2012

Phonotactics as Phonology: Knowledge of a Complex Restriction in Dutch

René Kager, *Utrecht University*
Joe Pater

Slightly revised version to appear in Phonology.

Phonotactics as phonology:
Knowledge of a complex restriction in Dutch

René Kager (Utrecht University) and Joe Pater (University of Massachusetts Amherst)

Abstract: The Dutch lexicon contains very few sequences of a long vowel followed by a consonant cluster, where the second member of the cluster is a non-coronal. We provide experimental evidence that Dutch speakers have implicit knowledge of this gap, which cannot be reduced to the probability of segmental sequences or to word-likeness as measured by neighborhood density. The experiment also suggests that the ill-formedness of this sequence is mediated by syllable structure: it has a weaker effect on judgments when the last consonant begins a new syllable. We provide an account in terms of Hayes and Wilson's Maximum Entropy model of phonotactics, using constraints that go beyond the complexity permitted by their model of constraint induction.

1. Introduction

1.1 General background

Phonological analysis and theorizing typically takes phonotactics as part of the data to be accounted for in terms of a phonological grammar. For example, Chomsky and Halle (1968: 382) aim to provide an account of voice assimilation in Russian as it is instantiated in morpheme alternations and also of the corresponding static phonotactic restriction: the absence of morpheme-internal sequences of obstruents that disagree in voicing (see Pater and Tessier 2006 for experimental support for a unified account of alternations and phonotactics). Ohala (1986) challenges the assumption that phonotactics require a phonological account, raising the possibility that knowledge of the shape of a language's words could be reduced to "knowledge of the lexicon plus the possession of very general cognitive abilities". As an example of the requisite general cognitive abilities, he points to Greenberg and Jenkins' (1964) phoneme substitution metric, which measures the closeness of a nonce word to the existing lexicon as proportional to the number of times that the replacement of each subset of the nonce word's segments yields an existing word.

Like many other models of word-likeness or probability, Greenberg and Jenkins' model of phonotactic knowledge is quite different from those constructed in phonological theory. The first difference is in the nature of the representations employed: Greenberg and Jenkins make use only of segments, while a phonological analysis of a language's phonotactics typically refers to segment-internal features as well as prosodic constituents such as syllables. While Greenberg and Jenkins and Ohala recognize the absence of features from the model as a potential flaw, purely segmental models of phonotactics continue to be common in psycholinguistics and natural language procesing. The second difference is in the nature of the computation used to evaluate a representation. As Ohala (1986) points out, the Greenberg and Jenkins model assumes no "abstracted ('derivative') knowledge about language-specific or language-universal sound patterns", that is, it makes no use of phonological constraints or rules. The third and final difference is in the nature of the classification of a representation that the model returns: the Greenberg and Jenkins model yields a range of scores for representations that would typically be classified as only either "well-formed" or "ill-formed" in a phonological analysis (this two-way classification is commonly assumed in phonological theory – see e.g. Hayes 2004, Prince and Tesar 2004, though cf. Chomsky and Halle 1968:417 and others cited just below). Ohala (1986) points to experimental data gathered by Ohala and Ohala (1986) that shows that nonce word judgments are gradient in the way that the Greenberg and Jenkins model predicts.

Phonologists have only recently responded to Ohala's (1986) challenge. The first type of response has involved the creation of models of phonotactics that provide gradient evaluation of well-formedness, but that exploit the representational vocabulary of phonological theory (e.g. Coleman and Pierrehumbert 1997, Bailey and Hahn 2001, Frisch, Pierrehumbert and Broe 2004, Coetzee and Pater 2006, 2008, Anttila 2008, Coetzee 2008, Hayes and Wilson 2008, Albright 2009). The second type of response has been to provide experimental evidence that phonotactic knowledge is encoded in terms of the constructs of phonological theory, and cannot be attributed solely to measures of word-likeness or probability calculated over segmental strings (e.g. Berent, Everett and Shimron 2001, Frisch and Zawaydeh 2001, Coetzee 2008, Albright 2009).

The present paper contributes a novel type of evidence in support of a phonological account of phonotactics. We show using a nonce word judgement task that Dutch speakers have knowledge of a phonotactic restriction against a sequence of a long vowel followed by a pair of consonants, the second of which is non-coronal (e.g. *[meːlk]). The argument that this knowledge cannot be reduced to segmental sequence probability or to word-likeness is three-fold. First, the experimental items were controlled for measures of both sequence probability and word-likeness. Second, as we discuss in the following section, the complexity of this restriction goes beyond the representational capacity of many models of sequence probability. And finally, our experiments indicate that the strength of this restriction is relative to syllable structure. When the trisegmental sequence at issue is contained within a syllable, as in *[meːlk], its effect on judgments is stronger than when it straddles a syllable boundary, as in *[meːl.kəl]. This asymmetry is not present the Dutch lexicon, in which the sequence is rare in both contexts (as we show in detail in section 2). It seems that in acquiring the phonotactics of Dutch, learners "over-phonologize" them, imposing (or enhancing) a phonological structure.

The Dutch data presented below speak not only to the question of whether phonotactics deserve a phonological treatment, but also to the form of the phonotactic grammar. In the next section, we briefly introduce probabilistic models of phonotactics, focusing on how the Dutch phonotactic restriction investigated here may pose challenges for them. We return to further discussion of Hayes and Wilson's (2008) Maximum Entropy model of phonotactics in section 4, where we present our account of our findings within that framework.

1.2 Probabilistic phonotactics and the Dutch restriction

The complexity of the Dutch phonotactic restriction that we are investigating poses a challenge to both what can be taken as a baseline probabilistic model of phonotactics, as well as to some versions of the more elaborate Maximum Entropy model proposed by Hayes and Wilson (2008).

A probabilistic model of phonotactics assigns to every word some probability over the space of possible words. The simplest probabilistic model of phonotactics is an *n-gram* model. An *n*-gram model of phonotactics is based on the frequency of each segmental string of length *n* in the words of a language. These frequencies yield a probability for each of the strings (the frequency of a string divided by the summed frequencies of all strings of length *n*). The probability of any string longer than *n* can then be calculated as the product of the probabilities of the substrings of length *n*. For example, a unigram model would calculate the probability of [paka] as $p([p]) \times p([a]) \times p([k]) \times p([a])$. A bigram model would calculate the probability of [#paka#] as $p([\#p]) \times p([pa]) \times p([ak]) \times p([ka]) \times p([a\#])$.

While appealingly simple, and broadly used in psycholinguistic research and natural language processing (see Jurafsky and Martin 2000), *n*-gram models are limited in terms of the types of pattern they can capture. One limit derives from the sparseness of attested strings over the space of possible strings when *n* gets larger than 2, making even trigrams typically unuseable as models of phonotactics (see Pierrehumbert 2003: 214 *ff.* for related discussion). A second limit derives from the inability of *n*-gram models to cross-classify strings, to express the non-independent restrictions imposed by constraints on various aspects of phonological representation (Hayes and Wilson 2008, Martin to appear).

The Dutch restriction against long vowels followed by a consonant cluster ending in a non-coronal (e.g. *[meːlk]) cannot be expressed in terms of bigrams, since both the sequence of a long vowel followed by a consonant (e.g. [meːl] 'flour'), as well as the consonant cluster ending in a non-coronal (e.g. [mɛlk] 'milk'), are well-attested. Section 2 documents these claims based on corpus statistics.

Pierrehumbert (2003: 218) raises the issue of whether triphone constraints are internalized by native speakers: "Possibly, people may have some implicit knowledge of these facts, but it would be difficult to demonstrate that any such knowledge goes beyond lexical neighborhood effects". Section 3 presents experimental results that show that Dutch speakers have indeed internalized the $*V{:}CC_{[-cor]}$ restriction. As documented by their comparative well-formedness judgments, subjects show a stronger dispreference for long vowels followed by a cluster than for long vowels followed by just a singleton consonant. The experimental items were controlled for bigram and trigram frequency, as well as for measures of lexical similarity, including the lexical neighborhood density measure mentioned by Pierrehumbert (which somewhat resembles the Greenberg and Jenkins model discussed above). These controls appear to rule out any model of phonotactics based purely on raw segment frequency as an account for these results.

To overcome the limits of simple probabilistic models of phonotactics like *n*-gram models, Hayes and Wilson (2008) propose a Maximum Entropy model, which calculates the probability of a word based on the weights of a set of phonological constraints (see Hayes and Wilson 2008, as well as section 4 below on the details of this calculation). There is no limit to complexity of the constraints that can be employed in a Maximum Entropy model, nor is there a requirement that they be independent. This probabilistic model of phonotactics can in principle accommodate any constraint that has been proposed in phonological theory (that is, so long as the constraint assigns to each representation a numerical score, such as the number of violations).

Hayes and Wilson (2008) also propose a method for learning the constraints, and this method does impose limits on constraint complexity. Hayes and Wilson's constraints are restrictions against sequences of segment types (classified in terms of natural classes), with an upper limit of three elements in the sequence. To express dependencies between non-adjacent segments, as needed to capture vowel harmony and stress placement regularities, Hayes and Wilson make use of projections, similar to those used in autosegmental theory. However, Hayes and Wilson's learner does not acquire constraints that refer to syllable structure or any other prosodic structure.

Violations of $*V{:}CC_{[-cor]}$ can occur both tautosyllabically, and across a syllable boundary (e.g. *[meːlkəl]). At issue is whether violations in these contexts are judged equally serious by human subjects. Our corpus study supports Kager's (1989) observation that this sequence is also highly underrepresented heterosyllabically, and gives little evidence for the constraint's sensitivity to syllable structure. Nevertheless, our experimental results show that syllable structure affects well-formedness judgments; $*V{:}CC_{[-cor]}$ has a stronger effect when the string is contained in a single syllable.

We take these results to support the view that phonotactics do require a phonological account, since no explicit model of "knowledge of the lexicon plus...general cognitive abilities" that we know of can account for them. They also support the view that phonotactic knowledge is not categorical, since the $*V{:}CC_{[-cor]}$ constraint does have exceptions, yet speakers still show evidence of having internalized the pattern, and they also distinguish between grades of ill-formedness. In section 4, we provide an account of our results in terms of Hayes and Wilson's Maximum Entropy model of phonotactic grammar, using a constraint set that goes beyond the one that would be induced by their learner. In particular, the constraint set must distinguish between violations of $*V{:}CC_{[-cor]}$ that occur word-finally and those that occur word-internally. In terms of purely linear constraints, this would exceed the three-element maximum imposed by Hayes and Wilson. In our account, the distinction is achieved by allowing constraints to be indexed to syllabic context; we choose this over including even larger segmental constraints because we suspect that it will be more useful in generalizing to other phenomena. We do not, however, provide an alternative model of constraint induction that would generate these constraints. The case of $*V{:}CC_{[-cor]}$ thus continues to stand as a challenge for models of the learning of phonotactics.

2. Corpus data

Dutch vowels fall into several classes based on distributional criteria (de Groot 1931, Moulton 1962, Cohen, Ebeling, Fokkema and van Holk 1959, Nooteboom 1972, Booij 1995, Gussenhoven 2009). We focus on monophthongs that can occur in stressed syllables (hence, we ignore schwa, diphthongs, and a class of long vowels that occurs in loanwords only). These vowels form two classes that have been traditionally referred to as Class A and Class B (Moulton 1962).

(1)	a.	Class A		b.	Class B		
		ɪ	ʏ		i	y	u
		ɛ	ɔ		eː	øː	oː
			ɑ				aː

These classes are based mainly on two distributional criteria (Moulton 1962, Trommelen 1983, Van der Hulst 1984, Kager 1989, Booij 1995, van Oostendorp 2000). First, Class A vowels never occur in word-final position nor before another vowel (these vowels are 'checked'), whereas Class B vowels freely occur in these positions (these vowels are 'free'). Second, Class A vowels freely occur before consonant clusters whose second member is either labial or velar (i.e., a non-coronal), whereas Class B vowels do not, with few exceptions (e.g. [twaːlf] 'twelve'). This distributional restriction on Class B vowels, which is the focus of our study, is illustrated in (2).

(2)	a.	VːC			b.	$\text{VCC}_{[\text{-cor}]}$			c.	$\text{*VːCC}_{[\text{-cor}]}$	
		paal	paːl	'pole'		*palm*	pɑlm	'palm'		*paalm*	*paːlm
		stoom	stoːm	'steam'		*stomp*	stɔmp	'dull'		*stoomp*	*stoːmp
		haar	haːr	'hair'		*harp*	hɑrp	'harp'		*haarp*	*haːrp
		meel	meːl	'flour'		*melk*	mɛlk	'milk'		*meelk*	*meːlk

We will follow the phonological tradition of referring to Class A vowels as 'short' and Class B vowels as 'long'. Most proposals about Dutch syllable structure have interpreted the phonotactic restriction illustrated in (2c) in terms of an abstract length property (Trommelen 1983, van der

Hulst 1984, Kager and Zonneveld 1986, Kager 1989). In such proposals, the explanation of this phonotactic restriction is that the maximum rime template in word-final position for Dutch has three positions, two of which are occupied by a Class B vowel, leaving only space for a single consonant, plus an extraprosodic word appendix. Alternatively, the rime template can be filled by a Class A vowel (one position) plus two consonants. Since the appendix is restricted to containing coronal obstruents, this explains the relaxation of the pre-cluster restriction for final coronals (e.g. [taːrt] 'cake').

Phonetically it is more accurate to interpret Class A and B vowels as 'lax' and 'tense', respectively (Cohen, Ebeling, Fokkema and van Holk 1959; cf. Rietveld and van Heuven's use of ATR). When unstressed, Class B vowels are phonetically short (with approximately the same duration as Class A vowels; Nooteboom 1972, Rietveld, Kerkhoff and Gussenhoven 2004). Additional phonetic evidence against a length-based analysis is that the high Class B vowels [i, u, y] are phonetically short even when stressed (e.g. *mie* [mi] 'Chinese noodles', *moe* [mu] 'tired', *cru* [kry] 'crude'; *riet* [rit] 'reed', *zoet* [zut] 'sweet', *fuut* [fyt] 'grebe'; Nooteboom 1972, Rietveld, Kerkhoff and Gussenhoven 2004), except being phonetically long before tautomorphemic and tautosyllabic singleton [r] (e.g. [biːr] 'beer', [buːr] 'farmer', [byːr] 'neighbor'; Gussenhoven 1993). Nevertheless, high tense vowels [i, u, y] participate in both of the phonotactic restrictions for Class B vowels, offering only a handful of exceptions to the pre-cluster restriction (e.g. *hielp* [hilp] 'helped', *stierf* [stirf] 'deceased', *wierp* [ʋirp] 'threw'; Moulton 1962, Booij 1995). Distributionally, it is relevant that these exceptions all contain short [i], even those where it occurs before tautomorphemic [r]. Identifying tenseness to be the core contrastive property, van Oostendorp (2000) proposes to specify Class B vowels as underlyingly tense and to derive the surface length of non-high tense vowels by open syllable lengthening.[1] Gussenhoven (2009) argues, on the basis of surface length pairs (e.g. short [i] in *wierpen* [ʋirpən] 'threw+plural' versus long [i] in *Kierkegaard* [kiːrkəgaːrt] idem), that length should nevertheless be represented in the phonology (cannot be due to phonetic implementation), although tenseness may suffice for lexical representations.

[1] Open syllable lengthening targets CV sequences in word-final and pre-vocalic position, as well as crucially word-final CVC (e.g. [paːl] 'pole'), whose final consonant is extrasyllabic. In order to predict lengthening in $CVCC_{[+cor]}$ (e.g. [taːrt] 'cake') but not in $CVCC_{[-cor]}$ (e.g. *[paːlm]), van Oostendorp assumes extraprosodicity of final coronals in addition to extrasyllabicity. Crucially, these devices are additive in $CVCC_{[-cor]}$.

Clearly, the issue of whether length, tenseness, or both, are the phonologically active features is a complex one, and we do not aim to resolve it here. Our own phonological account in section 4 is compatible with either view; all that is crucial is that there be some feature that separates the vowels into two classes. In our corpus study and our experiment, we include the high tense vowels that are phonetically short as instances of Class B/'long' vowels. If Dutch speakers do not in fact classify these vowels as belonging to this class, this would only weaken the difference between the two types of stimuli, rather than introduce a confound.

The following are the statistics (type counts) from a dictionary of 8,305 monomorphemic stems derived from the CELEX Dutch Phonological Lemmas database (Baayen, Piepenbrock and Gulikers 1995). In all counts presented below, consonants in position C_1 in VC_1 and VC_1C_2 were limited to liquids [l, r] and nasals [m, n, ŋ], and consonants in C_2 of VC_1C_2 to obstruents [p, b, f, v, k, g, x, ɣ, t, d, s, z] and nasals. This is a subset of the full set of clusters, with limits that are the same as we imposed on our experimental stimuli; this reflects an assumption that the subjects in our experiment may calculate the relevant probabilities over that part of the lexicon that has phonological properties similar to those of the test items (see further footnote 9). This choice does not affect our conclusions.[2] Long/tense vowels included [a, e, o, ø, i, u, y] and short/lax vowels [ɑ, ɛ, ɪ, ɔ, ʏ]. The CELEX database was searched for word-final strings without limitation to syllable number. For example, the search for the word-final sequence $VCC_{[-cor]}$ returned monosyllables such as *bonk* as well as disyllables such as *spelonk*.

We start with word-final position, where consonant clusters are tautosyllabic, as in the above examples. The expected values (E) are calculated from the joint probability of the two levels of a factor in this set of strings. For example, E=99 in the Vː CC cell comes from the overall probability of long vowels (0.39) times the overall probability of clusters (0.19) times the total number of strings (1340). O/E is the observed number (O) divided by E. When O/E is lower than 1, the observed value is lower than expected. Since there are only two long vowels in the context of word-final clusters that end in non-coronals, O/E is quite low (O/E = 0.02). A chi-square test measures the likelihood that the overall distribution arose from chance. The distribution in Table 1 is highly unlikely to have arisen from chance (Chi-square = 191.87, df = 1, $p < .001$).

[2] For example, if we include all CC clusters in a contingency table like that in Table 1, we still only have 3 examples of $Vː CC_{[-cor]}$ clusters, which leads to an O/E value of 0.03. The other O/E values also remain essentially the same, as do those in other tables.

	_C	_CC
V	563 (O/E = 0.85)	252 (O/E = 1.63)
V:	523 (O/E = 1.23)	2 (O/E = 0.02)

Table 1. Word-final position in monomorphemic Dutch stems, where C_2 is [-cor].

The underrepresentation of V:$CC_{[-cor]}$ sequences is thus not merely the product of the probability of long vowels in the _$CC_{[-cor]}$ context and the probability of postvocalic $CC_{[-cor]}$ sequences as compared with C. We can further note that while long vowels are somewhat underrepresented in all _CC contexts, the effect is far stronger with clusters that end in a non-coronal. That is, the *V:$CC_{[-cor]}$ restriction is not just a general restriction against long vowels before clusters. This can be seen in Table 2, where the coronal-final clusters receive an O/E score of 0.59.[3] The overall distribution is again highly unlikely to have arisen by chance (Chi-square = 240.38, df = 2, $p < .001$).

	_C	_$CC_{[+cor]}$	_$CC_{[-cor]}$
V	563 (O/E = 0.80)	264 (O/E = 1.22)	252 (O/E = 1.53)
V:	523 (O/E = 1.36)	70 (O/E = 0.59)	2 (O/E = 0.02)

Table 2. Word-final position in monomorphemic Dutch stems

We now turn to the distribution of clusters which occur in prevocalic position, and hence, are heterosyllabic. This investigation is motivated by claims in the literature (Moulton 1962, Kager 1989) that Class B vowels are restricted by a $CC_{[-cor]}$ cluster not only in word-final position but also before a vowel, where clusters are heterosyllabic (e.g. [aːrdə] 'earth' versus

[3] It remains possible that the *V:$CC_{[-cor]}$ restriction is due to the joint effects of a constraint against long vowels before clusters and a constraint against clusters that end in non-coronals. This seems unlikely, given the large number of $CC_{[-cor]}$ clusters following short vowels (see Wilson and Obdeyn 2009 for a Maximum Entropy method for addressing this and similar questions).

*[aːrbə]). Again, long vowels are highly underrepresented before non-coronal-final clusters. Compared with word-final position, the coronal/non-coronal difference is slightly smaller (that is, the degree of underrepresentation of long vowels before coronal-final clusters is higher in Table 3 than Table 2). The overall distribution is again highly unlikely to have arisen by chance (Chi-square = 1190, df = 2, $p < .001$).

	_C	$\_CC_{[+cor]}$	$\_CC_{[-cor]}$
V	408 (O/E = 0.49)	485 (O/E = 1.73)	492 (O/E = 1.80)
Vː	1124 (O/E = 1.61)	30 (O/E = 0.13)	10 (O/E = 0.04)

Table 3. Word-internal position in monomorphemic Dutch bisyllabic stems

The bisyllabic words in our experiment are limited to monomorphemic stems ending in a schwa-liquid cluster. Table 4 presents the figures for monomorphemic stems ending in a schwa-liquid cluster in the Dutch lexicon, which form a subset of the cases in Table 3. The pattern is largely the same, though long vowels are preferred in the _C context in table 3, while short vowels are more common in this same context in table 4. The overall distribution is once more highly unlikely to have arisen by chance (Chi-square = 23.48, df = 2, $p < .001$).

	_C	$\_CC_{[+cor]}$	$\_CC_{[-cor]}$
V	54 (O/E = 0.86)	66 (O/E = 1.07)	71 (O/E = 1.07)
Vː	14 (O/E = 2.66)	1 (O/E = 0.19)	1 (O/E = 0.18)

Table 4. Word-internal position in Dutch stems ending in VC(C)əL

In sum, long vowels are strongly underrepresented before $CC_{[-cor]}$ clusters that are both tautosyllabic with respect to the vowel (in word-final position) and that are heterosyllabic (in prevocalic position). To test the statistical reliability of the $*V{:}CC_{[-cor]}$ restriction, and to investigate whether syllable structure affects its strength, we performed a logistic regression on the lexical data presented above. We included only the cases with singleton consonants and $CC_{[-cor]}$ clusters, since these correspond to our experimental stimuli. Words were coded for the binary dependent variable of vowel length (long = 1, short = 0), and the binary explanatory variables of following consonantal context ($CC_{[-cor]}$ cluster = 1, singleton = 0) and syllabification (tautosyllabic = 1, heterosyllabic = 0). We expect an effect of context based on the relatively low frequency of long vowels in the $CC_{[-cor]}$ cluster context shown in the tables above. If this effect is mediated by syllable structure, there should be an interaction between context and syllabification.

The first analysis includes the tautosyllabic data from table 2 and the bisyllabic data from table 3 (that is, the broader set of bisyllables). The model was fitted using the GLM function in R (R Development Core Team 2010), and the significance values of the factors and their interaction shown in (3) are those that it returns.

(3) Result of logistic regression with full set of bisyllables

	Coefficient Estimate	*Standard error*	*P (>|z|)*
Intercept	1.01	0.06	< 0.001
Syllabification	–1.09	0.08	< 0.001
C-Context	–4.91	0.32	< 0.001
C-Context * Syllabification	0.15	0.78	0.851

The effects of both consonantal context (C-Context) and syllabification are statistically significant, while the relatively small effect of the interaction is not.

From the coefficient of the intercept, we can calculate the probability that the model gives a long vowel when the explanatory variables have a value of zero, that is, when a singleton consonant follows in a bisyllable ($1 / (1 + \exp–(1.01)) = 0.73$). This expected probability matches the frequency in the observed data (to several decimal places). The coefficients for each of the main effects allows us to calculate the probability given to a long vowel when a $CC_{[-cor]}$ cluster follows in a bisyllable ($1 / (1 + \exp–(1.01–4.91)) = 0.02$, observed frequency = 0.06), and when a singleton follows in a tautosyllabic context ($1 / (1 + \exp–(1.01–1.09)) = 0.48$, observed frequency = 0.48). The large main effect for context is what we expected based on the rarity of long vowels in the $CC_{[-cor]}$ context. The effect for syllabification arises from the general overrepresentation of long vowels in bisyllables relative to monosyllables.

With the coefficients from both of the main effects and the interaction term we get the predicted probability of a long vowel before a $CC_{[-cor]}$ cluster in a tautosyllabic context ($1 / (1 + \exp–(1.01–1.09–4.91+0.15)) = 0.01$, observed frequency = 0.01). The coefficient for the interaction term indicates a small effect that goes in the opposite direction from what would be predicted if the presence of $CC_{[-cor]}$ cluster had a greater influence in a tautosyllabic context than in a heterosyllabic context on the probability of a long vowel. Since the effect is so small and so far from statistically significant at the 0.05 level, this almost certainly indicates the absence of an influence of syllable structure on the strength of the restriction, rather than an influence in the unexpected direction.

In sum, the model gives long vowels significantly lower probability before *V:$CC_{[-cor]}$ clusters, and in tautosyllabic contexts than in heterosyllabic contexts, but these effects are independent: the strength of the *V:$CC_{[-cor]}$ restriction is not significantly greater in tautosyllabic contexts. We further confirmed this finding by fitting a model that does not include the interaction term: there was no significant difference in predictive power between the models as measured by a chi-square test ($p = 0.854$). With the interaction, the AIC score was 3409.3, and without it was 3407.3. On this measure, which includes a penalty for model complexity, lower is better, but a difference of 2 is not usually taken as choosing between models (Burnham and Anderson 2002: 446).

The second analysis includes the same tautosyllabic data, but the heterosyllabic data from Table 4, in which the rime shapes of the second syllables more closely match those of our experimental data. The analysis was otherwise conducted in an identical fashion.

(4) Result of logistic regression with subset of bisyllables

	Coefficient Estimate	*Standard error*	*P (>|z|)*
Intercept	–1.35	0.30	<0.001
Syllabification	1.28	0.31	<0.001
C-Context	–2.91	1.05	0.006
C-Context * Syllabification	–1.85	1.27	0.145

The picture is largely the same: the main effects are both statistically significant at the 0.01 level, while the interaction is not. The baseline expected probability for long vowels in heterosyllabic contexts with a following singleton is 0.21 (observed 0.21). With a following $CC_{[-cor]}$ cluster, the expected probability drops to 0.01 (observed 0.01). The expected probability of a long vowel with a following singleton in a tautosyllabic context is in this case higher than the baseline heterosyllabic case: 0.48 (observed 0.48). Finally, the expected probability of a long vowel with a following $CC_{[-cor]}$ cluster in a tautosyllabic context is 0.001. In this analysis, the coefficient for the interaction does have the expected sign: a $CC_{[-cor]}$ cluster does exert a greater negative influence on the probability of a long vowel in a tautosyllabic context. However, the lack of statistical significance of the interaction is again confirmed in model comparison: the chi-square test again fails to reach significance ($p = 0.198$). Comparison of the AIC scores again does not rule out the model with the interaction, which has only a slightly worse score (1615.1) than the model without it (1614.7).

In sum, in the larger dataset, there is no evidence at all that syllable structure affects the strength of the *V:$CC_{[-cor]}$ restriction. In the smaller dataset, there is only very weak evidence of such an effect. It is difficult to know which of these datasets more closely matches the data that underly the system that the subjects used to perform our experiment (though see footnote 9). We thus conclude that there is little or no evidence from the lexicon for the role of syllable structure.

3. Experimental data

To investigate (1) whether the *V:$CC_{[-cor]}$ restriction is internalized by native speakers, and (2) whether the internalized constraint is syllable-sensitive, we conducted a non-word judgment experiment. We addressed the first question in terms of whether our Dutch participants show a greater dispreference for words containing long vowels when a $CC_{[-cor]}$ cluster follows than when a singleton does. We might have also tested whether the long vowel dispreference is greater when a $CC_{[-cor]}$ cluster follows than when a $CC_{[+cor]}$ one does, but there is a confound introduced by morphology. The Dutch suffixes [-s] (plural) and [-t] (second and third person plural) are both single coronals that attach to consonant-final stems, and so any nonce word containing final [s] or [t] might be interpreted as bi-morphemic, which would exempt them from the restriction for a non-phonological reason. Omitting these consonants would overly limit the set of possible stimuli. We asked the second question in terms of whether this dispreference was affected by syllabification: whether the word being judged was a monosyllable, in which case the cluster is tautosyllabic and contained in a single syllable, or a disyllable, in which case the cluster is heterosyllabic and spans a syllable boundary. In statistical terms, we are asking whether there is a main effect of cluster presence on long vowel dispreference, and whether there is an interaction between cluster presence and syllabification.

The experiment involved a comparative well-formedness judgment task, in which participants were presented with pairs of nonce words and indicated which item in a pair sounded more like it could be a real Dutch word. Comparative well-formedness has been found to bring out finer-grained differences than judgments of single items on a scale (Ohala and Ohala 1986, Berent and Shimron 1997, Coetzee 2009, Daland *et al.* 2011).[4]

3.1 Design

Participants. Participants were 34 native speakers Dutch, all students of Utrecht University. None reported any hearing difficulties. They were paid a small amount for participation.

Materials. A total of 240 stimuli was included, monosyllabic and disyllabic. Monosyllabic stimuli consisted of four sets of 30 nonce words of the structural types CVC, CVCC, CVːC, and CVːCC, varying in vowel length (short versus long) and the presence of a cluster (cluster versus singleton consonant). The clusters consisted of a sonorant [l, r, m] in first position, and a voiceless obstruent [p, f, k, x] or a nasal [m] in second position. All clusters used [lp, lk, lm, rp, rf, rm, mp, mk] are attested in monomorphemic words; one cluster was attested only intervocalically ([mk]). Singleton consonants ending CV(ː)C nonce words were [l, r, m], matching postvocalic consonants of the CV(ː)CC items. Disyllabic stimuli consisted of 4 sets of 30 nonce words of the structural types CVCəC, CVCCəC, CVːCəC, and CVːCCəC, all ending in schwa plus liquid. These again had [l, r, m] in C_1, but a slightly wider range of choice for C_2, which also included the voiced obstruents [b, v, x]. Nonce words of the types CVːl(C)əl and CVːr(C)ər were avoided in order to rule out any possible influence from constraints disfavoring identical liquids. Stimuli came in short-long pairs: for each nonword containing a short vowel, there was another that was identical except having a long vowel. Moreover, stimuli came in

[4] We also ran a scalar judgment study, with a 1-7 word-likeness scale, using the same stimuli, and 20 different Dutch participants. We found a highly significant main effect for cluster presence, and an effect in the predicted direction for the interaction, which did not quite reach significance at the 0.05 level.

cluster-singleton pairs: for each nonword containing a consonant cluster, there was another that was identical except having a singleton, omitting the cluster's final consonant. Examples of nonword stimuli are provided in Table 5.

CVC	CV:C	CVCC	CV:CC
bɑm	ba:m	bɑmk	ba:mk
xɔl	xo:l	xɔlm	xo:lm

CVCəC	CV:CəC	CVCCəC	CV:CCəC
dɛmər	de:mər	dɛmxər	de:mxər
jɔlər	jo:lər	jɔlbər	jo:lbər

Table 5. Examples of monosyllabic and disyllabic nonce words.

Nonce words ending in clusters contained zero or low-frequency V(:)$CC_{[-cor]}$ portions. The rationale was that if speakers have internalized the restriction, then this should generalize to all V:$CC_{[-cor]}$ sequences, including ones unattested in Dutch (such as [e:mk]). See Appendix A for a complete list of stimuli.

Throughout the stimulus set, we controlled for lexical factors and phonotactic probability in the sets of long-short pairs, e.g. CVC versus CV:C, CVCC versus CV:CC, *etc*. Two lexical factors were controlled for between the conditions. The first was Lexical Neighborhood Density (LND), the sum of the logged token frequencies of a nonce word's neighbors, based on the CELEX Dutch Phonological Lemmas database. A neighbor was defined as any word that results from changing, inserting, or deleting a single segment. The second was Cohort Density (CD), the sum of logged token frequencies of a nonce word's cohort members, where a nonce word's cohort was defined as all words sharing its first three segments.

In addition to these lexical factors, we controlled for transitional probabilities of biphones (TPs). TP values were calculated from word types in the monomorphemic lexicon. The TP for a biphone xy is defined as p (y|x) = freq(xy)/freq(x), where freq(xy) is the frequency of a biphone xy, and freq(x) that of the phoneme x. Biphones included word-initial and word-final phonemes (e.g. [#b] for word-initial [b], or [p#] for word-final [p]). The transitional probabilities of stimuli as reported in the Appendix were defined as the averaged logged TP value of the biphones in each stimulus. For example, the transitional probability for [#baːm#] (−1.030) was calculated as follows. The TP values of its four constituent biphones [#b], [baː], [aːm], [m#] (0.081, 0.089, 0.057, 0.187, respectively) were logged (−1.092, −1.049, −1.248, −0.729), then summed (−4.118), and this sum was divided by four, the total number of biphones. In case of biphones with zero frequencies, smoothing was applied, resetting their frequency to 1. To the 240 stimuli, 124 fillers were added, monosyllabic and disyllabic nonce words, none of which contained clusters used in the test words.

The materials were spoken by a female native speaker who was naive to the purposes of the experiment, and were recorded digitally. Two versions of each stimulus were recorded. After listening to the recorded stimuli, we eliminated any that contained hesitation or noise, and generally selected the ones that sounded most natural. To form the test items, the stimuli were paired into 120 long-short pairs, such that each stimulus pair only minimally differed in vowel length, but was otherwise identical. Examples of test pairs are provided in Table 6.

	monosyllabic	**disyllabic**
CVCC - CVːCC	bɑmk - baːmk	bɑmvər - baːmvər
CVC - CVːC	bɑm - baːm	bɑmər - baːmər

Table 6. Examples of monosyllabic and disyllabic nonword test pairs used in the experiment.

The 120 test pairs were mixed with 120 filler pairs, which all consisted of one test item plus a filler item from the first experiment (e.g. [bɑmvər] - [xoːlər]). Each test pair and filler pair was presented in both orders, so that the total number of trials was 480. Stimulus order was randomized between participants.

Procedure. Participants were instructed to select the most word-like item for every pair of nonce words, using the following (translated) text: “You are going to listen to pairs of nonce words. Your task will be to determine how much these nonce words sound like real Dutch words. For each pair, you should select the one that sounds most Dutch-like word: if it is the first one, press the lefthand button; if it is the second one, press the righthand button”. Responses were recorded using a button box. Participants listened to stimuli through headphones at a comfortable level of loudness. The interstimulus interval (time elapsed between the members of the pairs) was 200 ms. The intertrial interval (time elapsed between the response and the next stimulus pair) was 500 ms. The response duration was 2500 ms. If no response was made at that point, the next pair was presented.

3.2 Results

In table 7, we present the proportion of long vowel responses for each of the conditions. Standard deviations, calculated over subjects, are given in parentheses. Examples of comparison pairs can be found in the corresponding cells of Table 6.

	monosyllabic	**disyllabic**
singleton	0.41 (0.17)	0.38 (0.20)
cluster	0.27 (0.17)	0.32 (0.18)

Table 7. Proportion of long vowel choices in short-long pairs, with standard deviations.

The frequency of long vowel responses was lower than chance in the singleton conditions. One possibility is that our distractors were not successful in stopping our subjects from developing a response bias, perhaps generalizing the ill-formedness of long vowels in the cluster conditions to others. Nonetheless, even with this possible uniform response bias, differences between conditions remain. The pattern of responses suggests positive answers to both of our research questions. Words with long vowels were chosen less frequently when a cluster followed, across monosyllables and disyllables. In addition, the difference between singleton and cluster was greater in the monosyllables (0.14) than the disyllables (0.06), suggesting an interaction between consonantal context and syllabification. This is shown graphically in the steeper slope for monosyllables in Figure 1.

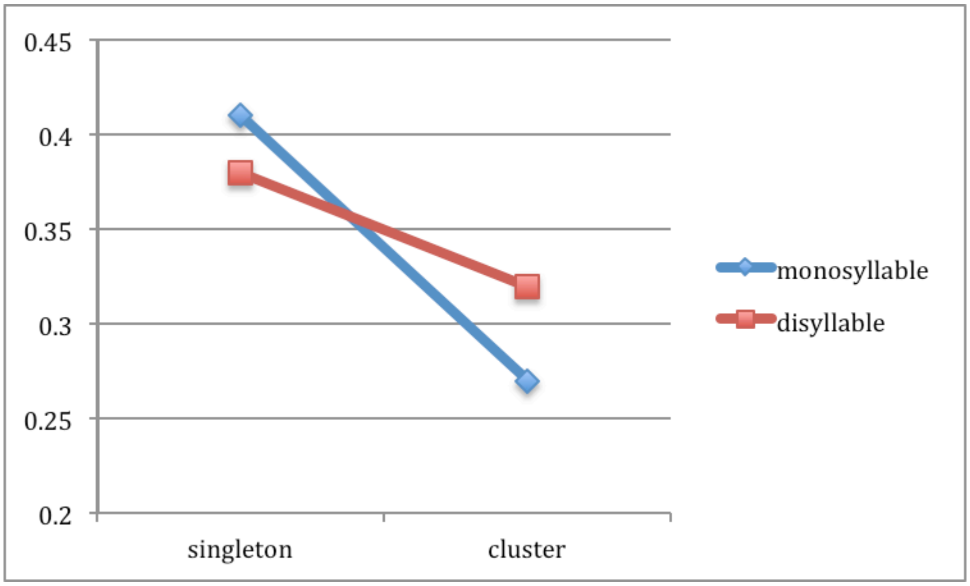

Figure 1. Proportion long vowel choices across consonantal contexts within syllable types.

To test if the observed main effects for cluster and interaction between syllabification and cluster are statistically reliable, we used a mixed effects logistic regression model (see Baayen 2008 and Jaeger 2008 on the advantages over ANOVAs for the type of data analysed here). As in the logistic regression analysis of the corpus data presented in section 2, the dependent variable is vowel length (long = 1, short = 0), though here it corresponds to subjects' choice on a test pair, rather than the length of a vowel in a word in the corpus. The binary explanatory variables are again consonantal context ($CC_{[-cor]}$ cluster = 1, singleton = 0) and syllabification (tautosyllabic = 1, heterosyllabic = 0). The predictor values were centered before the analysis was run. Experimental items were given random intercepts, and subjects were given random slopes for consonantal context and syllabification (these random effects were selected through model comparison). The model was fitted in R using the lmer function of the lme4 package (Bates and Maechler 2010). We report here only on the fixed effects, whose coefficient estimates, with standard errors and *p* values, are shown in (5).

(5) Result of mixed effects logistic regression on experimental data

	Coefficient Estimate	*Standard error*	*p (>|z|)*
Intercept	–0.81	0.16	< 0.001
Syllabification	–0.02	0.06	0.691
C-Context	–0.30	0.07	< 0.001
C-Context * Syllabification	–0.12	0.06	0.046

As expected from our descriptive statistics, C-Context has a negative effect on the probability of selecting a long vowel. Without the random effects, the baseline probability of selecting a long vowel (in a bisyllable with a following singleton) is 1 / (1 + exp–(–0.81)) = 0.31. This drops to 1 / (1 + exp–(–0.81–0.30)) = 0.25 when the vowel is followed by a consonant cluster. This effect is statistically reliable, as indicated by the very low *p* value for C-Context. Syllabification on its own has only a very small negative effect, which is not statistically reliable, but the interaction of C-Context and Syllabification is larger: the model gives a probability of 0.30 to a long vowel choice in the singleton context in a tautosyllabic case, and 0.22 in the cluster context in a heterosyllabic case. As shown in (5), the interaction reaches significance at the 0.05 level. We further confirmed the reliability of this effect by comparing this model with one that has the interaction removed (from the fixed and random effects). The model with the interaction has an AIC score of 8954.2, which is better than the score of 8957.1 for the model without it, and the chi-square test returns a *p* value of 0.025.The finding of a statistically significant effect for the interaction supports the hypothesis that syllable structure affects the strength of the *Vː$CC_{[-cor]}$ restriction.

An alternative account of the effect of Syllabification might be that the long vowels are phonetically longer in monosyllables than disyllables, and that this increased phonetic duration is responsible for the greater dispreference for long vowels with a following $CC_{[-cor]}$ cluster.[5] To test this alternative account, we first took measurements of the duration of all of the vowels in our stimuli. These measurements are summarized in table 8. Each cell contains the average difference between the long and short members of our test pairs, with standard deviations. These means indicate that the duration differences between the long and short vowels were in fact

[5] We thank an anonymous reviewer for bringing this possibility to our attention.

greater in the disyllables than in the monosyllables, counter to the premise of the alternative account.

	monosyllabic	**disyllabic**
singleton	0.07 (0.11)	0.08 (0.04)
cluster	0.05 (0.03)	0.06 (0.04)

Table 8. Mean vowel length differences in short-long pairs, with standard deviations in parentheses

To further explore the role of duration in explaining our experimental results, we performed mixed effects logistic regressions using the duration differences in the test pairs as an explanatory variable. In the first analysis we included only duration difference as a fixed effect, and found that the best model included a random slope for subjects, and a random intercept for items. Again we discuss only the fixed effect.

(6) Mixed effects logistic regression with phonetic duration

	Coefficient Estimate	*Standard error*	*P (>|z|)*
Intercept	–0.79	0.16	< 0.001
Duration difference	0.23	0.06	< 0.001

There is a significant effect for duration, but it goes in the opposite direction from that predicted by the alternative account of our results. Greater difference in duration between the long and short members of a pair leads to an increase in the probability of a long vowel choice. Even if one had an *a priori* reason to expect that greater duration differences should lead to more long vowel choices, this model does a poorer job of explaining the data than the one in (5), getting an AIC score of 9005. We also explored models that included both duration difference and consonantal context as fixed effects, and found that the best one (with duration difference as a random slope for subjects and a random intercept for items) got an AIC score of 8958, again worse than the 8954.2 for our preferred model. Because the effect goes in the opposite direction from predicted, and because models using it do not perform as well as the one incorporating Syllabification, we reject phonetic duration differences as an alternative account of those aspects of the data that we attribute to syllable structure. However, there is reason to seek an explanation for why duration difference has the effect it does. When duration difference is added to our preferred model, it continues to have a positive effect on the probability of a long vowel choice, with a p value < 0.001. This shows that its effect is not merely due to a correlation with the other explanatory variables (they continue to have the same direction of effect, with about same size and p value).[6]

Finally, we return to the question of whether the subject's judgments can be explained by segmental sequence probability and word-likeness. Our experimental items were controlled for measures of these constructs, but we here examine whether residual differences might play a role in explaining the data. We coded our experimental items for differences in the stimulus pairs for lexical neighbourhood density (LND), cohort density (CD), and transitional probability (TP) (see section 2 for definitions). We first performed a logistic mixed effects regression with vowel length choice as the dependent variable, and centered versions of these three measures and their interactions as fixed effects. Subject and items had random intercepts. In this analysis, we found that only the main effect of TP reached statistical significance at the 0.05 level ($p < 0.001$), as did the interaction between TP and CD ($p = 0.036$). The best model that we could find incorporating these measures (and none others) included only TP as a fixed effect, with TP as a random slope for subjects and a random intercept for items. This received an AIC score of 8989, much worse than that of our preferred model.

When we add TP to the model presented in (5), the result is as shown in (6). The AIC score is 8933, the best that we found in our model exploration. Along with TP, C-Context continues to have a highly reliable effect, and the effect of the interaction between C-Context and Syllabification now approaches statistical significance at the 0.01 level.

[6] One possible explanation for the pattern of duration differences is that our speaker slowed her speech rate in more marked contexts. It is also possible that subjects' responses reflect this slowing. We leave the investigation of the possible role of gradient ill-formedness on speech rate, and the possibility that this affects judgments, as a topic for further research.

(6) Mixed effects logistic regression including transitional probability (TP)

	Coefficient Estimate	*Standard error*	*P (>|z|)*
Intercept	–0.81	0.16	< 0.001
Syllabification	0.02	0.06	0.742
C-Context	–0.32	0.07	< 0.001
TP	–0.28	0.05	< 0.001
C-Context * Syllabification	–0.12	0.06	0.016

All of this indicates that TP did affect our subject's choices, but that this effect does not explain the same aspects of the data that consonantal context and its interaction with syllabification do.

4. A Maximum Entropy phonotactics account

Our experiment shows that Dutch speakers have knowledge of the $*V{:}CC_{[\text{-cor}]}$ restriction: their dispreference for long vowels is stronger before a cluster ending in a non-coronal than before a singleton consonant. The results also show that the force of $*V{:}CC_{[\text{-cor}]}$ is mediated by syllable structure: the dispreference for long vowels is stronger when the cluster is fully contained in the same syllable as the vowel. In this section, we provide an account of this aspect of Dutch phonology in terms of a Maximum Entropy model of phonotactics (Hayes and Wilson 2008) that incorporates constraints referring to prosodic structure.

Maximum Entropy phonotactics uses weighted constraints to define a probability distribution over the space of possible words. As an illustration, we can consider the probability distribution defined over a small space of monosyllabic words by an arbitrary weighting of two standard phonological constraints. The constraints penalize long vowels (*V:), and consonants in coda position (*CODA), assigning a violation score of –1 for each offending structure. The scores assigned by the constraints to the types of word under consideration are shown in the rows of the tableau in (7). The first word type, V:CC contains a long vowel, and thus scores –1 on *V:, as well as two coda consonants, which result in a –2 score on *CODA. The weights of the constraints are shown beneath the constraint names in the first row: *V: has a weight of 2, and *CODA has a weight of 1. The column labeled *H* shows the Harmony of each word type: the sum of the violation scores, each multiplied by the constraint's weight (see Smolensky and Legendre 2006 on the history of Harmony in linguistics). Words of the V:CC type have Harmony = (2 × –

1) + (1 × –2) = –4. The next column exp(*H*) shows the result of raising *e* (2.72) to the power of *H*. A word's probability is proportional to exp(*H*); this is shown in the final column labeled *p*.

(7) *Illustration of Maximum Entropy phonotactics*

	*V: 2	*CODA 1	*H*	exp(*H*)	*p*
V:CC	–1	–2	–4	0.02	0.01
VCC		–2	–2	0.14	0.08
V:C	–1	–1	–3	0.05	0.03
VC		–1	–1	0.37	0.22
V:	–1		–2	0.14	0.08
V			0	1.00	0.59

This example illustrates several properties of the model. First, well-formedness, defined in terms of probability, is gradient, ranging from 0.59 for a word with neither a long vowel nor coda, to 0.01 for the V:CC word type. Second, we can see how the constraint weights affect probability: because *V: has a higher weight than *CODA, words with long vowels (V:) have a lower probability (0.14) than words with codas (VC, 0.37). And finally, we can see the cumulative effect of constraint violation: V:CC words have the lowest probability not because they violate any one constraint that the others satisfy, but because they have a greater number of constraint violations than the others.

Our aim is to provide a Maximum Entropy phonotactic grammar that is compatible with our experimental findings that Dutch speakers have knowledge of the *V:$CC_{[\text{-cor}]}$ restriction, and that they are particularly sensitive to the restriction when the sequence is contained within a single syllable. We also aim to find the weights for the grammar by training a learner with the lexical data presented in section 2. This second goal might at first seem to present an insurmountable challenge: if as our regression analyses suggest, there is little or no evidence for syllable-sensitivity in the lexicon, how could a learner acquire a syllable-sensitive restriction on the basis of those data? The answer involves a particular kind of inductive bias: the learner posits a *V:$CC_{[\text{-cor}]}$ constraint that is relativized to prosodic context, and favors a grammar that gives a significant amount of weight to this constraint.

As the learning data, we used the lexical counts that were used in the second regression analysis in section 2. These are the data that most closely match the experimental data: they exclude coronal-final clusters, and the bisyllables are limited to ones ending in schwa-liquid clusters. The lexical data are shown in terms of counts and proportions/probabilities in Table 10.

Word-type	*Count*	*p*	*Word-type*	*Count*	*p*
V:CC$_{[-cor]}$#	2	0.001	V:CC$_{[-cor]}$V	1	0.001
VCC$_{[-cor]}$#	252	0.075	VCC$_{[-cor]}$V	71	0.048
V:C#	523	0.155	V:CV	14	0.009
VC#	563	0.167	VCV	54	0.036

Table 10. Learning data

The constraints penalize types of segmental sequence, as in Hayes and Wilson (2008), and much other research in phonology. We do not employ Hayes and Wilson's constraint induction algorithm, since it does not yield prosodically conditioned constraints. Instead, we employ the full set of constraints of a particular type that apply to our case, with the expectation that these may well be learned both by humans and by a learning algorithm on the basis of the observed forms of the language. Like Hayes and Wilson (2008), we have not tailored our constraint set to yield typological predictions: with different weights, they could yield implausible outcomes, such as a language that has lower probability of short than long vowels before tautosyllabic consonants. Our results could equally well be analyzed using a typologically tailored constraint set, though see Hayes and Wilson (2008), as well as Daland *et al.* (2011) and Pater (to appear), for discussion of how phonological typology might be accounted for if constraints are learned.

As a first pass at a set of constraints, we can consider the following "biphone"-type constraints, which penalize each of sequence types under consideration:

(8) *V:C — Assign –1 to a long vowel followed by a consonant
*VC — Assign –1 to a short vowel followed by a consonant
*CC$_{[-cor]}$ — Assign –1 to a consonant followed by a non-coronal

Because constraint interaction is cumulative in the Maximum Entropy model, one might think that this constraint set could account for the *V:CC$_{[-cor]}$ restriction as the cumulative effect of *V:C and CC$_{[-cor]}$. However, the biphone constraints in a Maximum Entropy model will not suffice to account for the Dutch data, for a reason parallel to why the bigram model fails to account for them (see the discussion in section 2). The problem is illustrated in (9).

(9) *The failure of biphone constraints*

	*V:C 2	*CC$_{[-cor]}$ 2	*VC 1	*H*	exp(*H*)	*p*
V:CC$_{[-cor]}$	–1	–1		–4	0.02	0.03
VCC$_{[-cor]}$		–1	–1	–3	0.05	0.09
V:C	–1			–2	0.14	0.24
VC			–1	–1	0.37	0.64

By granting *V:C a greater weight than *VC, and assigning a positive weight to *CC$_{[-cor]}$, we succeed in making the Harmony of V:CC$_{[-cor]}$ lower than that of the other word types, and thus making its probability the lowest. However, this also leads to V:C being at least as ill-formed relative to VC as V:CC$_{[-cor]}$ is to VCC$_{[-cor]}$. If we take well-formedness to be a function of Harmony (see Coetzee and Pater 2008 on phonotactics), then the difference between these two pairs of word types will be equivalent (equal to difference between the weights of *V:C and *VC). If we take well-formedness to be a function of exp(*H*) (Hayes and Wilson 2008), or of *p*, then the difference between V:C and VC will in fact be greater, since exponentiation makes the contribution of the V:C *vs.* VC difference lower when the words also violate *CC$_{[-cor]}$. Either way, the outcome is not the one we want.

We thus require a constraint that will pick out V:CC$_{[-cor]}$ as especially ill-formed. To produce this constraint as one instantiation of a constraint type, we expand our constraint set to include the triphone constraints in (10):

(10) *V:CC$_{[-cor]}$ Assign –1 to a long vowel followed by a consonant and a non-coronal
*VCC$_{[-cor]}$ Assign –1 to a short vowel followed by a consonant and a non-coronal

We also require a constraint to pick out V:CC$_{[-cor]}$# as opposed to V:CC$_{[-cor]}$V, so as to make the short vowel preference stronger when the *V:CC$_{[-cor]}$ violation is tautosyllabic. To produce the constraint type of which this is an instance, we include domain-specific versions of all of the constraints, which apply only when the sequence is contained in a single syllable. The domain-specific version of *V:$CC_{[-cor]}$ is shown in (11).

(11) $*_{\sigma}$V:$CC_{[-cor]}$ Assign –1 to a long vowel followed by a consonant and a non-coronal contained within a syllable

As we mentioned in the introduction, an alternative would have been to expand the segmental size of the constraints, relativizing V:$CC_{[-cor]}$ to the word-final and prevocalic environments. We chose to relativize it to prosodic context because we expect that prosodically conditioned constraints will in general prove more useful to a theory of phonotactics than would extremely long segmental constraints.

To find constraint weights, we used the L-BFGS-B method (Byrd *et al.* 1995) as implemented in R (R Development Core Team 2010).[7] The objective is to find a set of weights that minimizes the difference between two probability distributions over the space of word-types: the one supplied in the learning data, and the one defined by the Maximum Entropy grammar. We used Kullback-Leibler divergence as our measure of difference (Kullback and Leibler 1951).[8] The objective function also includes a regularization term or prior that penalizes weights as they depart from zero. We set a zero minimum on constraint weights and used an L2 or Gaussian prior, which with a zero minimum is equivalent to penalizing the sum of squared weights. The strength of the regularization term with respect to error minimization is expressed in terms of variance in the distribution of weights, so its strength is inversely correlated with its magnitude. We used two values: 10, which imposes a strong penalty on large weights, and 1,000,000, which imposes a very weak penalty.

The constraint weights for the two grammars are shown in (12), with 'Weak' and 'Strong' indicating which regularization term was used in learning (multiple runs get similar

[7] We thank Robert Staubs with providing us with the R script that we used for this purpose. Readers interested in replicating our analysis can find the necessary materials at http://blogs.umass.edu/pater/papers/.

[8] Hayes and Wilson (2008) state their objective function in terms of maximizing the likelihood of the observed forms, which is equivalent to our reformulation in terms of error minimization over probability distributions (thanks to Robert Staubs and Colin Wilson for discussion).

results). All other constraints were assigned zero weights.

(12)	*Weak*	*Strong*
$*V{:}CC_{[-cor]}$	4.67	0.60
$*CC_{[-cor]}$	0.58	0.58
$*_{\sigma}V{:}CC_{[-cor]}$	0	0.27

Our goal is for the learned grammar to reflect the pattern of data from our experiment. Table 11 repeats the probabilities of long vowel choices in each of the four experimental contexts in the column labeled '*p* long'. In none of the contexts does this probability reach 0.5, even though there is no general bias against long vowels in the lexicon. As we discussed in section 3, this is plausibly the result of an experimentally induced response bias. We treat it as such here, and instead of trying to model it in the grammar, we provide scaled probabilities in the column labeled '$p \times 1.22$' which multiply the probabilities by 0.5/0.41 = 1.22, resulting in the highest probability of long vowel choice being equal to 0.5. The grammatical probabilities of long vowel choices in each context were derived by dividing the probability that the grammar assigns the long vowel by the sum of the probabilities of it and the short vowel (i.e. by using the "Luce choice rule"; Luce 1959). The probabilities thus derived using the 'Strong' grammar come close to matching the scaled experimental probabilities; the grammar learned with the weak regularization term does not work as well.[9]

Context	*p* long	$p \times 1.22$	*Strong*	*Weak*
$\_CC_{[-cor]}\#$	0.27	0.33	0.295	0.001
_C#	0.41	0.5	0.5	0.5
$\_CC_{[-cor]}V$	0.32	0.39	0.354	0.001
_CV	0.38	0.46	0.5	0.5

Table 11. Experimental data and probability of long vowel choices derived from grammars

[9] We also ran the same learning simulations with the lexical data used for the first regression analysis, which included a broader range of bisyllables. The probabilities assigned by the resulting grammar did not match the experimental data as well as those reported in the text. This might indicate that the subjects calculated probabilities over just the subset of the lexicon that shared a similar second syllable.

The Strong grammar better matches the experimental probabilities for two reasons. First, the differences between the _CC$_{[-cor]}$ and _C contexts is not as sharp in the experimental data as it is in the lexicon. The Weak grammar, which more closely matches the lexicon, has a relatively high weight of 4.67 on the *Vː CC$_{[-cor]}$ constraint, which results in a very low probability of long vowel choice in the two _CC$_{[-cor]}$ contexts (0.001). Because the Strong grammar was learned with a much bigger penalty on high weights, the *Vː CC$_{[-cor]}$ constraint got a much lower weight of 0.60, which results in a higher probability for long vowel choices in the _CC$_{[-cor]}$ contexts. It's worth noting that regularization is not crucial here: the result of experimental judgments being more diffuse than lexical probabilities could have been obtained in other ways (see e.g. Hayes *et al.* 2009 who use a Temperature parameter for this purpose). Second, and more importantly, the Strong grammar differentiates between the two _CC$_{[-cor]}$ contexts because it assigns a non-zero weight to *$_{\sigma}$Vː CC$_{[-cor]}$, the version of the constraint relativized to the tautosyllabic context. This constraint acquired weight because the L2 prior penalizes the sum of squared weights, so that weight spread between *$_{\sigma}$Vː CC$_{[-cor]}$ and *Vː CC$_{[-cor]}$ is preferred to weight on *Vː CC$_{[-cor]}$ alone.

5. Conclusions

The results of our experiments show that native speakers have internalized a phonological constraint that refers to a sequence of three segments. The fact that this constraint cannot be construed as a combination of two biphone constraints poses problems for both *n*-gram models of phonotactics, discussed in the introduction, as well as for phonological models that induce constraints only of biphone size (see Pierrehumbert 2003).[10] A second aspect of our experimental results is that the internalized representation of this phonotactic constraint is stronger in the word-final than the word-internal context, which we take to indicate its sensitivity to syllable structure. While the Hayes and Wilson (2008) model of phonotactics is in principle compatible with syllable-based phonotactics, their learning model does not induce syllabically conditioned constraints. In our learning simulation, we showed that with such constraints, a Maximum Entropy learner trained on corpus data from Dutch ends up with a grammar that matches the

[10] Pierrehumbert (2003: 218) hedges on whether a purely biphone-based theory of phonotactics is feasible, suggesting that triphone constraints might sometimes be induced. Pierrehumbert does not offer a fully explicit account of the conditions under which triphone (or even biphone) constraints are posited, so it is difficult to know whether her theory would permit *Vː CC$_{[-cor]}$.

distinctions found in the human judgment data. This supports the general point made by both Pierrehumbert (2003) and Hayes and Wilson (2008) that probabilistic models, which are motivated by the gradience of phonotactic judgments, can and should operate over the sorts of representations developed in phonological theory.

References

Albright, Adam (2009). Feature-based generalisation as a source of gradient acceptability. *Phonology* **26**, 9-41.

Anttila, Arto (2008). Gradient phonotactics and the Complexity Hypothesis. *Natural Language and Linguistic Theory* **26**, 695–729.

Bailey, Todd, and Ulrike Hahn. 2001. 'Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods?', *Journal of Memory and Language* **44**, 568–591.

Booij, Geert (1995). *The Phonology of Dutch*. Oxford: Oxford University Press.

Douglas Bates, and Martin Maechler (2010). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-34. http://CRAN.R-project.org/package=lme4.

Baayen, Harald (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Berent, Iris, Daniel L. Everett, and Joseph Shimron. 2001. Do Phonological Representations Specify Variables? Evidence from the Obligatory Contour Principle. *Cognitive Psychology* 42, 1-60.

Boersma, Paul, and David Weenink (2009). Praat: doing phonetics by computer (Version 5.1.05) [Computer program]. Retrieved May 1, 2009, from http://www.praat.org/

Booij, Geert (1995). *The Phonology of Dutch*. Oxford: Oxford University Press.

Burnham, Kenneth P. and David R. Anderson (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. New York: Springer-Verlag.

Chomsky, Noam, and Morris Halle (1968). *The Sound Pattern of English*. New York: Harper Row.

Cohen, Antonie, Carl Ebeling, Klaas Fokkema, and André van Hoik. 1971. *Fonologie van het Nederlands en het Fries*. The Hague: Nijhoff.

Coetzee, Andries. (2009). Grammar is both categorical and gradient. In Steven Parker, ed., *Phonological Argumentation: Essays on Evidence and Motivation*. London: Equinox. 9-42.

Coetzee, Andries, and Joe Pater. (2006). Lexically ranked OCP constraints in Muna. Ms, University of Michigan and University of Massachusetts, Amherst.

Coetzee, Andries, and Joe Pater (2008). Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* **26**, 289–337.

Coleman, John, and Janet Pierrehumbert (1997). Stochastic phonological grammars and acceptability. In *Computational Phonology. Third Meeting of the ACL Special Interest Group in Computational Phonology*, Association for Computational Linguistics, Somerset, NJ, pp. 49–56.

Daland, Robert, Bruce Hayes, James White, Marc Garellek, Andrea Davis and Ingrid Norrmann. 2011. Explaining sonority projection effects. *Phonology* **28**, 197 - 234

de Groot, A.W. (1931). De wetten der phonologie en hun betekenis voor de studie van het Nederlands. *Nieuwe Taalgids* **25**, 225-243.

Frisch, Stephan, Nathan Large, and David Pisoni (2000). Perception of wordlikeness: Effects of segment probability and length on the processing of nonce words. *Journal of Memory and Language* **42**, 481–496.

Frisch, Stephan, and Bushra Zawaydeh (2001). The psychological reality of OCP-Place in Arabic. *Language* **77**, 91–106.

Frisch, Stephan, Janet Pierrehumbert, and Michael Broe (2004). Similarity avoidance and the OCP. *Natural Language and Linguistic Theory* **22**, 179–228.

Greenberg, Joseph H., and James J. Jenkins (1964). Studies in the psychological correlates of the sound system of American English. *Word* **20**, 157-177.

Gussenhoven, Carlos (2009). Vowel duration, syllable quantity, and stress in Dutch. In Kristin Hanson and Sharon Inkelas (eds). *The Nature of the Word. Essays in Honor of Paul Kiparsky*. Cambridge, MA/London, UK: MIT Press. 181-198.

Hayes, Bruce. (2004). Phonological acquisition in Optimality Theory: the early stages. In René Kager, Joe Pater and Wim Zonneveld (eds.) *Constraints in Phonological Acquisition. Cambridge*, Cambridge University Press. pp. 158-203

Hayes, Bruce, and Colin Wilson (2008). A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* **39**, 379–440.

Bruce Hayes, Kie Zuraw, Péter Siptár, and Zsuzsa Londe. 2009. Natural and Unnatural Constraints in Hungarian Vowel Harmony. *Language* 85. 822-863.

Hulst, Harry van der (1984). *Syllable structure and stress in Dutch*. Dordrecht: Foris.

Jaeger, T. Florian. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59, 434-446.

Jurafsky, Daniel, and James Martin. (2000). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Prentice-Hall.

Kager, René (1989). *A Metrical Theory of Stress and Destressing in English and Dutch*. Dordrecht: Foris.

Kager, René, and Wim Zonneveld (1986). Schwa, syllables, and extrametricality in Dutch. *The Linguistic Review* **5**, 197-221.

Kullback, Solomon and Richard Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics* 22(1): 79-86.

Luce, R. Duncan (1959). *Individual choice behavior: a theoretical analysis*. New York: Dover.

Martin, Andrew (to appear). Grammars leak: Modeling how phonotactic generalizations interact within the grammar. In *Language*.

Moulton, William (1962). The vowels of Dutch: Phonetic and distributional classes. *Lingua* **11**, 294–312.

Nooteboom, Sieb (1972). Production and perception of vowel duration: A study of durational properties of vowels in Dutch. Doctoral dissertation, Rijksuniversiteit Utrecht.

Ohala, John J. (1986). Consumer's guide to evidence in phonology. *Phonology Yearbook* 3, 2-36.

Ohala, John J., and Manjari Ohala (1986). Testing hypotheses regarding the psychological manifestation of morpheme structure constraints. In J. Ohala and J. Jaeger (eds.) *Experimental Phonology.* Orlando, FL: Academic Press. 239–252.

Oostendorp, Marc van (2000). *Phonological Projection*. Berlin: Mouton De Gruyter.

Pater, Joe (to appear). Canadian raising with language-specific weighted constraints. In *Language*.

Pater, Joe, and Anne-Michelle Tessier (2006). L1 phonotactic knowledge and the L2 acquisition of alternations. In Roumyana Slabakova, Silvina Montrul, and Philippe Prévost (eds.), *Inquiries in Linguistic Development: Studies in Honor of Lydia White.* John Benjamins Publishing Company*,* Amsterdam/Philadelphia. 115-131.

Pierrehumbert, Janet (2003). Probabilistic phonology: Discrimination and robustness. In R. Bod, J. Hay and S. Jannedy (eds.) *Probability Theory in Linguistics*. Cambridge MA: The MIT Press. 177–228.

Prince, Alan, and Paul Smolensky (1993/2004). *Optimality Theory: Constraint interaction in generative grammar*. Technical Report, Rutgers University and University of Colorado at Boulder, 1993. Revised version published by Blackwell, 2004.

Prince, Alan, and Bruce Tesar. 2004. Learning phonotactic distributions. In René Kager, Joe Pater and Wim Zonneveld (eds.) *Constraint in Phonological Acquisition*. Cambridge: Cambridge University Press. pp. 245-291.

R Development Core Team (2010). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org

Rietveld, A.C.M. and V. J. van Heuven. 2009. *Algemene Fonetiek.* Bussum: Coutinho.

Rietveld, Toni, Joop Kerkhoff, and Carlos Gussenhoven (2004). Word prosodic structure and vowel duration in Dutch. *Journal of Phonetics* **32**, 349-371.

Smolensky, Paul, and Géraldine Legendre (2006). *The Harmonic Mind.* Cambridge MA: MIT Press.

Trommelen, Mieke (1983). *The Syllable in Dutch: With special reference to diminutive formation*. Dordrecht: Foris Publications.

Wilson, Colin and Marieke Obdeyn. 2009. Simplifying subsidiary theory: statistical evidence from Arabic, Muna, Shona, and Wargamay. Ms, The Johns Hopkins University.

Appendix. Stimuli used for the nonword judgments

LND CVC	lexical neighborhood density of the short vowel item
LND CVːC	lexical neighborhood density of the long vowel item
CD CVC	cohort density of the short vowel item
CD CVːC	cohort density of the long vowel item
2TP short	biphone transitional probability of the short vowel item
2TP long	biphone transitional probability of the long vowel item
TP2 delta	difference in biphone transitional probability between short and long vowel items

CV(ː)C monosyllables

short	long	LND short	LND long	CD short	CD long	2TP short	2TP long	2TP delta
bɑm	baːm	2.234	1.849	6.38	6.51	-0.975	-1.030	0.055
dɛm	deːm	1.979	1.822	6.71	101.19	-1.210	-1.231	0.021
dɪl	dil	2.049	2.315	2.99	21.00	-1.145	-1.111	-0.034
fɑr	faːr	1.526	2.385	20.98	9.17	-1.210	-1.151	-0.059
fɔl	foːl	2.115	2.021	31.24	6.67	-1.143	-1.218	0.075
hʏm	hum	2.135	2.126	7.48	0.60	-1.044	-1.204	0.160
jɛl	jeːl	2.063	2.035	0.00	0.78	-1.193	-1.381	0.188
jɪr	jiːr	2.424	2.522	0.00	0.00	-1.572	-1.619	0.047
jʏr	jyːr	2.373	2.587	9.13	17.44	-1.180	-1.239	0.059
kɛm	keːm	1.735	2.034	23.17	1.26	-1.100	-1.191	0.091
kʏm	kum	2.021	2.017	1.15	5.42	-1.140	-1.209	0.069
lɪm	lim	1.834	2.470	11.11	18.54	-1.162	-1.221	0.059
lʏm	lum	2.057	1.854	8.04	1.90	-1.124	-1.307	0.183
lʏr	lyːr	1.788	2.107	3.18	1.70	-1.158	-1.199	0.041
nɪl	nil	1.749	2.416	0.00	0.00	-1.246	-1.274	0.028
nɔl	noːl	2.278	1.841	2.13	0.00	-1.243	-1.232	-0.011
pʏm	pum	1.660	1.765	3.89	2.16	-1.083	-1.221	0.138
rɪr	riːr	1.823	2.485	0.00	0.00	-1.248	-1.095	-0.153
sɑl	saːl	2.184	2.223	15.84	54.45	-1.020	-1.042	0.022
sɔr	soːr	1.705	2.207	18.57	25.18	-1.071	-1.156	0.085
ʃɔl	ʃoːl	1.886	2.050	0.00	0.00	-1.204	-1.232	0.028
tɑr	taːr	1.774	2.662	26.29	22.50	-1.063	-1.064	0.001
tɪr	tiːr	2.173	2.351	0.00	29.92	-1.327	-1.151	-0.176
xɑm	xaːm	2.123	1.962	8.54	5.84	-1.100	-1.132	0.032
ʋɛm	ʋeːm	2.120	1.764	0.60	16.76	-1.086	-1.139	0.053
ʋɔr	ʋoːr	2.299	2.614	93.27	68.61	-1.115	-1.267	0.152
xɛm	xeːm	1.654	1.763	5.82	19.27	-1.125	-1.192	0.067
xɔl	xoːl	2.112	1.874	52.43	12.47	-1.072	-1.118	0.046
zɑm	zaːm	2.113	1.873	1.93	3.94	-1.205	-1.225	0.020
zɛl	zeːl	2.493	2.124	225.47	15.06	-1.114	-1.070	-0.044
mean		**2.016**	**2.137**	**19.54**	**15.61**	**-1.156**	**-1.197**	**0.041**
t-test		p = 0.082		p = 0.65		p= 0.15		

CV(ː)CC monosyllables

short	long	LND short	LND long	CD short	CD long	2TP short	2TP long	2TP delta
bɑmk	baːmk	2.189	1.034	6.38	6.51	-1.288	-1.332	0.044
dɛrx	deːrx	2.458	1.307	59.20	20.96	-1.404	-1.419	0.015
dɪlp	dilp	0.727	1.845	2.99	21.00	-1.330	-1.303	-0.027
fɪlk	filk	2.636	1.870	85.32	57.77	-1.327	-1.323	-0.004
fɔlm	foːlm	1.755	2.731	31.24	6.67	-1.321	-1.380	0.059
hɑmk	haːmk	1.777	1.561	15.32	19.19	-1.323	-1.369	0.046
hɔlm	hoːlm	1.693	1.808	52.90	13.73	-1.212	-1.239	0.027
jʏlm	julm	4.043	0.670	4.04	9.50	-1.298	-1.447	0.149
jʏrm	jyrm	2.116	0.000	9.13	17.44	-1.290	-1.336	0.046
jɛrx	jeːrx	2.445	0.000	0.00	2.16	-1.473	-1.616	0.143
jɔlm	joːlm	0.778	0.000	0.60	5.43	-1.370	-1.313	-0.057
kɛmk	keːmk	2.039	4.391	23.17	1.26	-1.388	-1.461	0.073
kɪrf	kirf	1.044	1.632	6.39	13.24	-1.385	-1.277	-0.108
kɪrp	kirp	1.314	2.054	6.39	13.24	-1.397	-1.289	-0.108
lʏmk	lumk	0.000	0.699	8.04	1.90	-1.407	-1.553	0.146
lʏrm	lyrm	1.079	1.699	3.18	1.70	-1.272	-1.305	0.033
lʏrp	lyrp	0.602	1.699	3.18	1.70	-1.333	-1.366	0.033
mɪrp	mirp	0.984	2.220	2.87	28.19	-1.438	-1.272	-0.166
nɪmk	nimk	0.903	0.000	10.51	9.74	-1.562	-1.589	0.027
pɛrx	peːrx	2.231	1.541	217.41	49.87	-1.308	-1.354	0.046
pɔlm	poːlm	1.554	1.884	37.99	178.10	-1.248	-1.270	0.022
sɑmk	saːmk	0.301	2.660	5.46	275.11	-1.375	-1.404	0.029
tʏmp	tump	1.036	0.349	1.80	1.30	-1.147	-1.262	0.115
vɑmk	vaːmk	2.241	0.477	8.94	0.48	-1.414	-1.424	0.010
ʋɛmk	ʋeːmk	2.117	2.143	0.60	16.76	-1.377	-1.419	0.042
xɑmk	xaːmk	1.785	0.000	8.54	5.84	-1.388	-1.413	0.025
xɛmk	xeːmk	2.324	2.158	5.82	19.27	-1.408	-1.461	0.053
xɔlm	xoːlm	1.873	1.614	52.43	12.47	-1.264	-1.301	0.037
zɑmk	zaːmk	3.443	3.213	1.93	3.94	-1.472	-1.488	0.016
zɔlm	zoːlm	2.241	0.477	19.45	7.38	-1.329	-1.320	-0.009
mean		**1.724**	**1.579**	**23.04**	**27.40**	**-1.352**	**-1.377**	**0.025**
t-test		p = 0.30		p = 0.74		p= 0.29		

CV(ː)CəL disyllables

short	long	LND short	LND long	CD short	CD long	2TP short	2TP long	2TP delta
bɑmər	baːmər	2.046	2.056	6.38	6.51	-0.925	-0.961	0.036
dɛmər	deːmər	1.823	1.609	6.71	101.19	-1.081	-1.095	0.014
fɑrəl	faːrəl	2.061	1.914	20.98	9.17	-1.069	-1.030	-0.039
fɪrəl	fiːrəl	0.000	1.173	9.73	6.86	-1.167	-1.037	-0.130
hɛmər	heːmər	1.418	1.685	20.53	56.94	-0.974	-1.030	0.056
jʏrəl	jyːrəl	0.000	1.415	9.13	17.44	-1.050	-1.088	0.038
jɔlər	joːlər	1.412	1.190	0.60	5.43	-1.093	-1.046	-0.047
kʏmər	kumər	2.381	2.118	1.15	5.42	-1.034	-1.080	0.046
kɛmər	keːmər	1.714	1.822	23.17	1.26	-1.008	-1.069	0.061
kɪrəl	kiːrəl	2.076	1.942	6.39	13.24	-1.088	-0.998	-0.090
lʏmər	lumər	1.571	1.412	8.04	1.90	-1.024	-1.146	0.122
lʏrəl	lyːrəl	1.517	1.699	3.18	1.70	-1.035	-1.062	0.027
lɪmər	limər	1.508	1.707	11.11	18.54	-1.050	-1.089	0.039
nɑrəl	naːrəl	0.977	2.297	20.62	30.44	-1.085	-1.054	-0.031
nɪlər	nilər	1.610	1.628	0.00	0.00	-1.121	-1.139	0.018
nɔrəl	noːrəl	2.445	1.658	49.96	77.20	-1.065	-1.089	0.024
pʏmər	pumər	1.991	1.944	3.89	2.16	-0.997	-1.089	0.092
pɑlər	paːlər	1.286	1.677	36.30	47.61	-0.952	-0.971	0.019
pɪlər	pilər	1.386	1.400	14.59	18.63	-0.992	-0.990	-0.002
pɔlər	poːlər	1.346	1.721	37.99	178.10	-0.992	-1.010	0.018
rɛlər	reːlər	1.802	1.094	7.10	57.74	-1.001	-1.028	0.027
sɪrəl	siːrəl	2.263	1.643	59.36	44.90	-1.106	-0.937	-0.169
sɔlər	soːlər	1.132	1.372	52.40	37.16	-1.030	-1.055	0.025
sɔrəl	soːrəl	2.178	1.415	18.57	25.18	-0.976	-1.033	0.057
tʏmər	tumər	1.237	1.481	1.80	1.30	-1.048	-1.144	0.096
vɑmər	vaːmər	2.087	2.829	8.94	0.48	-1.030	-1.038	0.008
wɛlər	weːlər	1.509	1.516	253.73	14.82	-0.945	-0.986	0.041
wɛmər	weːmər	1.440	1.546	0.60	16.76	-0.999	-1.034	0.035
xɑmər	xaːmər	1.776	1.993	8.54	5.84	-1.008	-1.029	0.021
xɔlər	xoːlər	1.060	1.022	52.43	12.47	-1.005	-1.036	0.031
mean		**1.568**	**1.666**	**25.13**	**27.21**	**-1.032**	**-1.046**	0.014
t-test		p = 0.44		p = 0.85		p= 0.29		

CV(ː)CCəL disyllables

short	long	LND short	LND long	CD short	CD long	2TP short	2TP long	2TP delta
bɑmvər	baːmvər	0.000	0.000	6.38	6.51	-1.205	-1.236	0.031
dɛmxər	deːmxər	0.000	0.000	6.71	101.19	-1.479	-1.491	0.012
dɪlkər	dilkər	2.468	1.954	2.99	21.00	-1.153	-1.134	-0.019
dɪlpər	dilpər	0.000	2.564	2.99	21.00	-1.176	-1.157	-0.019
fɑlbər	faːlbər	0.000	0.000	10.12	14.53	-1.338	-1.306	-0.032
fɑrxəl	faːrxəl	0.000	0.000	20.98	9.17	-1.349	-1.315	-0.034
fɑrvəl	faːrvəl	0.000	0.000	20.98	9.17	-1.175	-1.141	-0.034
hʏmvər	humvər	0.000	0.000	7.48	0.60	-1.244	-1.336	0.092
jʏrməl	jyrməl	0.477	0.000	9.13	17.44	-1.130	-1.164	0.033
jɔlbər	joːlbər	0.000	0.000	0.60	5.43	-1.336	-1.295	-0.040
jɔlmər	joːlmər	0.000	0.000	0.60	5.43	-1.214	-1.173	-0.040
kʏmbər	kumbər	0.000	0.000	1.15	5.42	-1.098	-1.137	0.039
kʏmkər	kumkər	0.000	0.000	1.15	5.42	-1.241	-1.280	0.039
kɛmxər	keːmxər	0.699	0.000	23.17	1.26	-1.417	-1.468	0.052
kɪrfəl	kirfəl	0.000	0.000	6.39	13.24	-1.238	-1.161	-0.077
kɪrxəl	kirxəl	0.000	0.000	6.39	13.24	-1.365	-1.288	-0.077
kɪrpəl	kirpəl	0.000	0.000	6.39	13.24	-1.197	-1.120	-0.077
lʏrbəl	lyrbəl	0.000	0.000	3.18	1.70	-1.212	-1.235	0.023
lɪmxər	limxər	0.000	0.000	11.11	18.54	-1.452	-1.486	0.034
nɔrbəl	noːrbəl	0.000	1.869	49.96	77.20	-1.238	-1.258	0.021
pɛlbər	peːlbər	0.000	0.000	22.12	12.44	-1.211	-1.250	0.039
rɛlxər	reːlxər	0.000	0.000	7.10	57.74	-1.312	-1.335	0.023
sɔrvəl	soːrvəl	0.000	0.000	18.57	25.18	-1.095	-1.144	0.049
tʏmpər	tumpər	1.158	0.778	1.80	1.30	-1.045	-1.127	0.082
wɛmxər	weːmxər	0.000	0.000	0.60	16.76	-1.409	-1.439	0.030
wɛmkər	weːmkər	2.690	1.114	0.60	16.76	-1.210	-1.240	0.030
xɑmkər	xaːmkər	0.000	0.000	8.54	5.84	-1.218	-1.236	0.018
xɑmvər	xaːmvər	0.000	0.602	8.54	5.84	-1.276	-1.294	0.018
xɔlbər	xoːlbər	0.000	0.000	52.43	12.47	-1.260	-1.286	0.026
xɔlmər	xoːlmər	0.000	0.000	52.43	12.47	-1.138	-1.164	0.026
mean		**0.250**	**0.296**	**12.35**	**17.58**	**-1.248**	**-1.257**	**0.009**
t-test		p = 0.79		p = 0.29		p= 0.75		