vowel), the vast majority appear in the frames D __ D and L __ L. The data in Table VI therefore demonstrates quite clearly that [u], [y] and the neutral vowels behave as transparent vowels in non-initial position, while $D_{e,ʌ}$ and L vowels do not.

---

[16]   These results stand despite the potential confound introduced by the /ʌ o/ merger. In Table VI, the count for the sequence DLL would change from 29 to 15 and the sequence LLL from 152 to 166. The results of the Fisher exact tests maintain the significant/non-significant distinctions shown in Table VI.

	D__D$_{e,ʌ}$	L__L	D__L	L__D$_{e,ʌ}$
D$_{e,ʌ}$	118	0	0	0
y	2	1	0	0
u	75	45	0	1
i	43	39	3	1
ɨ	116	98	1	10
L	0	152	29	1

Table VI
Stem frequency with which various harmony types
appear between dark and light vowels.

	L	ɨ	i	u	y
D$_{e,ʌ}$	$< 2{\cdot}2 \times 10^{-16}$	$< 2{\cdot}2 \times 10^{-16}$	$< 2{\cdot}2 \times 10^{-16}$	$< 2{\cdot}2 \times 10^{-16}$	0·02
y	$< 2{\cdot}4 \times 10^{-4}$	1	1	1	
u	$< 2{\cdot}2 \times 10^{-16}$	0·08	0·06		
i	$< 2{\cdot}2 \times 10^{-16}$	0·11			
ɨ	$< 2{\cdot}2 \times 10^{-16}$				

Table VII
P-values from Fisher's exact test on pairs of rows from Table VI.

4.2.4 *Potential complications: the scarcity of* [y], *and* [o]-*raising.* While
the tests above indicate that both [u] and [y] are neutral vowels, the
confidence intervals for Fisher's exact tests are very large in some cases,
because there are so few [y] vowels in the corpus (only 46 forms contain
[y]). However, it is worth noting that none of these tests reveals a signifi-
cant difference between [y] on the one hand and [u] and neutral vowels on
the other for any of the cases under consideration. Also, Fisher's exact test
does reveal significant differences between non-initial [y] on the one hand
and non-initial dark and light vowels on the other, no significant differ-
ences between initial [y] and initial dark vowels, and significant differences
between initial [y] and initial light vowels. In other words, even though
there are few [y] vowels in the corpus, the results of these tests are 100%
consistent with the hypothesis that [y] is a neutral vowel, and inconsistent
with the hypothesis that it is a dark vowel. We further note that [y] appears
to pattern more closely with [u] than the other neutral vowels.

Both non-initial [y] and non-initial [u] appear to prefer to occur with dark vowels than with light vowels, at a ratio of approximately 2:1, whereas non-initial neutral vowels prefer dark vowels to light at only slightly more than a 1:1 ratio. In terms of ratios, [u] is actually slightly closer to neutral vowels than [y], despite appearing more divergent in the Fisher's exact tests.

Although the relative rareness of the vowel [y] in the corpus makes the hypothesis that [y] is neutral somewhat tentative, there is one additional piece of support for this hypothesis. If [y] is treated as dark, then the corpus contains three forms wherein the second and third syllables are disharmonic, two of which contain [y]. One is of the type L*y*L (*akycak-akycak* (13b)), while the other two are of the type LL*y*/LLD (*tataky-tataky* (13b) and *sæsætʌk-sæsætʌk* (said of noisy laughing and bantering)). Considering the rarity of non-initial [y], it would be surprising that it occurs in two out of three of these sequences if [y] is dark. On the other hand, the sequences L*y*L and LL*y* are not unexpected if [y] is treated as transparent in non-initial position.

In short, the results of the corpus indicate that the vowel [y], like [u], ought to be classified as a neutral vowel. This result supports a similar observation made in Lee (1984: n. 8) concerning [y]. Finally, if this hypothesis is correct, the addition of [y] would mean that the natural class of [+high] vowels make up the class of neutral vowels, rather than the unnatural class of [i ɨ u] to the exclusion of [y].[17]

Before continuing to the next section, we wish to address one further issue. An anonymous reviewer suggests that the raising of [o] to [u] in fast speech, which occurs especially in non-initial positions, might impinge upon the vowel-harmony facts (see also Lee 1992: 247ff). To test this, we checked all 265 of the disharmonic forms containing non-initial [u] on the Korean search engine Naver to see whether they appeared in its dictionary. Of these, only five, or 1·9%, were not found in the dictionary.[18] We also replaced [u] with [o] in all 265 of the disharmonic forms and checked whether they appeared in the Naver dictionary. A total of 67 [o] forms were listed in the dictionary. In one of these instances, the corresponding [u] form was not in the dictionary (though it was in the corpus). Of the 66 forms listed with both [o] and [u], all of the pairs contained at least one related meaning (with different connotations). In some cases, one of the forms contained an additional meaning not listed for the other form.

---

[17] Lee (1992: 252) also makes the observation that Modern Korean appears to be shifting towards having a neutral vowel class containing all [+high] vowels, but she does not consider the vowel [y] in her discussion (see note 4).

As observed in note 4, [y] is commonly pronounced as a diphthong, [ɥi] or [wi], which then simplifies to [u] or [i] in some dialects. If [y] is merging with these other vowels, then it is not surprising that it is becoming neutral.

[18] Among the disharmonic forms containing one of the initial vowels [a], [o] or [æ] and a non-initial [u], approximately 40–43% were labelled North Korean forms in the dictionary, which is slightly higher than the corpus as a whole (see note 8). Eight out of the nine forms containing initial [ø] and non-initial [u] were from the North Korean dialect.

Additionally, it was sometimes the case that one form was listed as a North Korean form and the other was not marked for dialect. In these cases, exemplified in (18), the form listed as North Korean was not consistently the [o] or [u] form.

(18) a. [o] *and* [u] *are both possible*

    palloŋ-palloŋ   said of the expansion and contraction of a flexible
    palluŋ-palluŋ   object

    p'æc'ok-p'æc'ok said of a wavy or jagged surface
    p'æc'uk-p'æc'uk

    kækol-kækol   'ribbit-ribbit' (sound of a frog)
    kækul-kækul

  b. [o] *and* [u] *are both possible, where choice depends on dialect*

    k'aok-k'aok   'kaw-kaw' (sound of a crow, raven)
    k'auk-k'auk (NK)

    tølok-tølok   said of an eyeball moving vigorously
    tøluk-tøluk (NK)

    hæc'ok-hæc'ok (NK) said of a cute or lovable type of smile
    hæc'uk-hæc'uk

While this data is not inconsistent with a hypothesis that speakers are raising [o], there are strong reasons to believe this is not the case. First, it is important to stress that 192 attested disharmonic forms containing [u] were not found in the dictionary when [u] was replaced with [o]. If these forms are simply the result of fast speech [o]-to-[u] raising, we would expect the dictionary to include the original [o] forms, but this is not the case. Moreover, the [o]-variants are frequently absent from naturalistic data as well, as confirmed by web searches (see (19)).

(19) [u] *cannot be replaced by* [o]

   kapult'ak-kapult'ak (NK) said of the quick bending and
   *kapolt'ak-kapolt'ak   straightening motion of an object

   komulak-komulak   said of frequent small movements
   *komolak-komolak

   kæut'uŋ-kæut'uŋ (NK)   said of a small object tilting back and
   *kæot'oŋ-kæot'oŋ   forth

Second, only two of the 265 disharmonic [u] forms were claimed to be mistakes in the corresponding [o] forms, while five of the harmonic [o] forms were listed in the dictionary as mistakes in [u] forms. Third, Lee (1992: 249) notes that, in cases where we see an [o] ~ [u] alternation, the alternative forms have different connotations (this was also reflected in the dictionary), which suggests they contain different vowels. Because of these facts, we do not believe the fast speech raising of [o] to [u] is an adequate explanation of the abundance of non-initial [u] vowels following light vowels. On the other hand, the claim that [u] is a neutral vowel predicts

that the frequency of its occurrence after light vowels should approach that of the neutral vowels [i] and [ɨ].

The data above argues against the hypothesis that synchronic raising of [o] to [u] can account for the seemingly disharmonic forms containing a non-initial [u]. Such an account would claim that all instances of disharmonic sound-symbolic stems containing a non-initial [u] must be derived from phonemic /o/, which would allow one to maintain that phonemic /u/ is actually dark. While this could hold in the examples provided in (18a), under the assumption that raising is optional, it would not explain why only the seemingly disharmonic variants in (19) are attested. One would need to claim that raising is sometimes required and sometimes optional, but unless the phonetic environment can be determined as the cause of the (non-)optionality of raising, this would be the same as positing a phonemic difference between the surface [u] forms in (18a) and (19). If, on the other hand, we accept that phonemic /u/ is neutral, then the pattern in (19) could be explained by positing an underlying /u/, given that [u]-to-[o] lowering has not, to our knowledge, be claimed to occur in Korean.[19] The optionality in (18a) may then indicate either the presence of a phonemic /o/ with optional [o]-to-[u] raising or that each form is derived from a (slightly) distinct underlying representation.[20]

4.3 The partial neutrality of [a]

One of the most salient pieces of data presented in Table I is the large number of mismatches involving the vowel [a]. In all, approximately 84% of the total mismatches involve [a]. An anonymous reviewer points out that the asymmetry between non-initial [a], which follows dark vowels approximately 11% of the time, and non-initial [o], which almost never

[19] We use the terms phonemic /o/ and phonemic /u/ purely for convenience, in order to distinguish two mental representations which may, in some cases, lead to the same pronunciation (here [u]). We do not wish, to make a claim about how the vowels in sound-symbolic morphemes are represented in the lexicon. To understand our reasoning for this, consider the final set of semantically related sound-symbolic forms from (3): [piŋkɨl, pæŋkɨl, pʌŋkɨl, paŋkɨl] 'with a smile'. This set appears to be based on a template /pVŋkɨl/, allowing V to be [i], [æ], [ʌ] or [a], but not [o], [u], [ɨ] or any other vowel. Because this contrast is not merely a light–dark alternation – which a simple rule or equivalent mechanism could account for – this hints at the possibility that vowels in sound-symbolic morphemes are underspecified or specified differently when compared to vowels in other words in the lexicon.

[20] An anonymous reviewer questions our conclusion that [o]-to-[u] raising cannot explain the patterning of [u], and that [u] should instead be classified as neutral. The reviewer states that the neutrality of [u] may have been caused by raising and provides the diachronic examples *hoto* > *hotu* 'walnut' and *cato* > *catu* 'plum'. The reviewer's comment itself does not refute the claim that [u] has become neutral, but rather it suggests that [o]-to-[u] raising is a cause for this change. We wish to clarify that we are not making any claims with regards to the *cause* of the shift of [u] from dark to neutral status, and we accept that diachronic [o]-to-[u] raising may have played a crucial role. We additionally accept that synchronic [o]-to-[u] raising – which is not limited to sound-symbolic forms – may also be influencing the shift of [u] toward neutrality.

does, matches their behaviour in verbal suffix harmony, as discussed in Hong (2008). The reviewer suggests that [a] is becoming a neutral vowel in Korean. This hypothesis deserves consideration, and we point out some important differences between [a] and the neutral vowels.

First, [a] is clearly a light vowel in initial syllables, occurring in mismatches in just under 1·4% of instances, whereas the neutral vowels (including [u] and [y]) have been demonstrated to be dark in initial syllables. Second, [a] is found much more frequently after light vowels than after dark vowels, whereas neutral vowels follow both dark and light vowels to relatively the same degree. Third, whereas neutral vowels appear in the frames DND and LNL with the same frequency (see Table VI), showing that they are transparent vowels, [a] never occurs in the frame D*a*D (unless we treat non-initial [u] as a dark vowel, as it is traditionally classified; even then, [a] only occurs in the frame D*a*D once, shown in (20)).

(20) tɨlk'apul-tɨlk'apul   said of shaking up and down

Instead, when [a] occurs in the second syllable of a three-syllable word containing an initial D vowel, it is almost always followed by [a] in the third syllable (29/30 roots; examples are provided in (21)). This strongly suggests that [a], if partially neutral, is not transparent, but rather opaque.

(21) sisitak-sisitak   said of speaking quietly with a light smile
     teŋkataŋ-teŋkataŋ (NK)   sound of a metallic object breaking or falling
     cʌlkatak-cʌlkatak (NK)   sound of large, solid objects lightly colliding
     pʰutatak-pʰutatak   said of a bird vigorously flapping its wings
                                    or a fish thrashing its tail back and forth,
                                    or the sound thereof

The sum of these facts suggests that [a] is fundamentally different from the traditional neutral vowels, as well as from [u] and [y]. While the vowels [i ɨ u y] are transparent, [a] instead appears to be partially opaque. The only issue with this characterisation is that if [a] is partially opaque, then why must the light vowel in D*a*L sequences be, without exception, [a], as opposed to any other light vowel? We leave this as an open question.

An alternative explanation for the high number of mismatches involving the vowel [a] is that it may be due to the presence of North Korean words in the corpus. While the majority of seemingly disharmonic stems with non-initial [a] were found in the Naver dictionary (94/99), a much larger percentage of the forms were labelled as North Korean than in our random searches on the entire corpus (58% compared to 33%). As mentioned in §4.1, all stems attested in the dictionary (27/31) containing the sequence #ʌ … a were North Korean forms, and are potentially not disharmonic in North Korean.

While the presence of dialectal variation should be unsurprising, what may be surprising is that in some dialects (e.g. standard South Korean),

non-initial [a] occurs in disharmonic stems only after certain initial dark vowels. While it is found after initial [e i u y ɨ], it appears to never occur after [ʌ] in the South Korean forms.[21] If we wish to maintain that [a] is partially neutral, it seems we must consider its 'neutrality' *vis à vis* different dark vowels individually. If [a]'s behaviour is relative to individual vowels, we should also expect other vowels to behave slightly differently, depending on the exact vowels with which they occur. For example, a quick look at initial [æ] suggests it may also be partially neutral. Focusing only on its occurrences with light or dark vowels, it occurs with [ʌ] in 7·9% of instances (alternatively, this may mean that [ʌ] is not entirely dark).[22] We return to these issues in §6.

5 Diphthongs

The inventory of Korean also contains a large number of opening diphthongs, with the onglides [j], [w] and [ɨ]. The full inventory is provided in (22), though we exclude [y] and [ø], which are pronounced as diphthongs in some dialects (see note 4).[23]

(22) *Korean diphthongs*
    a. *Onglide* [j]   [ja], [jo], [jæ], [jʌ], [je], [ju]
    b. *Onglide* [w]   [wa], [wæ], [wʌ], [we]
    c. *Onglide* [ɨ]   [ɨi]

We ignored the harmonic classification of diphthongs in the discussion above because the literature on Korean vowel harmony in sound-symbolic words does not generally discuss diphthongs. We might, however, expect diphthongs to fall into one of the three harmonic classes, based on either their onglide or nucleus. The data from the corpus indicates the latter suggestion is correct for the cases where the diphthong is found in the first syllable. When diphthongs occur in non-initial syllables, the results are equivocal.

---

[21] The sequence #e … a occurs in ten forms, only three of which are found in South Korean forms. Of these three, one (*teŋkaŋ*) is found in Naver's dictionary, while two (*cek'ak* and *c'ek'ak*) are not. These latter two are easily found in natural contexts across the internet with the meaning 'quickly', and appear to be alternate spellings of *cæk'ak* and *c'æk'ak*, both of which are found in dictionaries. When we consider that younger speakers of South Korean dialects rarely, if ever, distinguish [e] from [æ] (cf. Martin 1992: 25, Sohn 1999: 156), it is not clear that these forms should be considered disharmonic.

[22] Four of nine instances of initial [æ] followed by [ʌ] are South Korean, thus the North Korean /ʌ o/ merger may explain away half of these cases.

[23] Despite our use of the term diphthong, it is debatable whether onglides form part of the onset or nucleus in Korean. Yun (2004), for example, claims that onglides form part of the nucleus when an onset consonant is present, but otherwise they occur in onset position. We thank an anonymous reviewer for highlighting this point.

A \ B		L[a] 66	L[o] 5	L[æ] 1	D[ʌ] 33	N[y] 2	N[u] 42	N[i] 30	N[ɨ] 77	[ja] 9	[wa] 2	[jʌ] 23	[wʌ] 3	[ju] 1	[ɨi] 1
L[a]	2											2			
L[o]	7									3		4			
D[ʌ]	3							1 (0)		2 (0)		1			
D[e]	1											1			
D[y]	4											3		1	
D[u]	12									2	2	7	1		
D[i]	6											4	2		
D[ɨ]	1											1			
[ja]	60	9		1	1 (0)	1	28 (27)	10	17	2					
[jo]	5	1		2				1	1						
[jæ]	10			1		1	4 (3)	4	1						
[wa]	86	50			1		1	8	37 (36)						
[wæ]	11	6		1			3		3						
[jʌ]	6				1		1	3	1						
[wʌ]	24				21 (19)			2 (1)	6 (5)						
[we]	6				4 (3)		2		1 (0)						
[ju]	4						2		2						
[ɨi]	15		1		5 (4)		1	1	8						1

Table VIII

Number of stems with AB sequences, where A is a vowel in the initial syllable and B is a vowel in a non-initial syllable, and either A or B is a diphthong. Initial vowels are indicated in the left column, and non-initial vowels are indicated in the top row. The numbers in the left column and top row indicate the total number of stems containing each vowel in initial and non-initial position respectively. The shaded cells indicate a vowel-harmony mismatch. Numbers in parentheses indicate the counts after the removal of suspected non-sound-symbolic forms.

Table VIII shows the number of vowels which occurred in forms with diphthongs.[24] The shaded cells indicate where a vowel-harmony mismatch has occurred, under the assumption that diphthongs fall into harmonic classes as determined by their nuclei.[25]

Though the data is sparse, it strongly supports the conclusion that diphthongs in the initial syllable pattern with their nuclei, with only three harmonic mismatches (one of which we determined was not a legitimate sound-symbolic form) compared to 96 stems occurring with two light or dark vowels. It should also be noted that only three tokens contained two diphthongs, in each instance the diphthong in initial position matching the diphthong in non-initial position. The low number of co-occurrences is perhaps surprising when we consider that, as Table I shows, many tokens are found which contain the same monophthong in initial and non-initial positions.

If the conclusion that diphthongs pattern with their nuclei is correct, the above data provides further support for the claim that [y] and [u] are neutral rather than dark. [y] follows two light diphthongs, but does not follow any dark diphthongs, which is expected only if [y] is neutral (having already shown [y] is clearly not light). Similarly, [u] follows both light and dark diphthongs; indeed, it follows light diphthongs with greater frequency. Again, this is only expected if [u] is neutral.

Diphthongs in non-initial position do not appear to pattern with their nuclei. While the diphthongs [ɨi], [ju] and [wʌ] could be argued to do so, this cannot be the case for [ja], [jʌ] and [wa]. Moreover, non-initial [ja] and [jʌ], seemingly disharmonic in 44% and 26% of tokens respectively, are the only forms with enough data to draw any (even tentative) conclusions. It may be the case that diphthongs are all neutral in non-initial position, thereby patterning with their onglides, but this conclusion remains highly speculative.

To summarise, the corpus data suggests that diphthongs in initial positions fall into the classes of dark and light corresponding to their nuclei. In contrast, non-initial diphthongs do not appear to do so.

6 Discussion

The corpus study in the preceding sections reveals a much more nuanced picture of sound-symbolic vowel harmony in Korean than earlier research

---

[24] As before, the numbers in the left column and the top row indicate the total number of tokens, and not instances, in which the sound in question was found in the position in question; thus the sequence [wa a a] represents one token of [a] in a non-initial syllables, despite containing two instances of [a].

[25] To verify that the data was valid, we checked all 265 forms in the same manner as before. We found twelve forms (4·5%) which were neither in Naver's dictionary nor attested on the internet, and we indicate this below by placing the revised counts in parentheses. Of the remaining 252 forms, 89 (35·3%) were listed as North Korean forms, which is comparable to the case of monophthongs. Note that all [jæ]-initial forms were North Korean.

suggests. The back high rounded vowel [u], while exhibiting the dual functionality of the other neutral vowels, is preceded in non-initial position by far fewer light vowels than one would expect, given the distributions of the other neutral vowels. Also, some of disharmonic stems exhibit regularities. Specifically, the light vowel [a] precedes dark vowels in slightly more than 10 % of [a]-initial stems, but it does not precede all dark vowels to the same extent (perhaps subject to dialectal variation). The 'messiness' of the data is reminiscent of the study of Hungarian vowel harmony by Hayes & Londe (2006), where they demonstrate that the actual distribution of vowels in Hungarian exhibits patterned regularities in forms otherwise taken to be exceptional.

The above facts raise questions about whether classifications like dark, light and neutral are appropriate. Should vowel categories instead reflect each unique vowel distribution? If *any* statistically significant difference in *any* aspect of a vowel's distribution is the basis for forming a new category, then it is plausible that at least in the corpus study above, every vowel may, in fact, end up being the sole member of its category.

There are two points to be made in this regard. First, the distribution of individual vowels reflects the culmination of several different factors, not all of them being aspects of the grammar. For example, we mentioned the role history may have played in the distribution of [u]. If the goal is to determine aspects of the nature of Korean speakers' grammar from distributional studies of corpora, we need to be able to set aside non-grammatical factors. Without theories of those non-grammatical factors which make quantitative predictions, this is especially difficult.

Second, one purpose of categorising sounds into groups is to capture common elements in their behaviour. There are common elements in *major* aspects of the distributions of the vowels above, which we believe are captured well by the traditional classification of light, dark and neutral vowels. These classifications, however, cannot distinguish *minor* aspects of the vowels' distributions. These minor aspects cannot be captured on a linear scale, for the simple reason that the major categories are defined by a cluster of properties.

For example, in Korean sound-symbolic harmony, neutral vowels (i) behave like dark vowels syllable-initially, and (ii) are transparent non-initially. We consider both of these properties to be major aspects of the distribution of neutral vowels. But property (ii) can be broken down into occurrences in several different environments, such as #D__, #L__, #D__D and #L__L. For example, we said that the vowel [u] was 'less' neutral because it occurs much less frequently than other neutral vowels in the #L__ environment than in the #D__ environment. We consider this to be a minor aspect of the distribution of vowels because it is one aspect of property (ii). In this respect, the clustering of properties discussed here is reminiscent of the cluster properties associated with other grammatical categories in other domains, such as subjecthood in syntax (Keenan 1976).

The extent to which these lexical frequencies are internalised in the grammars of native speakers is an open question. Many researchers

have argued in favour of such a hypothesis, including Coleman & Pierrehumbert (1997), Zuraw (2000) and Hayes & Londe (2006). On the other hand, recent research has documented 'surfeit of the stimulus' effects (Becker *et al*. 2011). The important question, of course, is *which* lexical frequencies are tracked by speakers. Therefore, it is clear that future research on the vowel-harmony pattern of Korean ought to include experiments with Korean speakers looking for psycholinguistic evidence that that they have knowledge of the minor aspects of the distributions of the vowels. Modelling this knowledge with grammars and accounting for its learnability are additional worthwhile goals.

7 Conclusion

Our goal at the outset of this research was to examine the major aspects of the distribution of vowels with respect to their traditional classification as dark, light or neutral. By examining a corpus of 3972 sound-symbolic stems, we confirmed that neutral vowels have a dual nature, behaving as dark vowels in initial position and as transparent vowels in non-initial position. Furthermore, this study has concluded that the vowels [u] and [y] behave like neutral vowels, as claimed by Lee (1984), Lee (1992) and Cho (1994). It was observed that [u] was distinct from the traditional neutral vowels [i] and [ɨ] in minor aspects of its distribution. We also showed that non-initial [a] behaves like a neutral vowel to some degree, but unlike the other neutral vowels in Korean, it is best described as opaque in these cases. Next, we demonstrated that when Korean sound-symbolic words contain diphthongs in the initial syllables, they pattern with the harmony class of their nuclei; however, when diphthongs occur in non-initial syllables, it is unclear how they should be classified.

Taken together, these results show that the vowel-harmony pattern in sound-symbolic words in Korean, like the patterns in Finnish and Hungarian, is much more nuanced than previously thought. This research lays the groundwork for complementary studies which can probe speakers' fine-grained knowledge of these vowel patterns, and can model the learnability and nature of this knowledge.

Appendix A

Total sequence counts for monophthongs (D = [e ʌ u y i ɨ], $D_{e,\Lambda}$ = [e ʌ])

1 Two-syllable patterns

$DD_{e,\Lambda}$ 602
  $D_{e,\Lambda}D_{e,\Lambda}$ 260
  $ND_{e,\Lambda}$ 153
  $uD_{e,\Lambda}$ 181
  $yD_{e,\Lambda}$ 8

DL   71
  $D_{e,\Lambda}L$ 23
  NL   24
  uL   23
  yL   1

DN 596
  $D_{e,\Lambda}N$ 207
  NN   209
  uN   176
  yN   4

Du 323
  $D_{e,\Lambda}u$ 126
  Nu   86
  uu   96
  yu   15

Dy   4
  $D_{e,\Lambda}y$ 3
  uy   1

$LD_{e,\Lambda}$ 15

LL   668

LN   521

Lu   191

Ly   1

2 Three-syllable patterns

$DD_{e,\Lambda}D_{e,\Lambda}$ 118
  $D_{e,\Lambda}D_{e,\Lambda}D_{e,\Lambda}$ 86
  $ND_{e,\Lambda}D_{e,\Lambda}$ 9
  $uD_{e,\Lambda}D_{e,\Lambda}$ 22
  $yD_{e,\Lambda}D_{e,\Lambda}$ 1

$DD_{e,\Lambda}L$ 0

$DD_{e,\Lambda}N$ 6
  $D_{e,\Lambda}D_{e,\Lambda}N$ 4
  $uD_{e,\Lambda}N$ 2

$DD_{e,\Lambda}u$ 10
  $D_{e,\Lambda}D_{e,\Lambda}u$ 6
  $ND_{e,\Lambda}u$ 2
  $uD_{e,\Lambda}u$ 2

$DD_{e,\Lambda}y$ 1
  $D_{e,\Lambda}D_{e,\Lambda}y$ 1

$DLD_{e,\Lambda}$ 0

DLL   29
  $D_{e,\Lambda}LL$ 18
  NLL   4
  uLL   7

DLN   0

DLu   1
  NLu   1

DLy   0

$DND_{e,\Lambda}$ 159
  $D_{e,\Lambda}ND_{e,\Lambda}$ 71
  $NND_{e,\Lambda}$ 48
  $uND_{e,\Lambda}$ 39
  $yND_{e,\Lambda}$ 1

DNL   4
  NNL   4

DNN   66
  $D_{e,\Lambda}NN$ 9
  NNN   26
  uNN   31

DNu   10
  $D_{e,\Lambda}Nu$ 3
  NNu   7

DNy   0

$DuD_{e,\Lambda}$ 75
  $D_{e,\Lambda}uD_{e,\Lambda}$ 21
  $NuD_{e,\Lambda}$ 18
  $uuD_{e,\Lambda}$ 33
  $yuD_{e,\Lambda}$ 3

DuL   0

DuN   2
  uuN   2

Duu   43
  $D_{e,\Lambda}uu$ 12
  Nuu   11
  uuu   17
  yuu   3

Duy   0

$DyD_{e,\Lambda}$ 2
  $D_{e,\Lambda}yD_{e,\Lambda}$ 1
  $NyD_{e,\Lambda}$ 1

DyL   0

DyN   0

Dyu   0

Dyy   0

$LD_{e,\Lambda}D_{e,\Lambda}$ 0

$LD_{e,\Lambda}L$ 0

$LD_{e,\Lambda}N$ 0

$LD_{e,\Lambda}u$ 0

$LD_{e,\Lambda}y$ 0

$LLD_{e,\Lambda}$ 1

LLL   152

LLN   8

LLu   8

LLy   1

$LND_{e,\Lambda}$ 11

LNL   137

LNN   69

LNu   6

LNy   0

$LuD_{e,\Lambda}$ 1

LuL   45

LuN   0

Luu   14

Luy   0

$LyD_{e,\Lambda}$ 0

LyL   1

LyN   0

Lyu   0

Lyy   0

Appendix B

Vowel sequences of forms not containing diphthongs (excepting [y] and [ø])

1 Two-syllable patterns

a	a	258	æ	u	53	ɨ	æ	1	ø	a	4	ʌ	e	11
a	æ	13	æ	ʌ	1	ɨ	e	6	ø	i	1	ʌ	i	72
a	e	3	e	a	6	ɨ	i	32	ø	ɨ	2	ʌ	ɨ	119
a	i	97	e	i	7	ɨ	ɨ	43	ø	o	11	ʌ	u	113
a	ɨ	166	e	ɨ	9	ɨ	u	18	ø	u	7	ʌ	ʌ	229
a	o	57	e	u	13	ɨ	ʌ	77	u	a	21	ʌ	y	3
a	u	92	e	ʌ	20	o	a	172	u	æ	2	y	a	1
a	ʌ	5	i	a	15	o	æ	8	u	e	8	y	i	1
a	y	1	i	i	35	o	e	3	u	i	86	y	ɨ	3
æ	a	38	i	ɨ	99	o	i	74	u	ɨ	90	y	u	15
æ	æ	2	i	o	3	o	ɨ	81	u	u	96	y	ʌ	8
æ	i	22	i	u	68	o	o	86	u	ʌ	173			
æ	ɨ	78	i	ʌ	70	o	u	39	u	y	1			
æ	o	19	ɨ	a	5	o	ʌ	3	ʌ	a	17			

2 Three-syllable patterns

a	a	a	84	æ	æ	ɨ	1	i	i	u	4	o	a	o	2	u	u	ʌ	33
a	a	i	1	æ	æ	u	3	i	i	ʌ	7	o	i	a	17	u	ʌ	i	2
a	a	ɨ	3	æ	æ	ʌ	1	i	ɨ	ɨ	11	o	i	i	3	u	ʌ	u	2
a	a	u	2	æ	i	a	4	i	ɨ	u	2	o	i	ɨ	2	u	ʌ	ʌ	22
a	a	y	1	æ	i	u	2	i	ɨ	ʌ	20	o	ɨ	a	18	ʌ	a	a	14
a	i	a	14	æ	ɨ	a	12	i	u	u	11	o	ɨ	ɨ	22	ʌ	i	ɨ	1
a	i	i	3	æ	ɨ	ɨ	6	i	u	ʌ	15	o	o	a	9	ʌ	i	u	3
a	i	ɨ	1	æ	ɨ	u	1	i	ʌ	u	1	o	o	æ	1	ʌ	i	ʌ	11
a	i	o	4	æ	ɨ	ʌ	6	i	ʌ	ʌ	8	o	o	ɨ	1	ʌ	ɨ	ɨ	8
a	i	u	3	æ	o	u	1	i	y	ʌ	1	o	o	o	16	ʌ	ɨ	ʌ	51
a	i	ʌ	1	æ	u	a	4	ɨ	a	u	1	o	u	a	19	ʌ	u	u	11
a	ɨ	a	68	æ	u	u	8	ɨ	i	i	1	o	u	u	1	ʌ	u	ʌ	20
a	ɨ	ɨ	32	æ	u	ʌ	1	ɨ	i	ɨ	1	ø	o	u	1	ʌ	ʌ	i	1
a	ɨ	ʌ	4	e	a	a	4	ɨ	i	u	1	ø	u	u	1	ʌ	ʌ	ɨ	3
a	o	o	5	e	e	u	2	ɨ	i	ʌ	3	u	a	a	7	ʌ	ʌ	u	3
a	u	a	22	e	i	ʌ	2	ɨ	ɨ	a	1	u	i	i	4	ʌ	ʌ	ʌ	74
a	u	u	4	e	ɨ	ʌ	7	ɨ	ɨ	ɨ	13	u	i	ɨ	3	ʌ	ʌ	y	1
a	y	a	1	e	u	u	1	ɨ	ɨ	ʌ	18	u	i	ʌ	19	ʌ	y	ʌ	1
æ	a	a	14	e	u	ʌ	1	ɨ	u	ʌ	3	u	ɨ	ɨ	24	y	ɨ	ʌ	1
æ	a	ɨ	1	e	ʌ	u	1	ɨ	ʌ	u	1	u	ɨ	ʌ	20	y	u	u	3
æ	a	o	1	e	ʌ	ʌ	12	ɨ	ʌ	ʌ	1	u	u	i	1	y	u	ʌ	3
æ	a	u	1	i	a	a	4	o	a	a	19	u	u	ɨ	1	y	ʌ	ʌ	1
æ	æ	a	1	i	i	a	3	o	a	i	1	u	u	u	17				

Appendix C

Vowel sequences of forms containing diphthongs (other than those with [y] or [ø])

1 Two-syllable patterns

a	jʌ	2	ja	i	7	jo	ɨ	1	u	jʌ	7	wæ	ɨ	1
e	jʌ	1	ja	ɨ	15	jo	o	2	u	wa	2	wæ	u	3
i	jʌ	4	ja	ja	2	ju	ɨ	2	u	wʌ	1	we	u	2
i	wʌ	2	ja	o	1	ju	u	2	ʌ	ja	1	we	ʌ	2
ɨ	jʌ	1	ja	u	19	jʌ	i	3	ʌ	jʌ	1	wʌ	i	1
i̯i	i	1	jæ	i	4	jʌ	ɨ	1	wa	a	27	wʌ	ɨ	1
i̯i	ɨ	7	jæ	ɨ	1	jʌ	u	1	wa	i	3	wʌ	ʌ	11
i̯i	o	1	jæ	o	1	jʌ	ʌ	1	wa	ɨ	10	y	ju	1
i̯i	u	1	jæ	u	3	o	ja	3	wa	u	1	y	jʌ	3
i̯i	ʌ	1	jo	a	1	o	jʌ	4	wæ	a	3			
ja	a	4	jo	i	1	u	ja	2	wæ	æ	1			

2 Three-syllable patterns

i̯i	i̯i	ʌ	1	ja	u	u	2	wa	i	i	1	we	ʌ	ʌ	1
i̯i	ʌ	ɨ	1	ja	u	ʌ	1	wa	i	ɨ	2	wʌ	i	ʌ	1
i̯i	ʌ	ʌ	2	ja	y	u	1	wa	ɨ	a	5	wʌ	ɨ	ɨ	1
ja	i	a	1	jæ	y	u	1	wa	ɨ	ɨ	18	wʌ	ɨ	ʌ	3
ja	i	u	2	ʌ	i	ja	1	wa	ɨ	ʌ	1	wʌ	ʌ	ɨ	1
ja	ɨ	a	1	wa	a	a	15	wæ	a	a	1	wʌ	ʌ	ʌ	5
ja	ɨ	ɨ	1	wa	a	ɨ	1	wæ	ɨ	a	2				
ja	u	a	3	wa	i	a	2	we	ɨ	ʌ	1				

REFERENCES

Becker, Michael, Nihan Ketrez & Andrew Nevins (2011). The surfeit of the stimulus: analytic biases filter lexical statistics in Turkish laryngeal alternations. *Lg* **87**. 84–125.
Cho, Mi-Hui (1994). *Vowel harmony in Korean : a grounded phonology approach*. PhD dissertation, Indiana University.
Coleman, John & Janet B. Pierrehumbert (1997). Stochastic phonological grammars and acceptability. In John Coleman (ed.) *Proceedings of the 3rd Meeting of the ACL Special Interest Group in Computational Phonology*. Somerset, NJ : Association for Computational Linguistics. 49–56.
Goldsmith, John & Jason Riggle (2012). Information theoretic approaches to phonological structure : the case of Finnish vowel harmony. *NLLT* **30**. 859–896.
Hayes, Bruce & Zsuzsa Cziráky Londe (2006). Stochastic phonological knowledge : the case of Hungarian vowel harmony. *Phonology* **23**. 59–104.

Hong, Sung-Hoon (2008). Variation and exceptions in the vowel harmony of Korean suffixes. *The Journal of Studies in Language* **24**. 405–428.

Hong, Sung-Hoon (2010). Gradient vowel cooccurrence restrictions in mono-morphemic native Korean roots. *Studies in Phonetics, Phonology and Morphology* **16**. 279–295.

Karlsson, Fred (1999). *Finnish : an essential grammar*. 2nd edn. London & New York : Routledge.

Keenan, Edward L. (1976). Towards a universal definition of 'subject'. In Charles N. Li (ed.) *Subject and topic*. New York : Academic Press. 303–333.

Kim-Renaud, Young-Key (1976). Semantic features in phonology : evidence from vowel harmony in Korean. *CLS* **12**. 397–412.

Ko, Seongyeon (2010). A contrastivist view on the evolution of the Korean vowel system. *MIT Working Papers in Linguistics* **61**. 181–196.

Kwak, Chung-Ku (2003). The vowel system of contemporary Korean and direction of change. [In Korean.] *Kwukehak* **41**. 59–91.

Lee, Jin-Seong (1992). *Phonology and sound symbolism of Korean ideophones*. PhD dissertation, Indiana University.

Lee, Sang-Oak (1984). An overview of issues in the vowel system and vowel harmony of Korean : disharmony among the hypotheses of vowel harmony. *Ohak Yongu* [*Language Research*] **20**. 417–451.

McCarthy, John J. (1983). Phonological features and morphological structure. In John F. Richardson, Mitchell Marks & Amy Chukerman (eds.) *Papers from the parasession on the interplay of phonology, morphology, and syntax*. Chicago : Chicago Linguistic Society. 135–161.

Martin, Samuel E. (1992). *A reference grammar of Korean : a complete guide to the grammar and history of the Korean language*. Singapore : Tuttle.

Park, Sayhyon (1990). Vowel harmony in Korean. *Ohak Yongu* [*Language Research*] **26**. 469–499.

Pierrehumbert, Janet (1993). Dissimilarity in the Arabic verbal roots. *NELS* **23**. 367–381.

R Development Core Team (2012). R: a language and environment for statistical computing. Vienna : R Foundation for Statistical Computing. Available at http://www.r-project.org.

Ringen, Catherine O. & Orvokki Heinämäki (1999). Variation in Finnish vowel harmony : an OT account. *NLLT* **17**. 303–337.

Rounds, Carol (2001). *Hungarian : an essential grammar*. London & New York : Routledge.

Sohn, Ho-Min (1999). *The Korean language*. Cambridge : Cambridge University Press.

Sohn, Hyang-Sook (1986). Toward an integrated theory of morphophonology : vowel harmony in Korean. *Studies in the Linguistic Sciences* **16:2**. 157–184.

Vago, Robert M. (1974). *Hungarian generative phonology*. PhD dissertation, Harvard University.

Vago, Robert M. (1976). Theoretical implications of Hungarian vowel harmony. *LI* **7**. 243–263.

Yang, Byunggon (1992). An acoustical study of Korean monophthongs produced by male and female speakers. *JASA* **91**. 2280–2283.

Yun, Yungdo (2004). *Glides and high vowels in Korean syllables*. PhD dissertation, University of Washington.

Zuraw, Kie (2000). *Patterned exceptions in phonology*. PhD dissertation, University of California, Los Angeles.