CrossMark

J. Linguistics **51** (2015), 3–39. © Cambridge University Press 2014
doi:10.1017/S002222671400022X First published online 18 June 2014

Phonotactic constraints and sub-syllabic structure: A difficult relationship[1]

THOMAS BERG

Department of English, University of Hamburg

CHRISTIAN KOOPS

Department of Linguistics, University of New Mexico

(Received 18 January 2013; revised 28 January 2014)

Of late, a controversy has arisen over the internal structure of Korean syllables. While there is general agreement that non-phonotactic criteria argue for left-branching, Lee & Goldrick's (2008) left-branching phonotactic analysis is contradicted by Berg & Koops's (2010) claim as to a phonotactically symmetrical syllable structure. A comparison of the methodologies of the two studies, a revisit of the previous data and a new analysis cement the conclusion that there is neither a left-branching nor a right-branching phonotactic effect in Korean syllables. An investigation of the phonotactic structure of Finnish CVC syllables, which exhibit a psycholinguistic left-branching bias much like Korean, reveals that word-initial syllables possess a largely symmetrical organization whereas word-final syllables tend to show a right-branching slant. This curious set of results is consistent with the following three hypotheses: (i) The phonotactic criterion has an inherent VC bias. (ii) Symmetrical syllable structures represent a compromise between left- and right-branching effects. (iii) The strength of phonotactic constraints increases from earlier to later portions of words. The bottom line of this analysis is that, contra all previous claims, phonotactic constraints cannot be used as an argument for sub-syllabic constituency. We discuss the proposal that the basis of the left-branching bias in Korean syllables is instead to be found in the high degree of coarticulation between the onset consonant and the following vowel.

1. Introduction

The study of the internal structure of the syllable has long played an important part in suprasegmental phonology. A time-honoured method of examining syllable structure is through phonotactic analysis (see Pike & Pike 1947 for an early example). The following three observations have proved essential: phonotactic constraints are naturally sensitive to distance, i.e. they operate more forcefully between adjacent than between non-adjacent segments (and more forcefully between tautosyllabic than between heterosyllabic segments). In addition to linear and hierarchical

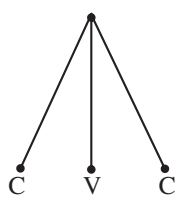
[1] We extend our gratitude to Florian Biermann and especially to Frauke Gebauer for performing the Herculean task of coding the Finnish data as well as to You-Mi Seo and Frauke Gebauer for their native-speaker advice on languages which we are not so familiar with. The reports of the *JL* referees are also acknowledged with gratitude.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

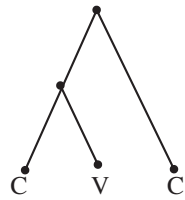
distance, the phonological nature of the segments represents a significant influence. Two neighbouring consonants evince a larger number of phonotactic restrictions than a consonant and a neighbouring vowel. A further factor was discovered in a comparison of (tautosyllabic) consonant–vowel (CV) and vowel–consonant (VC) sequences. The former were found to be subject to fewer phonotactic constraints than the latter (e.g. Fudge 1987 and Kessler & Treiman 1997 on English).

These facts were taken as direct evidence for the organization of the syllable. The three basic structural types of CVC syllables are illustrated in (1). Diagram (1a) illustrates a flat organization, (1b) hierarchical left-branching and (1c) hierarchical right-branching.

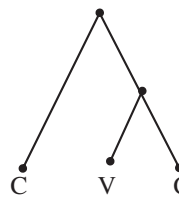
(1) (a)



(b)



(c)



The general logic of the phonotactic argument is straightforward and seemingly impeccable. Phonotactic constraints are understood to be a reflection of syllable-internal constituency. A tighter syllable organization imposes tighter phonotactic constraints than a looser organization. More specifically, the first factor (distance) identifies the syllable as the relevant (though not the only) domain of phonotactic constraints. The second factor gives rise to nodes for consonant clusters and the third factor provides the basis for the claim that the structure of the (English) CVC syllable is right-branching. In fact, the syllable is divided into the two immediate constituents called 'onset' (C) and 'rime' (VC), thereby reflecting the closer association of the vowel with the following rather than the preceding consonant.

Among the numerous lines of evidence that have been brought to bear on syllable structure, phonotactic constraints have always played a pre-eminent role. This is because they provide structural information at the segmental level, which is of the same nature as the information one wishes to obtain at the syllabic level. The link between phonotactic constraints and syllable organization thus appears to be especially direct, at least more direct than that between syllable structure and other types of evidence. Take poetic rhymes as an example (as also discussed in Fudge 1987). While the nature of poetic rhymes may certainly shed light on syllable-internal constituency, there is no guarantee that poetic rhymes must correspond to phonological rimes. That is, a poetic rhyme may have come into being for reasons that are unrelated to the existence of a phonological rime node. Differently put, the presence or absence of a poetic rhyme does not prove the presence or absence of a phonological rime. The relation between the two notions thus is somewhat indirect.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

In view of the potential uncertainty surrounding individual bits of evidence (and in order to minimize the risk of methodological artifacts), the analysis of syllable structure should be informed not by a single criterion, but by a number of criteria. Of course, one would hope that different criteria converge on the same structural option. However, what are we to do if two different criteria favour mutually exclusive conclusions, with one criterion arguing for a left-branching, and the other for a right-branching, structure? Even though such conflicts exist, relatively little effort has been expended on resolving them. Can branching direction be decided by majority rule? Is it justified to ignore the odd one out? Moreover, how do we explain symmetrical representations? Can they be understood as a compromise between left- and right-branching effects?

Berg & Koops (2010) examined one such conflict in detail. In an analysis of the internal structure of Korean CVC syllables, they applied standard arguments used to determine branching direction in suprasegmental phonology, including word games, reduplication patterns and phonotactic constraints. They found that while the bulk of the evidence weighs in favour of a left-branching structure, phonotactic constraints are about equally distributed across the body and the rime domains. Thus, the phonotactic criterion suggests a symmetrical syllable structure, i.e. option (1a) above. This is very much unlike the situation in English CVC syllables where all criteria without exception speak in favour of right-branching, i.e. option (1c) above (e.g. Anderson & Port 1994, Kessler & Treiman 1997, Da Cara & Goswami 2002, Nimmo & Roodenrys 2004, Kapatsinski 2009).

In their account of this unexpected finding, Berg & Koops questioned a methodological assumption which had never been explicitly spelt out but which had always been taken for granted. This assumption states that all criteria standardly used in the analysis of sub-syllabic structure have the same baseline probability of supporting a particular branching direction. Rather, each line of evidence may be biased in favour of a particular analysis. Specifically, they argued that the phonotactic criterion has an inherent VC bias. In Korean syllables, where the phonotactic evidence reveals equally strong CV and VC effects, the right-branching phonotactic bias appears to be offset by that language's general tendency toward left-branching, which is seen in all the psycholinguistic evidence available (see Yi 1999, Yoon & Derwing 2001, Yoon et al. 2002, Kim 2007, Lee, Lee & Kim 2010). This generates the observed phonotactic symmetry within the syllable, which thereby reflects a kind of compromise result.

Furthermore, Berg & Koops (2010: 44–46) put forward the strong claim that the hypothesized phonotactic VC bias is universal in nature because it follows from general aspects of speech processing. Since speech is auditorily processed from left to right, earlier parts of words are more important for lexical access than later ones. Applied to CVC words, the CV sequence does relatively more 'work' than the VC sequence. Phonotactic constraints, which reduce the combinatory power of contrasting phonemes, are better placed later in the syllable than at the beginning, where the maintenance and full exploitation of all available

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

contrasts is most needed. The VC portion of CVC words by definition occurs later than the CV portion and therefore plays a less crucial part in language comprehension. This leads to the inherent right-branching bias of the phonotactic criterion. We will return to this point in the discussion in Section 6 below.

Lee & Goldrick (2008) present an independent investigation of Korean phonotactic structure on the basis of earlier work by Lee (2006a, b). In contrast with Berg & Koops's study, their analysis points to a left-branching phonotactic bias. This is a potentially reassuring result, as it creates perfect harmony among the different criteria used to determine sub-syllabic structure, and thereby strengthens the case for left-branching in Korean syllables. Of course, even in the case of complete agreement among different criteria, these may still have different baseline probabilities of supporting a particular branching direction. However, as long as all of them point in the same direction, there is little incentive to examine their possible differences, which obviates the need for extensive methodological reflections like those undertaken by Berg & Koops.

The controversy summarized above raises two critical issues which will be addressed in the present paper. First, it is imperative to confront the contradictory empirical findings of Lee & Goldrick (henceforth L&G) and Berg & Koops (henceforth B&K). This is obviously a pressing task because as long as there is no consensus on the empirical facts, it makes little sense to construct theoretical models or to launch methodological debates.

Secondly, it is necessary to put B&K's model to the test. As it stands, this model is a *post hoc* explanation of an unexpected result and therefore requires further support. There is also a worrying discrepancy between its universal nature and the fact that it has been developed on the basis of a single language. The specific prediction to be tested is that a language cannot show an unequivocal phonotactic CV effect, even if it displays other left-branching phonological effects. To this end, we consider Finnish syllables here, which resemble Korean syllables in that they evidence a psycholinguistic left-branching bias. If the structure of Finnish syllables displays no left-branching phonotactic effect, B&K's account would receive independent empirical support.

The third aim of this paper is a theoretical one. If the connection between phonotactics and branching direction is less secure than has hitherto been assumed, the underpinnings of branching direction have to be reconsidered. It will be argued in this paper that phonotactics and branching direction are not causally related at all. As an alternative to phonotactic patterns, we discuss a phonetic explanation for branching structure which is grounded in coarticulatory effects.

2. Another look at Berg & Koops (2010) and Lee & Goldrick (2008)

The discrepant results obtained by B&K and L&G invite a careful reconsideration of each author team's methodological decisions and their possible implications. It is worth noting at the outset that the two studies take a very similar approach

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

to the problem. Methodologically, both follow the landmark study of probabilistic phonotactic effects in English by Kessler & Treiman (1997). Each author team performed a statistical analysis of probabilistic phonotactic constraints in Korean. Specifically, both calculated the strength of association between consonants and vowels in CV and VC sequences contained in CVC words. The prediction of both studies was that overall the association between the onset consonant and the vowel should be significantly greater than that between the vowel and the coda consonant.

In the following, we discuss five points of contrast between B&K's and L&G's analyses: (i) the database, (ii) the criteria used for data selection, (iii) the level of abstractness of the analysis, (iv) the assumed phoneme inventory, and (v) the statistical procedure. Interestingly, none of these points appears particularly likely to produce disparate results. Phonotactic structure can be assumed to be such a fundamental feature of language that it is revealed equally well in any dataset sufficiently large to prevent a type II error (i.e. a spurious null result). The same goes for the criteria for data selection. Given the pervasiveness of phonotactic structure, only criteria which severely distort the true picture would be capable of either concealing or falsely creating phonotactic effects. A minor explanatory potential may lie in statistical testing. If one test is more sensitive than another, the latter may yield a null result in the case of a weak effect while the former leads to a rejection of the null hypothesis.

The following discussion is designed to shed as much light as possible on potential reasons for the divergent results of B&K and L&G. At the same time, our discussion serves to set the stage for the analysis of Finnish CVC syllables, in which we will adopt the same general methodological approach.

2.1 *Database*

Both B&K and L&G performed statistical analyses of a subset of the Korean lexicon, viz. CVC words. B&K extracted all CVC entries from a large monolingual dictionary of Korean (Kim et al. 1991). L&G used a word list obtained from a website of the National Institute of the Korean Language (see Lee 2006a: 52f.). The two resulting datasets are of quite different size: 3,001 items in the case of B&K and 939 items in the case of L&G.

If the different size of the datasets is responsible for the divergent results, it would have to be argued that large databases can cover up an effect which is present in smaller databases. Let us uncontroversially assume that a large dictionary contains a higher proportion of infrequent and obsolete entries than a smaller one. If we go along with Treiman et al.'s (2000) claim that infrequent items are less prone to develop a branching direction asymmetry than frequent ones, we might have an explanation for the null result in B&K and the left-branching bias found by L&G. Although we do not find this line of argument particularly convincing, we believe that it is worth testing whether the size of

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

the database makes a difference. Below we present the results of a new analysis based on a different Korean dictionary, one that is smaller than that originally used by B&K.

We add in parentheses that L&G's word list was not put out in a typical publishing format so that it is not clear whether the compilers adhered to conventional lexicographic standards. We can do little to address this issue here other than to note that previous lexical analyses were usually based on published dictionaries (e.g. Kessler & Treiman 1997).

2.2 *Data selection criteria*

Following Kessler & Treiman (1997), both B&K and L&G compiled datasets consisting of lexical items of immediate relevance to the internal structure of CVC syllables, viz. all monosyllabic CVC words of Korean. In doing so, their selection criteria were in many ways parallel. Both studies counted homophonous lexical entries separately. Neither study excluded loanwords. However, some differences remain. L&G excluded proper nouns whereas B&K did not. More importantly, B&K included inflected verb and adjective roots like *mək-* 'eat' and *cak-* 'be small'. In dictionaries, these are listed with the infinitival suffix *-ta* (e.g. *mək-ta* 'eat-INF', *cak-ta* 'be small-INF'). On the other hand, L&G restricted their set of CVC words to roots which are not morphologically bound. This means that their database was essentially limited to nouns and numerals, e.g. *mok* 'neck' and *sam* 'three'. This might help to explain some of the difference in the size of the two datasets. However, it cannot be ruled out that bound verb and adjective roots are phonologically unusual and therefore create a special phonotactic bias. This possibility will be addressed below.

2.3 *Level of abstractness*

In Korean, phonemic representations (as found in dictionaries) often differ substantially from their phonetic realization in speech. One cause of this mismatch is the sweeping phonological process of coda obstruent neutralization (e.g. Park 1967, Kim-Renaud 1986, Kim & Jongman 1996, Kaplan 2011) whereby the wealth of 15 distinctive stops and fricatives collapses into a simple three-way stop contrast [p, t, k]. To borrow Lee's (2006a: 198) example, *nac^h* 'day', *nas* 'scythe' and *nat^h* 'face' are all realized as [nat]. In addition, there is a large number of other neutralization and assimilation processes affecting the surface form of both obstruents and sonorants (Sohn 1994), which further work to divorce the phonetic and phonemic levels.

B&K conducted independent analyses targeting both levels of representation. Their examination of abstract forms was based on the dictionary data described above, in which all underlying contrasts are maintained. Their analysis of surface forms was based on a collection of 12,500 speech syllables obtained from the Korean Telephone Conversation Speech Corpus (Ko 2003a, b) and from a

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

recorded Korean radio talk show segment. Syllables were identified through the phonetic surface form in which they appeared in the speech data, i.e. after the application of all phonological processes such as resyllabification, neutralization, and assimilation. In contrast, L&G employed what might be seen as a hybrid approach. They used dictionary entries as their starting point but converted them into the allophones resulting from coda obstruent neutralization (Lee 2006a: 198–199; see also pages 220–221, where all 76 VC sequences that entered the calculation are listed). Thus, for example, each of the three words *nac*^{h}, *nas* and *nat*^{h} was analyzed as *nat*.

L&G's hybrid approach is a possible alternative to B&K's separation of levels. However, it is not obvious why a left-branching effect should emerge in one but not the other approach. In order to clarify this question, we test the possible impact of the distinction between neutralized and non-neutralized coda obstruents in our new analysis below.

2.4 *Segmental inventory*

The Korean vowel and consonant inventories contain a few contrasts that are arguably obsolete. Decisions about the status of these cases may potentially affect phonotactic analyses. For the most part, B&K and L&G made identical assumptions. Neither team recognized the historical contrast between long and short vowels. Both assumed the contrast between the front vowels /e/ and /ɛ/.

A major complicating issue that was treated differently by B&K and L&G is the phonological status of the prevocalic glides /w/ and /j/. These occur in sequences like /ja/ and /wə/ in which they may be interpreted either as part of a diphthong (in a 'nucleus analysis') or as an independent onset consonant (in an 'onset analysis'). L&G adopted the nucleus view (Kang 2003). B&K did not commit themselves on this question and carried out three parallel analyses of their data: one treating glides as parts of diphthongs, another treating them as onsets, and a third one in which words containing prevocalic glides were excluded. Given that L&G's analysis is identical to one of B&K's subanalyses and that this subanalysis failed to bring out a CV effect, the disparate results are not attributable to any decision about the phonological status of glides. However, since we still regard the glide issue as unresolved, our new analysis below follows B&K's three-way approach.

2.5 *Statistical analysis*

The last point concerns the different statistical procedures adopted by B&K and L&G. To quantify the strength of association between adjacent segments in CVC syllables, both author teams obtained the frequency values *a*, *b*, *c* and *d* shown in (2) for all CV and VC sequences contained in their respective collections of CVC words.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

(2)

	V	Other V
C	*a*	*b*
Other C	*c*	*d*

The observed frequency of a `given phoneme sequence (value *a*) may be unexpectedly low or unexpectedly high in the light of the frequency with which each phoneme occurs in combination with other segments (values *b* and *c*). Either of these possibilities constitutes a probabilistic phonotactic effect. Where B&K's and L&G's procedures differ is on the question of how to quantify individual CV and VC effects and how to compare their combined strength.

B&K performed one-tailed Fisher Exact tests over the four frequency values in (2) for every CV and VC combination in order to identify the combinations which occur significantly more or less often than expected by chance. For example, their set of CVC syllables included 39 instances of the CV sequence /cʰə/, as in *cʰən* 'thousand'. To determine the probability of this observation, they used (i) the number of times /cʰ/ and /ə/ occur together as onset and nucleus (39 times; value *a*), (ii) the number of times /cʰ/ is followed by other vowels (97 times; value *b*), (iii) the number of times /ə/ is preceded by onsets other than /cʰ/ (376 times, value *c*), and (iv) the number of times onsets other than /cʰ/ combine with nuclei other than /ə/ (2301 times, value *d*). Fisher's Exact test takes these four values as input and yields an extremely small likelihood of obtaining this particular distribution ($p = .000014$). Thus, the observed frequency of 39 (value *a*) is significantly higher than the expected value (20 in this example). B&K used the Fisher Exact test for these calculations because in some cases the numbers per cell were very small, including values of zero. For cell values below 5, the better known Chi-squared test does not yield reliable results, and the non-parametric Fisher Exact test is generally preferred. In addition, given the large number of tests carried out on the same data set, the significance level was appropriately lowered using the Bonferroni correction. In the example above, the obtained probability is below the adjusted significance level of .000139 (.05 divided by 360, where 360 is the number obtained by multiplying 18 onset consonants with 20 vowels).

In a second step, B&K calculated the share of significant CV and VC pairs out of the total number of theoretically possible CV and VC pairs and compared these again using Fisher's Exact test. To illustrate, one of their subanalyses examined 360 theoretically possible CV pairs and 320 theoretically possible VC pairs. In the CV set, they found 16 pairs occurring either significantly more or significantly less often than expected, and 344 which did not differ from chance. In the VC set, 10 pairs showed such exceptional frequencies, while the other 310 did not. The Fisher Exact test reveals a chance distribution ($p = .244$). B&K found no significant difference between the rate at which significant associations occur in CV and VC sequences in any of their subanalyses. This led them to reject the hypothesis that Korean syllables exhibit a left- or right-branching phonotactic bias.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

L&G assessed the strength of association for each attested CV and VC combination by means of Pearson's r_Φ, a correlation statistic which Perruchet & Peereman (2004) advocated in their investigation of French syllables. The r_Φ formula is given in (3).

$$(3) \quad r_\Phi = \frac{ad - bc}{\sqrt{(a+b) \times (c+d) \times (a+c) \times (b+d)}}$$

To the extent that two given segments are correlated, in the sense that they do or do not tend to occur together, the r_Φ value will positively or negatively diverge from zero. Having determined two sets of r_Φ values, one for all CV associations and one for all VC associations, L&G then used a Mann–Whitney U-test (also known as the Wilcoxon rank sum test) to determine whether there was a significant overall difference in the r_Φ values. Their CV set displayed significantly greater absolute values overall, which they interpreted as supporting a left-branching phonotactic structure.

It is possible that L&G's statistical procedure is more sensitive than B&K's so that the former picks up a relatively weak effect which passes unnoticed in the latter. In B&K's mode of calculation, all sequences which did not reach statistical significance in the first round of calculations were discarded and did not enter the overall CV–VC comparison. In L&G's method, by contrast, the effect of all associations, no matter how strong each one was individually, was compared as a whole, which allows even weak associations to affect the overall distribution. Arguably, this makes L&G's method more sensitive than B&K's.

Given these methodological considerations, we decided to re-calculate the series of analyses reported in the B&K study using L&G's method of calculation. The results are presented at the beginning of the next section. We then go on to present our new Korean analysis, which addresses the other factors discussed above.

3. Korean syllable structure revisited

We begin with a re-analysis of B&K's Korean data in an attempt to determine if it was simply the different statistical procedures that caused the disparate findings. The datasets that entered into B&K's subanalyses were subjected to L&G's statistical procedure. Note that B&K's study consisted of 12 parallel subanalyses designed to control for a variety of alternative methodological decisions and their combinations. The results of the re-analysis are summarized in Table 1.

Columns 1, 2 and 3 of Table 1 specify the criteria which define each subanalysis (for definitions and justification of these criteria, see B&K 2010: 39–40). Column 4 gives the number of individual CV and VC types for which an r_Φ value was calculated in each subanalysis. The outcome of the overall CV–VC comparison is reported in column 5. In accordance with L&G's (2008: 159) mode of presentation, we provide the mean absolute r_Φ value followed by the standard deviation in parentheses. By comparing the CV and VC

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Analysis	Subanalysis	Status of pre-vocalic glides		N	Mean absolute r_Φ (SD)		Wilcoxon test
Lexicon analysis		Nucleus analysis	CV	209	0.032	(0.028)	$W=18051$, $p=.19$
			VC	160	0.029	(0.027)	
		Onset analysis	CV	147	0.033	(0.03)	$W=8383$, $p=.3$
			VC	106	0.03	(0.025)	
		Analysis without onglides	CV	137	0.032	(0.025)	$W=7639$, $p=.49$
			VC	106	0.031	(0.025)	
Speech syllable analysis	Token analysis	Nucleus analysis	CV	184	0.056	(0.069)	$W=7679$, $p=.81$
			VC	85	0.056	(0.06)	
		Onset analysis	CV	141	0.06	(0.072)	$W=3163$, $p=.28$
			VC	50	0.067	(0.065)	
		Analysis without onglides	CV	135	0.06	(0.073)	$W=2863$, $p=.11$
			VC	50	0.07	(0.066)	

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Type analysis 1	Nucleus analysis	CV	183	0.045	(0.04)	$W=8430$, $p=.27$
		VC	85	0.039	(0.034)	
	Onset analysis	CV	140	0.046	(0.043)	$W=3326$, $p=.6$
		VC	50	0.047	(0.039)	
	Analysis without onglides	CV	134	0.045	(0.039)	$W=3152$, $p=.54$
		VC	50	0.048	(0.038)	
Type analysis 2	Nucleus analysis	CV	154	0.056	(0.054)	$W=6301$, $p=.26$
		VC	75	0.047	(0.041)	
	Onset analysis	CV	129	0.055	(0.051)	$W=3366$, $p=.5$
		VC	49	0.05	(0.048)	
	Analysis without onglides	CV	123	0.054	(0.047)	$W=3175$, $p=.59$
		VC	49	0.051	(0.047)	

Table 1
Re-analysis of Berg & Koops' (2010) Korean data on the basis of Lee & Goldrick's (2008) statistical procedure.

means, it is possible to assess whether there is a tendency towards left- or right-branching. Column 6 reports the results of the Wilcoxon rank sum test carried out to determine whether the difference in absolute r_{Φ} values is statistically significant.

Table 1 presents a highly consistent picture. Not a single subanalysis yields a significant effect. Even if a different and arguably more sensitive statistical technique is used, no phonotactic asymmetry can be observed. These findings fully replicate B&K's (2010) results. We may conclude that the choice of one technique or the other did not cause the disparate results.

To address the other methodological issues raised in the preceding section, we performed a new lexicon analysis based on a smaller monolingual dictionary of Korean (Lee & Seo 1998). With approximately 50,000 entries, this dictionary is one fifth the size of that used by B&K. The CV and VC sequences contained in this new collection of CVC words were subjected to a series of subanalyses in order to examine every possible combination of relevant methodological decisions. As discussed above, these decisions encompass the choice of morphological root types (Section 2.2), neutralized or non-neutralized coda obstruents (Section 2.3), and the three ways of treating pre-vocalic glides (Section 2.4). Additionally, in each case we applied both statistical procedures. Tables 2A and 2B present the outcomes of the resulting 24 calculations. Table 2A provides the results calculated according to B&K's method, and Table 2B shows the results using L&G's procedure.

Columns 1, 2 and 3 of Table 2A specify the selection criteria applied to our collection of CVC syllables in each subanalysis. Column 4 gives the number of syllables included in each calculation, followed in parentheses by the proportion of excluded syllables. Columns 5 and 6 report the results of the statistical analysis of each CV and VC combination. Column 5 contains the number of CV and VC pairs whose frequency was significantly above or below chance level and, following the slash, the number of pairs that showed no significant effect. This is followed in parentheses by the percentage of significant pairs out of the total number of possible pairs. A comparison of the percentages reveals whether there is a trend towards a CV or VC effect. Column 6 reports the result of the one-tailed Fisher Exact test applied to the values in column 5. Table 2B is organized in the same way as Table 1.

The gist of Tables 2A and 2B can be summarized easily. None of the various combinations of data selection criteria produce a significant difference between the association strength of CV or VC pairs overall. Again, the two statistical procedures generate equivalent results. This conclusion confirms that of B&K's earlier study. There is no phonotactic evidence of left-branching in Korean CVC syllables. Our prior conclusion stands even if all potentially relevant methodological decisions are carefully considered and independently tested.

Finally, it is worth noting that the CV/VC asymmetry in Lee's (2006a) analysis is rather weak. L&G (2008: 158) report a value of $p<.05$. In order to be

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

certain about the exact strength of the effect, we performed the same statistical test as L&G (2008) over the r_Φ values listed in Appendix 1 of Lee's (2006a: 215–221) dissertation. Specifically, we performed a Wilcoxon rank sum test over the r_Φ values associated with the 152 Korean CV types and the 76 VC types.[2] The asymmetry only just reaches statistical significance ($p = .0438$). Such a value should be interpreted with caution because very small changes in the dataset could eliminate the significance of the result.

In fact, we did find an error in Lee's (2006a) dataset, which was carried over into L&G's (2008) article. Lee's (2006a) Appendix 1 lists all CV and VC types on which the overall r_Φ comparison is based. The CV list includes pairs containing 17 different vowels, while the VC list includes only pairs containing 13 different vowels. All VC types containing the diphthongs /wi/, /wɛ/, /wa/ and /we/ were omitted from Lee's (2006a) calculation.[3] Given the fact that the CV and VC pairs were drawn from the same collection of CVC words, the set of vowels should be identical across conditions. Taken together with the marginal statistical significance of their result, we cannot rule out the possibility that a rectification of this oversight might push the *p*-value above the .05 threshold.

4. Psycholinguistic evidence for left-branching in Finnish

After having established the absence of a phonotactic CV effect in Korean CVC syllables, we proceed to examine the generality of this result. After all, the phonotactic symmetry might be an idiosyncrasy of Korean and hence of limited cross-linguistic significance. However, as mentioned in the introductory section, B&K's explanation for this finding is of a language-independent nature. Their account predicts that other languages with a general left-branching bias should similarly fail to show a phonotactic CV effect. This prediction will now be put to the test.

There are only few languages besides Korean for which a left-branching syllable structure has been postulated, and even fewer for which the evidence for left-branching is compelling. Probably the best known case is Japanese. Kubozono (e.g. 1995) investigated the breakpoint in a classic blending task and found a strong preference for a body-coda structure. Unfortunately, it does not seem possible to compare this finding with the behaviour of other languages because Japanese has morae as immediate constituents of syllables and these morae impose a different kind of (rhythmic) organization on the language (Otake et al. 1993). Left-branching has also been hypothesized for Italian and Spanish syllables (Kubozono 1995). However, this claim is contradicted for

[2] The same values, i.e. 152 and 76, are given in L&G (2008: 159).

[3] Following B&K (2010), we use Sohn's (1994) romanization system to write Korean phonemes. Sohn's system differs from the way in which Korean phonemes appear in Lee's (2006a) dissertation. In terms of the symbols in Lee's appendix (pages 115–121) the missing vowels are *wi*, *wA*, *wa* and *O*.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Verb and adjective roots	Coda neutralization	Status of pre-vocalic glides	Number of syllables (% data reduction)		Significant/ non-significant pairs (% significant pairs)		Fisher Exact test
V and Adj roots Included	Coda obstruents not neutralized	Nucleus analysis	1205 (5%)	CV	5/283	(1.7%)	$p=.28$
				VC	2/254	(0.8%)	
		Onset analysis	1154 (9%)	CV	7/153	(4.4%)	$p=.18$
				VC	2/118	(1.7%)	
		Analysis without onglides	1086 (15%)	CV	6/138	(4.2%)	$p=.1$
				VC	1/119	(0.8%)	
	Coda obstruents neutralized	Nucleus analysis	1205 (5%)	CV	5/283	(1.7%)	$p=.63$
				VC	2/110	(1.8%)	
		Onset analysis	1154 (9%)	CV	7/153	(4.4%)	$p=.73$
				VC	2/54	(3.6%)	
		Analysis without onglides	1086 (15%)	CV	6/138	(4.2%)	$p=.71$
				VC	2/54	(3.6%)	

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

V and Adj roots excluded	Coda obstruents not neutralized	Nucleus analysis	972 (24%)	CV	4/284	(1.4%)	$p=.1$
				VC	0/224	(0%)	
		Onset analysis	920 (28%)	CV	3/157	(1.9%)	$p=.22$
				VC	0/104	(0%)	
		Analysis without onglides	858 (33%)	CV	3/141	(2.1%)	$p=.19$
				VC	0/104	(0%)	
	Coda obstruents neutralized	Nucleus analysis	972 (24%)	CV	4/284	(1.4%)	$p=.81$
				VC	1/111	(0.9%)	
		Onset analysis	920 (28%)	CV	3/157	(1.9%)	$p=.7$
				VC	1/55	(1.8%)	
		Analysis without onglides	858 (33%)	CV	3/141	(2.1%)	$p=.37$
				VC	0/56	(0%)	

Table 2A
New phonotactic analysis of Korean CVC syllables, following Berg & Koops' (2010) procedure.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Verb and adjective roots	Coda neutralization	Status of pre-vocalic glides	Number of syllables (% data reduction)		N	Mean absolute r_Φ (SD)		Wilcoxon test
V and Adj roots Included	Coda obstruents not neutralized	Nucleus analysis	1205 (5%)	CV	151	0.046	(0.038)	$W=10549$, $p=.15$
				VC	127	0.037	(0.028)	
		Onset analysis	1154 (9%)	CV	129	0.042	(0.037)	$W=6186$, $p=.59$
				VC	92	0.036	(0.027)	
		Analysis without onglides	1086 (15%)	CV	120	0.04	(0.034)	$W=5697$, $p=.69$
				VC	92	0.036	(0.026)	
	Coda obstruents neutralized	Nucleus analysis	1205 (5%)	CV	151	0.046	(0.038)	$W=6511$, $p=.62$
				VC	83	0.04	(0.028)	
		Onset analysis	1154 (9%)	CV	129	0.042	(0.037)	$W=3100$, $p=.43$
				VC	52	0.041	(0.026)	
		Analysis without onglides	1086 (15%)	CV	120	0.04	(0.034)	$W=2837$, $p=.35$
				VC	52	0.042	(0.027)	

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

V and Adj roots excluded	Coda obstruents not neutralized	Nucleus analysis	972 (24%)	CV	148	0.049	(0.039)	$W=8545$, $p=.18$
				VC	105	0.041	(0.031)	
		Onset analysis	920 (28%)	CV	127	0.044	(0.038)	$W=4863$, $p=.68$
				VC	74	0.04	(0.032)	
		Analysis without onglides	858 (33%)	CV	118	0.043	(0.034)	$W=4612$, $p=.51$
				VC	74	0.039	(0.03)	
	Coda obstruents neutralized	Nucleus analysis	972 (24%)	CV	148	0.049	(0.039)	$W=6334$, $p=.48$
				VC	81	0.042	(0.03)	
		Onset analysis	920 (28%)	CV	127	0.044	(0.038)	$W=3125$, $p=.57$
				VC	52	0.044	(0.031)	
		Analysis without onglides	858 (33%)	CV	118	0.043	(0.034)	$W=2890$, $p=.55$
				VC	52	0.044	(0.03)	

Table 2B
New phonotactic analysis of Korean CVC syllables, following Lee & Goldrick's (2008) procedure.

both languages by the work of Bertinetto (2001) who used the same experimental task as Kubozono. Further evidence corroborates right-branching in the Romance languages (Berg 2009: 206).[4] Finally, there is the case of Chinese, which has generated some controversy about whether it is an umbrella term for a group of separate languages or one language with a number of quite distinct dialects. Derwing (2007) reported left-branching effects for Minnan, the Chinese dialect spoken in Taiwan, while Shen (1993) argued for a right-branching organization of the syllable in Mandarin Chinese. In the light of the conflicting evidence in the Romance languages, it is difficult to judge whether Minnan and Mandarin Chinese do indeed have different branching directions or whether methodological artifacts are responsible for the (seemingly) contradictory results.

One of the safest bets for left-branching syllable structure currently appears to be Finnish. While linguistic analyses have not apparently addressed syllable-internal constituency (e.g. Karlsson 1999, Suomi, Toivanen & Ylitalo 2008, Sulkala & Karjalainen 2012), a convincing case for an asymmetry within the syllable can be made on the basis of psycholinguistic evidence. Two well-known types of psycholinguistic data will be considered – speech errors and language games. Niemi & Laine (1997) present an analysis of Finnish slips of the tongue. The strongest line of evidence comes from blend errors, in which the (most common) breakpoint is widely assumed to divide a given unit into its immediate constituents. Refer to (4). The slashes indicate the breakpoints.

(4) vamput ← va/nttuut × tu/mput 'mittens' (from Niemi & Laine 1997: 168)

Example (4) shows a blend of the two synonyms *vanttuut* and *tumput*. As is evident, the lexical items are split at the body-coda boundary. As there are no cases in Niemi & Laine's corpus which are broken up at the onset-rime boundary, slips like (4) for instance may be taken as evidence for a left-branching structure in Finnish syllables.

This conclusion is buttressed by errors in the linear order of speech, as exemplified in (5).

(5) viik sä menttii. For: meek + k sä vintti + i
go + CLITIC you attic + to
'Will you go upstairs?'
(from Niemi & Laine 1997: 167)

[4] On the basis of Ventura et al.'s (2001) experimental data, left-branching might be claimed to characterize Portuguese syllables. A division into body and coda occurred in CVC syllables which had an orthographic CVCV structure (with the final vowel grapheme being a mute <e>). In contrast, syllables with a match between orthographic and phonological units showed a right-branching bias. The nature of this pattern of results is not quite clear. It seems that Ventura et al.'s subjects divided the <CVCe> stimuli at the orthographic-syllable boundary. If this interpretation is correct, the evidence for left-branching in Portuguese is more apparent than real.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Case (5) illustrates the reversal of the two bodies *me* and *vi* (with some additional phonological accommodation). Again, no errors are found in which the rime is misordered.[5] There are quite a few tongue slips in Niemi & Laine's data in which the vowel is identical in the interacting bodies. Theoretically, this opens up the possibility of interpreting these cases as onset rather than body slips. However, their number is too high to warrant such an analysis. While a repeated vowel effect is known to facilitate consonantal slips of the tongue, it does not dramatically increase error rates (Stemberger 1994). The fact that identical vowels occur in two thirds of the Finnish exchange errors suggests that a large percentage of these seeming onset slips are in fact body slips, thereby providing further support for the left-branching analysis in Finnish.

The other type of data, viz. language games, resembles speech errors, even though the resemblance is a rather superficial one. Whereas language games require intentional alteration of a string of words, slips of the tongue are by definition unintentional. Two language games will be introduced, Pig German and *konttikieli*. As a matter of fact, Pig German generates an output which looks quite similar to exchange errors such as (5) above. The rule of the game states that the bodies of successive words are reordered (see e.g. Campbell 1980, Vago 1988).

(6) Saksalaisia hätyytettiin → häksäläisiä satuutettiin
'The Germans were attacked.'

(from Campbell 1980: 247)

Number (6) exemplifies the interchange of the bodies *sa* and *hä*. The vowel harmonic processes attendant upon this swapping need not concern us here. The interpretation of the Pig German data is as in the discussion of speech errors: the immediate constituents of syllables are bodies and codas.

The other language game is *konttikieli*, which has been described by Campbell (1980, 1986) and Vago (1985, 1988). It is similar to Pig German in that it also involves the reversal of word-initial bodies. The only difference is that the word *kontti* is inserted after each word, whereupon the exchange is always between *ko* and some other CV sequence.

Importantly, no language games which assign a role to the rime have been reported in the pertinent literature. It may therefore be concluded that the body is the most salient unit below the syllable level. This is tantamount to claiming that Finnish syllables are organized in a left-branching manner (see also Berg 2007).

Against the backdrop of the preceding analysis, we may now ask whether the phonotactic evidence is consistent with the evidence from speech errors and language games. B&K's account of the Korean data predicts that this is not the case. This prediction will be examined in the following section.

[5] The same is true of Hokkanen's (2001) Finnish error collection.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

5. A PHONOTACTIC ANALYSIS OF FINNISH CVC SYLLABLES

Our view of phonotactics was expounded in B&K (2010: 38–39). Critically, we consider combinatory rather than positional constraints. Whether or not a particular phoneme is allowed in a particular position is irrelevant to the internal structure of the syllable. As we are interested in possible asymmetries in syllable organization, we focus on constraints on the interaction of adjacent phonemes. The term 'interaction' is understood here as the likelihood with which a given phoneme A is followed or preceded by a given phoneme B, while factoring out the frequency of occurrence of the individual phonemes A and B.

Phonotactic analyses, especially those dealing with English (e.g. Malone 1936, Dell, Juliano & Govindjee 1993, Kessler & Treiman 1997), have focused primarily on monosyllabic words. It would seem natural to apply the same method to Finnish. However, this language runs short on monosyllabic words. An examination of 1,110 randomly selected entries in the small dictionary *Langenscheidt Universal-Wörterbuch Finnisch* (2004)[6] uncovered only 13 such items, amounting to 1.2%. Even high-frequency items such as *olla* 'to be' and *kyllä* 'yes' are disyllabic. This makes monosyllabic words an unsuitable data source. Exclusive reliance on such words would not only make a fine-grained statistical analysis impossible, but also create a sample that represents only a very small fraction of the lexicon that may not be representative of the phonotactics of Finnish words in general. We therefore decided to examine polysyllabic words without, however, ignoring the monosyllabic items.

Our focus on polysyllabic words necessitates a decision on which syllable types to select for closer scrutiny. Our choice fell on word-initial and word-final syllables. This move involves a complete analysis of disyllabic items and a partial analysis of longer items. An investigation of medial syllables is hampered by the fact that these syllables are difficult to compare across different word sizes. In structural terms, the medial syllable of, let us say, a trisyllabic word is not strictly comparable to the second or third syllable of a quadrisyllabic word. The focus on initial and final syllables also allows us to compare phonotactic effects at the beginning and end of words.

We extracted all entries containing either an initial or a final CVC syllable, or both, from Katara & Schellbach-Kopra (1997), a large monolingual Finnish dictionary containing more than 100,000 items. Our dictionary-based approach restricts the phonotactic analysis to words in their citation forms, which may roughly be equated with a phonemic level of representation. This means that the outcome of certain phonetic processes such as vowel epenthesis (e.g. *paljas* → *palajas* 'bare'), which affect syllable boundaries and which are commonly found in speech, cannot be taken into consideration. In other words, the analysis is confined to phonological CVC syllables. Note also that Finnish has

[6] All entries on every tenth double page (left and right), beginning with the second page, were taken into account.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

tautosyllabic and heterosyllabic vowel sequences. Of course, only syllables with genuine diphthongs such as /ai/ and /au/ were included in the analysis.

The extraction of CVC syllables from multisyllabic words required us to determine syllable boundaries. This is a relatively straightforward task as Finnish does not have ambisyllabic consonants. We employed the syllabification rules laid down in Sulkala & Karjalainen (2012). Syllable boundaries were inserted such that the subsequent syllable began with a single onset consonant. Geminate consonants were distributed across two syllables.

In addition to the position of a syllable within a word, we coded the data for three additional variables. First, a CVC syllable may be part of a morpheme which also occurs in other lexical entries. For example, two different compounds may share identical heads and hence identical word-final syllables (e.g. *soittoesitys* 'musical performance' and *kantaesitys* 'first performance'). There are two ways of dealing with these cases. Either the repeated morpheme is counted only once (type count) or as often as it occurs in different words (token count). The former option reflects the perspective of a type phonology, the latter option introduces a morphological token effect into the phonological analysis. Since it is not immediately obvious which option is to be preferred, we decided to examine both.

A second problem is that some syllables straddle a morphological boundary (e.g. a suffix boundary as in *kaksoset* 'twins', where the final /t/ marks plurality, even though *kaksose-* is not a free morpheme). Most cases of this kind are 'pluralia tanta' much like *scissors* in English. As it is not clear exactly how morphology interacts with phonotactics, we introduced separate categories for final syllables which contain a morpheme boundary and those which do not. Third, and primarily for the sake of completeness, we also included the very small set of Finnish monosyllabic CVC words and coded them separately.

In our analysis of the data, we tested all possible combinations of the different data selection criteria by running parallel tests on subsets of CVC syllables. Tables 3A and 3B report the results for the word-initial syllables, and Tables 4A and 4B those for the word-final syllables. The data were subjected to the two statistical treatments previously discussed. Tables 3A and 4A report the results of B&K's (2010) procedure, and Tables 3B and 4B those of L&G's (2008) method. The format of the tables is identical to that of the earlier ones.

Let us begin with the most important finding. In none of the sixteen subanalyses did we find evidence for left-branching. No matter which subset of the data or which of the two statistical procedures is considered, the outcome is never a CV effect. This stands in marked contrast to the psycholinguistic results discussed in the previous section.

We are thus left with a binary choice between a VC effect and no branching effect at all. Which of the two options is observed depends in the main on three factors: the scoring of repeated morphemes, the position of the syllable in a word and the statistical test chosen. All other factors are negligible. The overall result is that the likelihood of a VC effect increases (i) when syllables in repeated

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Repeated morphemes	Morpheme boundaries	Word types	Number of syllables (% data reduction)		Significant/ non-significant pairs (% significant pairs)		Fisher Exact test
Syllables in repeated morphemes counted separately	Syllables containing a morpheme boundary included	Polysyllabic words only	7313 (0.3%)	CV	37/490	(7%)	$p=.24$
				VC	44/483	(8.3%)	
		Mono- and polysyllabic words	7335 (0%)	CV	37/490	(7%)	$p=.32$
				VC	42/485	(8%)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	7161 (2.4%)	CV	38/489	(7.2%)	$p=.28$
				VC	44/483	(8.3%)	
		Mono- and polysyllabic words	7180 (2.1%)	CV	38/489	(7.2%)	$p=.32$
				VC	43/484	(8.2%)	
Syllables in repeated morphemes counted once	Syllables containing a morpheme boundary included	Polysyllabic words only	3803 (48.2%)	CV	15/512	(2.8%)	$p=.35$
				VC	12/515	(2.3%)	
		Mono- and polysyllabic words	3821 (47.9%)	CV	14/513	(2.7%)	$p=.42$
				VC	12/515	(2.3%)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	3760 (48.7%)	CV	14/513	(2.7%)	$p=.34$
				VC	11/516	(2.1%)	
		Mono- and polysyllabic words	3776 (48.5%)	CV	14/513	(2.7%)	$p=.34$
				VC	11/516	(2.1%)	

Table 3A
Phonotactic analysis of Finnish word-initial CVC syllables, following Berg & Koops' (2010) procedure.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Repeated morphemes	Morpheme boundaries	Word types	Number of syllables (% data reduction)		N	Mean absolute r_Φ (SD)		Wilcoxon test
Syllables in repeated morphemes counted separately	Syllables containing a morpheme boundary included	Polysyllabic words only	7313 (0.3%)	CV	291	0.026	(0.025)	$W = 34650$, $p = .66$
				VC	233	0.026	(0.024)	
		Mono- and polysyllabic words	7335 (0%)	CV	293	0.026	(0.024)	$W = 35217$, $p = .59$
				VC	234	0.025	(0.024)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	7161 (2.4%)	CV	285	0.026	(0.025)	$W = 32393$, $p = .84$
				VC	225	0.026	(0.026)	
		Mono- and polysyllabic words	7180 (2.1%)	CV	287	0.025	(0.025)	$W = 32978$, $p = .88$
				VC	228	0.026	(0.025)	
Syllables in repeated morphemes counted once	Syllables containing a morpheme boundary included	Polysyllabic words only	3803 (48.2%)	CV	290	0.023	(0.025)	$W = 29278$, $p = .02$
				VC	229	0.026	(0.023)	
		Mono- and polysyllabic words	3821 (47.9%)	CV	292	0.023	(0.025)	$W = 29847$, $p = .02$
				VC	231	0.026	(0.023)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	3760 (48.7%)	CV	285	0.024	(0.025)	$W = 28801$, $p = .06$
				VC	224	0.026	(0.024)	
		Mono- and polysyllabic words	3776 (48.5%)	CV	287	0.023	(0.025)	$W = 29143$, $p = .04$
				VC	227	0.026	(0.023)	

Table 3B
Phonotactic analysis of Finnish word-initial CVC syllables, following Lee & Goldrick's (2008) procedure.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Repeated morphemes	Morpheme boundaries	Word types	Number of syllables (% data reduction)		Significant/ non-significant pairs (% significant pairs)		Fisher Exact test
Syllables in repeated morphemes counted separately	Syllables containing a morpheme boundary included	Polysyllabic words only	4084 (0.4%)	CV	63/312	(16.8%)	$p = .08$
				VC	34/116	(22.7%)	
		Mono- and polysyllabic words	4102 (0%)	CV	63/327	(16.2%)	$p = .1$
				VC	33/123	(21.2%)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	2798 (31.8%)	CV	52/293	(15.1%)	$p = .5$
				VC	20/118	(14.5%)	
		Mono- and polysyllabic words	2816 (31.4%)	CV	52/308	(14.4%)	$p = .42$
				VC	19/125	(13.2%)	
Syllables in repeated morphemes counted once	Syllables containing a morpheme boundary included	Polysyllabic words only	731 (82.2%)	CV	9/336	(2.6%)	$p < .001$
				VC	17/121	(12.3%)	
		Mono- and polysyllabic words	749 (81.7%)	CV	9/351	(2.5%)	$p < .001$
				VC	15/129	(10.4%)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	649 (84.2%)	CV	12/318	(3.6%)	$p < .001$
				VC	17/115	(12.9%)	
		Mono- and polysyllabic words	667 (83.7%)	CV	12/333	(3.5%)	$p < .001$
				VC	16/122	(11.6%)	

Table 4A
Phonotactic analysis of Finnish word-final CVC syllables, following Berg & Koops' (2010) procedure.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Repeated morphemes	Morpheme boundaries	Word types	Number of syllables (% data reduction)		N	Mean absolute r_Φ (SD)		Wilcoxon test
Syllables in repeated morphemes counted separately	Syllables containing a morpheme boundary included	Polysyllabic words only	4084 (0.4%)	CV	172	0.075	(0.098)	$W=3678$, $p=.08$
				VC	51	0.122	(0.155)	
		Mono- and polysyllabic words	4102 (0%)	CV	179	0.073	(0.096)	$W=4253$, $p=.088$
				VC	56	0.115	(0.15)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	2798 (31.8%)	CV	136	0.102	(0.131)	$W=2677$, $p=.541$
				VC	42	0.139	(0.173)	
		Mono- and polysyllabic words	2816 (31.4%)	CV	145	0.097	(0.127)	$W=3234$, $p=.601$
				VC	47	0.128	(0.165)	
Syllables in repeated morphemes counted once	Syllables containing a morpheme boundary included	Polysyllabic words only	731 (82.2%)	CV	139	0.067	(0.068)	$W=2573$, $p=.048$
				VC	46	0.111	(0.121)	
		Mono- and polysyllabic words	749 (81.7%)	CV	149	0.064	(0.065)	$W=3098$, $p=.032$
				VC	52	0.106	(0.114)	
	Syllables containing a morpheme boundary excluded	Polysyllabic words only	649 (84.2%)	CV	129	0.071	(0.077)	$W=1767$, $p=.005$
				VC	39	0.141	(0.147)	
		Mono- and polysyllabic words	667 (83.7%)	CV	139	0.068	(0.075)	$W=2254$, $p=.005$
				VC	45	0.131	(0.138)	

Table 4B

Phonotactic analysis of Finnish word-final CVC syllables, following Lee & Goldrick's (2008) procedure.

morphemes are counted only once, (ii) when the syllable occurs in word-final rather than word-initial position, and (iii) when L&G's statistical procedure is used instead of B&K's procedure. These factors cross-cut one another. Let us consider each of them in turn.

A somewhat unexpected result displayed in Tables 3A–B and 4A–B is the emergence of a VC effect in three of the four conditions counting repeated morphemes once while no such effect emerged in any of the other conditions. It is striking that the number of words containing identical morphemes is quite high. As the column 'Number of syllables' in Tables 3 and 4 indicates, a considerable proportion of the data is lost when repeated morphemes are excluded (approximately every other initial syllable and more than 80% of the final syllables). Inspection of the column 'Significant/non-significant pairs' in Table 4A further reveals that the VC effect is brought about by a massive reduction of significant CV pairs but a much smaller reduction of significant VC pairs in word-final syllables as we move from the token to the type count. However, no such difference is observed in word-initial syllables (see Table 3A). Thus, this differential reduction of significant adjacent-segment pairs is a local effect. Because significant effects emerge only in the type count, the following analysis will be restricted to this set of conditions.

There is weak support for a VC effect in initial syllables but stronger support for the same effect in final syllables. Two lines of evidence back up this claim. In final syllables, the VC effect is brought to the fore by both B&K's and L&G's procedure. However, it manifests itself only inconsistently in initial syllables. By and large, it emerges using L&G's method but fails to do so using B&K's method. The other line of evidence pertains to effect size. As noted in Section 2.5 above, the r_{Φ} values measure the association strength of segments, so they may be directly inspected for effect size. The column 'Mean absolute r_{Φ} values' shows a relatively small difference between CV and VC pairs in initial syllables (0.003 or less for all four conditions; see Table 3B). By contrast, this difference is much larger in final syllables (between 0.049 and 0.07; see Table 4B). In two of the four conditions, the r_{Φ} values for VC pairs are almost double those for CV pairs (e.g. 0.0147 vs. 0.077). Hence, as revealed by association strength, the VC bias is stronger in final than in initial syllables.

The two statistical tests return relatively similar results. It is reassuring that they yield identical results for word-final syllables. For word-initial syllables, by contrast, Fisher's Exact test generates non-significant results throughout whereas Pearson's test produces non-significant results for the token count but significant results for most of the type-count conditions (see above). It should, however, be added that one of the four conditions yields only a marginally significant result ($p = .06$) and another yields a weakly significant result ($p = .04$).

How can we deal with this inconsistency? In our earlier discussion (see Section 2.5), we noted the possibility that L&G's method may be a more powerful indicator of phonotactic CV/VC asymmetries because it does not restrict itself to comparing only the significant cases of phonotactic association. Rather, it takes

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

all associations into account, regardless of their individual strength. We may now confirm this suspicion and argue that L&G's procedure is more sensitive than B&K's. Whenever B&K's method yields significant results, so does L&G's. However, the opposite is not true. L&G's test may yield significant results where B&K's does not.

To summarize Sections 4 and 5, we have hit upon a most curious pattern of results regarding the internal structure of Finnish CVC syllables. While the psycholinguistic evidence points towards left-branching, the phonotactic evidence argues for a largely symmetrical organization of word-initial syllables (with some traces of a VC effect) and a right-branching bias in word-final syllables (type count). This result poses the greatest challenge one can imagine: all possible options are attested. In the next section, an attempt will be made to meet this challenge.

6. Theoretical discussion

We begin with a discussion of the Korean data and then turn to the Finnish data. Our primary objective is to provide a unified account which accommodates the seemingly inconsistent evidence provided by the two languages. The upshot of this discussion is a fundamental reconsideration of the role of phonotactics in the analysis of sub-syllabic structure. Specifically, we suggest severing the widely accepted link between phonotactic constraints and branching direction. This also leads us to reject the idea that phonotactic effects allow speakers to infer sub-syllabic structure in the first place (L&G 2008). Consequently, the root cause of phonological right- or left-branching structure must lie elsewhere. We discuss a novel perspective based on phonetic cues, notably CV coarticulation.

6.1 *An account of the Korean and Finnish data*

We believe that the debate over the phonotactic structure of Korean monosyllables can be put to rest. There is now fairly robust evidence for the lack of a phonotactic asymmetry in Korean CVC words – a result on which all our sub-analyses converge. This finding runs counter to the highly consistent experimental evidence arguing for a left-branching analysis of Korean syllables (see Yoon & Derwing 2001 and the references cited above), which receives additional support from reduplication patterns (Jun 1994).

How can the phonotactic and the experimental evidence be reconciled? As noted at the beginning of this paper, B&K's (2010: 44–46) earlier discussion of this question suggested that the criteria traditionally used to determine branching direction may contain inherent biases. In particular, the phonotactic criterion was hypothesized to have a universal right-branching bias. If this right-branching bias meets a (phonologically) left-branching language like Korean, and if the two biases are of similar strength, they effectively cancel each other out. The net result is the lack of a CV or VC effect, as we observed in Korean CVC words.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

How can this mode of reasoning be extended to Finnish, where the empirical picture is slightly different? Specifically, how do we explain the relative lack of a phonotactic asymmetry in Finnish word-initial syllables and a VC effect in word-final ones? We suggest that the postulated right-branching bias is sensitive to word position: while word-final syllables display a phonotactic VC effect, word-initial syllables are more symmetrical. To see why this should be the case, we need to reconsider the nature of phonotactic constraints in the context of polysyllabic words.

Let us adopt the dynamic perspective of word recognition. Lexical processing begins immediately upon hearing the first segment of a word (Marslen-Wilson 1975) and then proceeds in a stepwise fashion whereby the number of mental addresses that match the auditory input quickly diminishes (Marslen-Wilson & Tyler 1980). In fact, lexical access is often unambiguously possible before the end of a word has been perceived because the incomplete phoneme string already diverges from all others in the lexicon. It follows that a phonotactic constraint placed beyond the uniqueness point implies a loss of dispensable phonological contrasts. Hence, the later in a word phonotactic constraints occur, the lower the threat to the distinctive function of phonemes.

There is another way in which the listeners' perspective favours word-final phonotactic constraints. As the listeners' task is to identify the unfolding word as quickly as possible, it can be assumed that they will exploit every available cue to the identity of the unfolding word, including the immediately upcoming segments. Phonotactic constraints provide such a cue in so far as they increase the predictability of upcoming segments. They do so by decreasing the number of following possibilities. To take a hypothetical example, upon hearing /t/, listeners who know that /t/ is rarely or never followed by /i/ will be able to adjust their expectations of possible continuations and hence be more likely to venture a correct guess at the unfolding word. Note that this strategy is least effective at the word onset and more effective the further advanced the word recognition process is. At an early stage of word recognition, the number of competing lexical candidates is typically still high. The small amount of added predictability that a phonotactic effect provides will be of very limited value in this situation because the cohort of lexical candidates is so large that taking a successful guess at the word is not a realistic option. However, at later stages of hearing a word, especially as the process of lexical access nears completion, a guess is more likely to be correct if the pool of lexical candidates is small.

We have offered two reasons why phonotactic constraints are more effectively placed far from the onset and as close as possible towards the end of words. This is consistent with our results of Finnish phonotactics. In word-initial syllables, the inherently right-branching nature of the phonotactic bias is relatively weak. It is just strong enough to counteract the psycholinguistic left-branching bias. As in Korean monosyllables, the clash of these opposed forces results in a largely balanced distribution of phonotactic effects. At the end of Finnish polysyllables, however, the phonotactic bias exerts itself more forcefully. In fact, it is so strong

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

that it overcomes the language's general left-branching bias. As a consequence, word-final syllables evince a phonotactic VC effect.

Convincing alternatives to the above processing account are hard to find. It is true that there are phonological differences between initial and final syllables, but it is doubtful that these are at the bottom of the phonotactic patterns. In particular, initial syllables have been claimed to be more perceptually salient and to code a higher number of contrasts than final ones (e.g. Beckman 1997). Let us assume, for the sake of the argument, that this is true of Finnish. This assumption ties in well with the fact that Finnish word-initial syllables are always stressed and word-final syllables always unstressed, given that stressed syllables are 'perceptual anchors' (e.g. Cole & Jakimik 1978). A possible hypothesis that emerges from this difference is that the absence of stress leads to a VC-effect and the presence of stress to no such effect. As far as we can tell, there is no evidence for such a claim in Finnish. If anything, the opposite is more likely to be true. It is well-known that segmental structure is more complex in stressed than in unstressed syllables. Analogically, we would expect suprasegmental structure to behave likewise. Since Finnish syllables show the opposite pattern, we are reluctant to hold stress (or contrast, for that matter) accountable for the empirical differences reported above.

6.2 *Severing the link between phonotactic constraints and branching direction*

The results of our analyses demonstrate a striking disconnect between phonotactic constraints and sub-syllabic constituency.[7] Judging by all the evidence other than the phonotactic criterion, both Korean and Finnish have 'basically' left-branching syllables. However, nowhere is this left-branching nature reflected in the phonotactic structure of CVC syllables. At best, it can be observed indirectly where the right-branching phonotactic bias neutralizes a 'basically' left-branching bias, resulting in the lack of a phonotactic asymmetry. Thus, the phonotactic criterion fails to predict phonological branching direction. Phonotactics and phonological branching structure should accordingly be viewed as independent phenomena.

The reason why phonotactics has loomed large in previous analyses of branching direction now emerges as a historical accident. The languages on which earlier analyses were performed are all basically right-branching ones. As the inherent right-branching bias of phonotactic constraints happened to match the right-branching syllable structure of languages such as English and German, students of these languages (including ourselves) could be easily tempted into treating phonotactics as valid evidence for branching direction. It is time to break with tradition on this point.

[7] We are therefore reluctant to accept Grimes's (2010) phonotactically-based conclusion that Hungarian has a flat syllable structure. The presence of phonotactic symmetry within the Hungarian syllable, alongside some trends towards right-branching, ties in rather nicely with the above results for Korean and Finnish.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

6.3 *A phonetic account of sub-syllabic branching direction*

Lee & Goldrick's (2008) short-term memory experiments demonstrate that individual CV and VC sequences which are strongly associated phonotactically are recalled more easily than weakly associated ones. This supports the notion that speakers keep track of the probabilities with which segments co-occur within syllables. What is less clear is whether their knowledge of segment-level phonotactic effects can give rise to a general phonological branching structure, for example a body node, as Lee & Goldrick (2008, 2011) propose. As long as the phonotactic evidence dovetails with other indicators of branching structure, their account appears convincing. However, when the two phenomena are at odds with each other, as we have shown here for Korean and Finnish, their theory falls apart.

With phonotactics removed as an explanatory factor, what is responsible for phonological left-branching in languages such as Korean and Finnish? In the following, we will discuss and reject one explanation that has been advanced in the relevant literature on Korean and subsequently suggest a novel one. It should be noted from the outset that our discussion of branching direction and its underlying causes is limited to the phonetic-phonological level. That is, we will not consider possible spill-over effects from other levels, such as 'cross-level harmony' between branching direction asymmetries in syntax and other levels.

Kim (2007) put forward an explanation of the body-coda preference which is based on the frequency of different syllable types. He argues that because CV syllables occur in Korean far more frequently than CVC, V, or VC syllables, the higher exposure to CV syllables leads Korean speakers to treat CV as a unit. Note that in this proposal language users are implicitly assumed to equate an entire CV syllable with the body of a CVC syllable. While not implausible, this hypothesis cannot be held up. It is well-known that there is a strong asymmetry between CV and VC syllables. Typologically, almost all languages have CV syllables and a sizeable number of them have only CV syllables while only very few have only VC syllables (e.g. Kaye & Lowenstamm 1981, Clements & Keyser 1983, Breen & Pensalfini 1999). Moreover, research on child language acquisition has shown that CV syllables are acquired prior to VC syllables (e.g. Levelt, Schiller & Levelt 2000). The attempt of relating the 'unmarked' nature of CV syllables to the issue of sub-syllabic constituency would lead us to expect a predominance of left-branching in the phonologies of the world's languages. However, the opposite is true (Berg 2009). To give only one example, Spanish has an absolute majority of CV syllables (Lloyd & Schnitzer 1967) but shows relatively strong signs of being a phonologically right-branching language. Thus, the predominance of CV syllables does not provide an account of branching direction.

We pursue this line of argument further and focus on CVC syllables in Finnish. The general logic of the frequency argument is that units which are used together fuse together (Bybee & Scheibman 1999: 578). The first subanalysis of Table 4A

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

above was based on 4,084 CVC syllable tokens, yielding 4,084 CV and 4,084 VC tokens. The CV sequences distribute across 172 CV types and the VC sequences across 51 VC types. Hence, the average token frequency of CV types is 23.7 while that of VC types is 80.1. This difference leads us to expect VC to be a more cohesive unit in Finnish than CV is. However, this prediction is incompatible with the left-branching nature of Finnish syllables (see Section 4). We thus cannot help severing the link between token frequency and branching direction.

We propose to shift the burden of explanation from phonotactics to phonetics. Let us return to the essence of branching direction and recall that branching structure is a hierarchical way of representing varying degrees of cohesion. For example, a right-branching syllable structure implies more cohesiveness between V and C than between C and V. At the phonetic level, cohesiveness can be brought about by coarticulation, i.e. the extent to which two adjacent segments overlap in terms of articulatory gestures and their acoustic consequences. If it can be shown that Korean CV sequences evince a particularly high degree of coarticulation, we might have a phonetic basis for the left-branching bias.

Evidence to this effect comes from the literature on the phonetics of the three-way contrast of the Korean voiceless stops, often called tense, lax, and aspirated, which only exists in onset position. Cho, Jun & Ladefoged (2002) studied the extent to which the stop contrasts are reflected in the voice quality of the following vowel. They found that coarticulatory cues, specifically short- and mid-range spectral slope as well as fundamental frequency (f0), co-vary systematically with the onset stop type. As Cho et al. discuss in detail, while the existence of CV coarticulation in f0 is not surprising, the temporal extent to which the consonantal information is carried over into the vowel in Korean is unusual. Crucially, the relevant differences are not only seen near the vowel onset, as regularly found in other languages, but spread well into the vowel. Similarly, Kim, Beddor & Horrocks (2002) found that phonatory cues contained in the following vowel, in particular f0, are especially strong in Korean, with as much as a 30 Hz difference in f0 distinguishing the lax from the tense and aspirated stops even at the vowel offset. The authors also tested the role of the vowel in the perception of the stop contrasts and found that Korean listeners are able to correctly identify manner of articulation when the stop portion is artificially removed. Moreover, when presented with contradictory information in the form of cross-spliced CV stimuli, the vocalic information usually overrides the consonantal information.

We submit that coarticulation is a better candidate for giving rise to sub-syllabic constituency than is phonotactics. To begin with, coarticulation is a given. Gestural overlap is an inevitable side effect of fluent articulation. The links which coarticulatory effects create between adjacent phonemes are directly observable in the phonetic signal. This means that groupings of segments established by coarticulation, especially where the degree of coarticulation is high, need not be 'discovered'. They are part and parcel of speech processing. Strongly coarticulated segments can thus be seen as natural processing units. In addition, coarticulatory effects are local in the sense that they can be observed

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

in individual words in ignorance of all others. Remarkably, phonotactic effects are entirely different. Phonotactic constraints can only be defined against the background of the complete lexicon. Nothing at all can be said about the phonotactic structure of an individual word viewed in isolation. In this sense, phonotactic constraints are of a global nature. They represent the outcome of a computational process over all items in a lexicon. This derivative nature of phonotactics contrasts sharply with the local nature of coarticulation. It may thus be argued that coarticulation is the simpler, the more salient, the more reliable and probably also the stronger cue for branching direction than phonotactic constraints.

We therefore tentatively suggest that the strong coarticulation found in Korean CV units may underlie the left-branching nature of CVC syllables. To be sure, such an account is no more than a first attempt at a phonetic explanation. The high degree of coarticulation discussed above has only been demonstrated for CV sequences in which the onset consonant is a stop. While it is true that the 12 Korean stops make up the bulk of the 19 consonants, our case would be strengthened if it could be demonstrated that the same holds for all CV sequences. Also, the phonetic studies cited above only speak to CV coarticulation. A direct comparison between CV and VC coarticulation would be desirable. However, this is impossible because, as mentioned, the relevant stop contrasts are all neutralized in coda position.

The hypothesis that coarticulation is a determinant of sub-syllabic structure generates predictions for other languages. For example, it leads us to expect that coarticulation within VC units is stronger than within CV units in phonologically right-branching languages. There is some evidence to bear this out. As summarized by Browman & Goldstein (1995), phonetic research on English sonorants shows that there is more temporal overlap between the vowel and the consonant gesture in nasals and laterals in coda position than in nasals and laterals in onset position (see Sproat & Fujimura 1993 on laterals, Krakow 1999 on nasals).

For obstruents in CV sequences the picture is less clear, however. In the case of stop–vowel sequences, it is well known that the formant transitions near the vowel onset contain crucial information about the stop's place of articulation (e.g. Liberman 1957, Liberman et al. 1967). Unfortunately for our purposes, phonetic studies of CV coarticulation have rarely directly compared the relative degree of coarticulation in tautosyllabic CV and VC sequences. A notable exception is the study by Sussman et al. (1997), who report less coarticulation between vowels and coda stops than between vowels and onset stops (see also Tabain, Breen & Butcher 2004). The notion that vowels and onset consonants form a unit is also central to the work of Nam & Saltzman (2003) and Nam, Goldstein & Saltzman (2010), who have modelled syllables as self-organizing structures in which CV patterns are intrinsically more stable than other patterns of gestural organization. Thus, pending further work on the hypothesized link between coarticulation and syllable structure, our line of argument has to remain highly tentative.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Finally, we hasten to add that the available evidence from coarticulation pertains only to Korean. We are unaware of a comparable study of CV coarticulation in Finnish, let alone a study comparing coarticulation in Finnish CV and VC sequences. We would not even want to commit ourselves to the claim that coarticulation is the only possible basis from which to infer sub-syllabic branching structure. Other phenomena such as the relative timing of consonants and vowels in CV and VC units, as seen in rime isochrony in English, might also play a part. The relevant parameters need not all be phonetic, but could also be phonological or psycholinguistic in nature.

7. Conclusion

A major objective of this paper was a modest test of our hypothesis of a universal phonotactic VC bias. We examined the prediction that no language should display a general phonotactic CV effect in CVC syllables. This prediction was borne out in our case study of Finnish, despite the fact that this language is one of the most likely candidates to display such an effect given prior assumptions about the relationship between phonotactic constraints and branching direction. Thus, we can now be more confident that the posited phonotactic VC bias will always exert itself, either directly or indirectly, no matter what decision a language has come to with regard to its sub-syllabic structure.

The analysis of Finnish has led us to extend the scope of probabilistic phonotactic analysis to polysyllabic words. We found that Finnish not only lacks a phonotactic CV effect but actually shows a VC effect in word-final syllables. Our theoretical interpretation of this effect has been conducive to the hypothesis that the phonotactic VC bias may not only be universal but also be universally stronger toward the end of words than at word beginnings. The predictions from this hypothesis are clear and certainly call for further testing. To do so, it is not even necessary to study languages with an 'unusual' syllable structure such as Korean or Finnish. Any language that uses CVC syllables in polsyllabic words, including, for example, English, can serve as a test case.

A somewhat surprising overall conclusion of the present paper is that phonotactics is a poor indicator of sub-syllabic structure. Other lines of evidence, whether naturalistic or experimental, appear to have a more direct bearing on the question of sub-syllabic constituency. Conceivably, these lines of evidence are the consequence rather than the cause of syllable structure. While we do not deny that there is an interaction between phonotactics and branching direction, the former cannot serve as the basis for explaining the latter. Instead, each requires its own explanation.

We have argued that the phonotactic VC bias and its variable strength in different word positions is grounded in the dynamics of speech processing, specifically the simultaneous goals of maximizing the power of phonemic contrasts and facilitating lexical access. This conclusion seems relatively secure. However, what determines left-branching in the syllabic phonology of some languages remains

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

something of a mystery. While phonotactics can be ruled out as an explanatory factor, a viable explanation of left-branching in CVC syllables seems to be a long way away.

We envisage two directions for future research. Despite a flurry of work on the role of phonotactics in language acquisition (e.g. Zamuner, Gerken & Hammond 2004), adult language processing (e.g. Walker & Dell 2006) and theoretical phonology (e.g. Hayes & Wilson 2008), we still have astonishingly few data on the phonotactic structure of different languages, in particular on relative phonotactic constraints. As noted before, the focus has typically been on monosyllabic words. However, our investigation of Finnish demonstrated that polysyllabic words have to be studied in their own right. It is far from clear that patterns observed in monosyllabic words carry over to polysyllabic ones. Furthermore, previous work has been limited to uninflected words. While this methodological decision may be justified for English in view of its impoverished inflectional morphology, it eliminates a very high percentage of word tokens in strongly inflected languages to the point of possibly creating a non-representative sample. Note that it is not at all self-evident that inflected words show the same phonotactic biases as uninflected ones. On the phonetic plane, we need more studies comparing the extent of coarticulation between CV and VC portions of CVC syllables in different languages. It is hoped that advances in these areas will provide a better understanding of the intricate relationship between coarticulation, phonotactics, and sub-syllabic structure.

REFERENCES

Anderson, Sven & Robert Port. 1994. Evidence for syllable structure, stress and juncture from segmental durations. *Journal of Phonetics* 22, 283–315.

Beckman, Jill N. 1997. Positional faithfulness, positional neutralization and Shona vowel harmony. *Phonology* 14, 1–46.

Berg, Thomas. 2007. A typology of suprasegmental structure. *MIT Working Papers in Linguistics* 53, 53–73.

Berg, Thomas. 2009. *Structure in language: A dynamic perspective*. New York: Routledge.

Berg, Thomas & Christian Koops. 2010. The interplay of left- and right-branching effects: A phonotactic analysis of Korean syllable structure. *Lingua* 120, 35–49.

Bertinetto, Pier Marco. 2001. Blends and syllable structure: A four-fold comparison. In Mercé Lorente, Núria Alturo, Emili Boix, Maria-Rosa Lloret & Lluis Payrató (eds.), *La gramática i la semántica en l'estudi de la variació*, 59–112. Barcelona: Promociones y Publicaciones Universitarias.

Breen, Gavan & Rob Pensalfini. 1999. Arrernte: A language with no syllable onsets. *Linguistic Inquiry* 30, 1–25.

Browman, Catherine P. & Louis Goldstein. 1995. Gestural syllable position effects in American English. In Fredericka Bell-Berti & Raphael J. Lawrence (eds.), *Producing speech: Contemporary issues*, 19–33. New York: AIP Press.

Bybee, Joan & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37, 575–596.

Campbell, Lyle. 1980. The psychological and sociological reality of Finnish vowel harmony. In Robert M. Vago (ed.), *Issues in vowel harmony*, 245–270. Amsterdam: John Benjamins.

Campbell, Lyle. 1986. Testing phonology in the field. In John J. Ohala & Jeri J. Jaeger (eds.), *Experimental phonology*, 163–173. New York: Academic Press.

Cho, Taehong, Sun-Ah Jun & Peter Ladefoged. 2002. Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics* 30, 193–228.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Clements, George N. & Samuel Jay Keyser. 1983. *CV phonology: A generative theory of the syllable*. Cambridge, MA: MIT Press.
Cole, Ronald A. & Jola Jakimik. 1978. Understanding speech: How words are heard. In Geoffrey Underwood (ed.), *Strategies of information processing*, 67–116. London: Academic Press.
De Cara, Bruno & Usha Goswami. 2002. Similarity relations among spoken words: The special status of rimes in English. *Behavior Research Methods, Instruments, & Computers* 34, 416–423.
Dell, Gary S., Cornell Juliano & Anita Govindjee. 1993. Structure and content in language production: A theory of frame constraints in phonological speech errors. *Cognitive Science* 17, 149–195.
Derwing, Bruce L. 2007. What's in CVC-like things? In Maria-Josep Solé, Patrice S. Beddor & Manjari Ohala (eds.), *Experimental approaches to phonology*, 325–338. Oxford: Oxford University Press.
Fudge, Erik. 1987. Branching structure within the syllable. *Journal of Linguistics* 23, 359–377.
Grimes, Stephen M. 2010. *Quantitative investigations in Hungarian phonotactics and syllable structure*. Ph.D. dissertation, Indiana University.
Hayes, Bruce & Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39, 379–440.
Hokkanen, Tapio. 2001. *Slips of the tongue: Errors, repairs, and a model*. Helsinki: Finnish Literature Society.
Jun, Jongho. 1994. Metrical weight consistency in Korean partial reduplication. *Phonology* 11, 69–88.
Kang, Kyung-Shim. 2003. The status of onglides in Korean: Evidence from speech errors. *Studies in Phonetics, Phonology and Morphology* 9, 1–15.
Kapatsinski, Vsevolod. 2009. Testing theories of linguistic constituency with configural learning: The case of the English syllable. *Language* 85, 248–277.
Kaplan, Abby. 2011. How much homophony is normal? *Journal of Linguistics* 47, 631–671.
Karlsson, Fred. 1999. *Finnish: An essential grammar*. London: Routledge.
Katara, Pekka & Ingrid Schellbach-Kopra. 1997. *Suomi–saksa-suursanakirja* [Finnish–German dictionary]. Poorvo: Söderström.
Kaye, Jonathan D. & Jean Lowenstamm. 1981. Syllable structure and markedness theory. In Adriana Belletti, Luciana Brandi & Luigi Rizzi (eds.), *Theory of markedness in generative grammar*, 287–315. Pisa: Scuola Normale Superiore.
Kessler, Brett & Rebecca Treiman. 1997. Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language* 37, 295–311.
Kim, Hyunsoon & Allard Jongman. 1996. Acoustic and perceptual evidence for complete neutralization of manner of articulation. *Journal of Phonetics* 24, 295–312.
Kim, Minsu, Yengkun Ko, Hongpin Im & Sungcay Lee (eds.). 1991. *Kukətɛsacən* [Korean dictionary]. Seoul: Kum Sung Publishing Co.
Kim, Mi-Ryoung, Patrice S. Beddor & Julie Horrocks. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30, 77–100.
Kim, Young-Suk. 2007. Phonological awareness and literacy skills in Korean: An examination of the unique role of body-coda units. *Applied Psycholinguistics* 28, 69–94.
Kim-Renaud, Young-Key. 1986. The syllable in Korean phonology. In Young-Key Kim-Renaud (ed.), *Studies in Korean linguistics*, 31–44. Seoul: Hanshin.
Ko, Eon-Suk. 2003a. *Korean telephone conversations: Transcripts*. Philadelphia, PA: Linguistic Data Consortium.
Ko, Eon-Suk. 2003b. *Korean telephone conversations: Speech*. Philadelphia, PA: Linguistic Data Consortium.
Krakow, Rena A. 1999. Physiological organization of syllables: A review. *Journal of Phonetics* 27, 23–54.
Kubozono, Haruo. 1995. Perceptual evidence for the mora in Japanese. In Bruce Cornell & Amalia Arvaniti (eds.), *Papers in Laboratory Phonology IV*, 141–156. Cambridge: Cambridge University Press.
Langenscheidt Universal-Wörterbuch Finnisch. 2004. Berlin: Langenscheidt.
Lee, Chang Hwan, Yoonhyoung Lee & Kyungil Kim. 2010. The role of antibody in Korean word recognition. *Journal of Psycholinguistic Research* 39, 457–464.
Lee, Sang-seop & Sang-kyu Seo. 1998. *Yonsei Korean dictionary*. Seoul: Doosan Dong'a.
Lee, Yongeun. 2006a. *Sub-syllabic constituency in Korean and English*. Ph.D. dissertation. Northwestern University, Evanston.

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

Lee, Yongeun. 2006b. Statistical regularities in the distribution of consonants and vowels in the Korean lexicon – their implications for the internal structure of Korean syllables. In Jacqueline Bunting, Sapna Desai, Robert Peachey, Christopher Straughn & Zuzana Tomková (eds.), *Papers from the main session of the 42nd Meeting of the Chicago Linguistic Society* (CLS 42), 197–211. Chicago: Chicago Linguistic Society.
Lee, Yongeun & Matthew Goldrick. 2008. The emergence of sub-syllabic representations. *Journal of Memory and Language* 59, 155–168.
Lee, Yongeun & Matthew Goldrick. 2011. The role of abstraction in constructing phonological structure. Presented at the Annual Meeting of the LSA, Pittsburgh, PA.
Levelt, Clara C., Niels O. Schiller & Willem J. M. Levelt. 2000. The acquisition of syllable types. *Language Acquisition* 8, 237–264.
Liberman, Alvin M. 1957. Some results of research on speech perception. *Journal of the Acoustical Society of America* 29, 117–123.
Liberman, Alvin M., Franklin S. Cooper, Donald P. Shankweiler & Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74, 431–461.
Lloyd, Paul M. & Ronald D. Schnitzer. 1967. A statistical study of the structure of the Spanish syllable. *Linguistics* 37, 58–72.
Malone, Kemp. 1936. The phonemic structure of English monosyllables. *American Speech* 11, 205–218.
Marslen-Wilson, William D. 1975. Sentence perception as an interactive process. *Science* 189, 226–228.
Marslen-Wilson, William D. & Lorraine K. Tyler. 1980. The temporal structure of spoken language understanding. *Cognition* 8, 1–71.
Nam, Hosung, Louis Goldstein & Elliot Saltzman. 2010. Self-organization of syllable structure: A coupled oscillator model. In Iona Chitoran, François Pellegrino, Egidio Marisco & Christophe Coupé (eds.), *Approaches to phonological complexity*, 1–29. Berlin: Mouton de Gruyter.
Nam, Hosung & Elliot Saltzman. 2003. A competitive, coupled oscillator model of syllable structure. *15th International Congress of Phonetic Sciences* (ICPhS 15), Barcelona, 2253–2256.
Niemi, Jussi & Matti Laine. 1997. Slips of the tongue as linguistic evidence: Finnish word initial segments and vowel harmony. *Folia Linguistica* 31, 161–175.
Nimmo, Lisa M. & Steven Roodenrys. 2004. Investigating the phonological similarity effect: Syllable structure and the position of common phonemes. *Journal of Memory and Language* 50, 245–258.
Otake, Takashi, Giyoo Hatano, Anne Cutler & Jacques Mehler. 1993. Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language* 32, 258–278.
Park, Chang-Hai. 1967. The structure of Korean phonemics and phonotactics. Ms.
Perruchet, Pierre & Ronald Peereman. 2004. The exploitation of distributional information in syllable processing. *Journal of Neurolinguistics* 17, 97–119.
Pike, Kenneth L. & Eunice Victoria Pike. 1947. Immediate constituents of Mazatec syllables. *International Journal of American Linguistics* 13, 78–91.
Shen, Jiaxuan. 1993. Slips of the tongue and the syllable structure of Mandarin Chinese. In Shun-chin Yan (ed.), *Essays on the Chinese language by contemporary scholars*, 139–162. Paris: Editions Langages Croisés.
Sohn, Ho-min. 1994. *Korean*. London: Routledge.
Sproat, Richard & Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21, 291–311.
Stemberger, Joseph P. 1994. Rule-less morphology at the phonology–lexicon interface. In Susan D. Lima, Roberta L. Corrigan & Gregory K. Iverson (eds.), *The reality of linguistic rules*, 147–169. Amsterdam: John Benjamins.
Sulkala, Helena & Merja Karjalainen. 2012. *Finnish*. London: Routledge.
Suomi, Kari, Juhani Toivanen & Riikka Ylitalo. 2008. *Finnish sound structure: Phonetics, phonology, phonotactics and prosody*. Oulu: University of Oulu.
Sussman, Harvey H., Nicola Bessell, Eileen Dalston & Tivoli Majors. 1997. An investigation of stop place of articulation as a function of syllable position: A locus equation perspective. *Journal of the Acoustical Society of America* 101, 2826–2838.
Tabain, Marija, Gavan Breen & Andrew Butcher. 2004. VC vs. CV syllables: A comparison of Aboriginal languages with English. *Journal of the International Phonetic Association* 34, 175–200.
Treiman, Rebecca, Brett Kessler, Stephanie Knewasser, Ruth Tinkoff & Margo Bowman. 2000. English speakers' sensitivity to phonological patterns. In Michael B. Broe & Janet

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press

B. Pierrehumbert (eds.), *Papers in Laboratory Phonology V*, 269–282. Cambridge: Cambridge University Press.
Vago, Robert M. 1985. The treatment of long vowels in word games. *Phonology Yearbook* 2, 329–342.
Vago, Robert M. 1988. Vowel harmony in Finnish word games. In Harry van der Hulst & Norval Smith (eds.), *Features, segmental structure and harmony processes*, part II, 185–205. Dordrecht: Foris.
Ventura, Paulo, Régine Kolinsky, Carlos Brito-Mendes & José Morais. 2001. Mental representations of the syllable internal structure are influenced by orthography. *Language and Cognitive Processes* 16, 393–418.
Walker, Jill A. & Gary S. Dell. 2006. Speech errors reflect newly learned phonotactic constraints. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 32, 387–398.
Yi, Kwangoh. 1999. The internal structure of Korean syllables. *2nd International Conference on Cognitive Science and the 16th Annual Meeting of the Japanese Cognitive Science Society Joint Conference*, 978–981. Tokyo: The Japanese Cognitive Science Society.
Yoon, Hye-Kyung, Donald J. Bolger, Oh-Seek Kwon & Charles A. Perfetti. 2002. Sub-syllabic units in reading: A difference between Korean and English. In Ludo Verhoeven, Carsten Elbro & Pieter Reitsma (eds.), *Precursors of functional literacy*, 139–163. Amsterdam: John Benjamins.
Yoon, Yeo Bom & Bruce L. Derwing. 2001. A language without a rhyme: Syllable structure experiments in Korean. *Canadian Journal of Linguistics* 46, 187–237.
Zamuner, Tania S., LouAnn Gerken & Michael Hammond. 2004. Phonotactic probabilities in young children's speech production. *Journal of Child Language* 31, 515–536.

Authors' addresses: *(Berg)*
Department of English, University of Hamburg,
Von-Melle-Park 6, 20146 Hamburg, Germany
thomas_berg@uni–hamburg.de

(Koops)
Department of Linguistics, 1 University of New Mexico,
MSC03 2130, Albuquerque, NM 87131, USA
ckoops@unm.ed

https://doi.org/10.1017/S002222671400022X Published online by Cambridge University Press