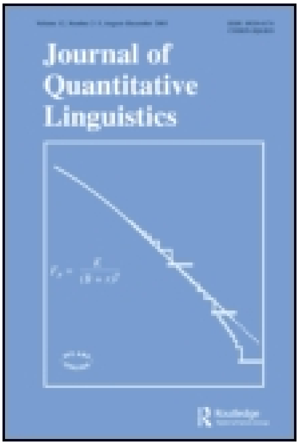

Journal of Quantitative Linguistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/njql20

Syllabification of American English: Evidence from a Large-scale Experiment. Part II

David Eddington [a], Rebecca Treiman [b] & Dirk Elzinga [a]
[a] Brigham Young University
[b] Washington University in St. Louis , USA
Published online: 22 Apr 2013.

Journal of Quantitative Linguistics, 2013
Vol. 20, No. 2, 75–93, http://dx.doi.org/10.1080/09296174.2013.773136

Syllabification of American English: Evidence from a Large-scale Experiment. Part II[1*]

David Eddington[1], Rebecca Treiman[2] and Dirk Elzinga[1]
[1]Brigham Young University; [2]Washington University in St. Louis, USA

ABSTRACT

4990 bi-syllabic English words were syllabified by about 22 native speakers who choose between different slash divisions (e.g. *photon*: FOW / TAHN, FOWT / AHN). Results for test items with one medial consonant are reported in Eddington, Treiman, & Elzinga (2013). In the present paper, the regression analysis of words with two, three, and four medial consonants are discussed. A model in which syllables are made to be as word-like as possible is supported; syllables are made that begin and end in the same phonemes and graphemes that are legal word-initially and finally. Syllabifications also coincide with morphological boundaries. In words with two medial consonants, stressed syllables attract consonants, and a sonorant first consonant is more likely to be placed in the coda than an obstruent. Clusters comprised of /s/C differ from other two consonant clusters that are legal word-initially; the former tend to be divided between syllables while the latter are placed in the onset.

1. INTRODUCTION

In Eddington et al. (2013) we describe an online experiment in which subjects chose between different divisions of bi-syllabic English words (e.g. *photon:* FOW / TAHN, FOWT / AHN). The purpose of that study is to determine what predictor variables influence syllabification preferences of words, such as *photon*, that contain a single medial consonant. In the present paper, we continue our discussion by analysing words, such as *pastor*, *outgrow* and *obstruct*, that have from two to four medial consonants.

*Address correspondence to: David Eddington, Dept. of Linguistics and English Language, 4064 JFSB, Provo, UT 84660, USA. Tel. 801-422-7452. Email: eddington@byu.edu

We investigate how a number of variables affect syllable division in words with internal consonant clusters.

2. SYLLABIFICATION OF WORDS WITH TWO MEDIAL CONSONANTS

The total number of words with two medial consonants that were analysed was 1941. There were 42,888 responses to these words, from which the 212 *I'm not sure* responses were eliminated before analysis, as were the 161 responses to the seven ambi-morphemic words. In 15% of the responses, both medial consonants were placed in the second syllable (.CC responses); in 79%, the two consonants were placed in separate syllables (C.C responses), and in 5%, both consonants were placed in the first syllable (CC. responses). Eighty percent of the words were syllabified in the same way by 80% or more of the subjects. Three separate mixed effects logistic regression analyses were performed to determine what factors influence .CC syllabifications, C.C syllabifications and CC. syllabifications. Test item and subject were included as random intercepts.

For these analyses, we considered the characteristics of the vowel in the first and second syllables as we did in the analysis of the words with a single medial consonant. The first medial consonant was identified as either sonorant or obstruent; further subdividing into nasal, lateral, etc. could not be done since it would result in too many empty cells in the analysis. Medial consonant sequences were put into one of two categories: (1) clusters that are phonologically legal word-initially such as [tɹ] (compare to an illegal initial cluster such as [bv]); or (2) clusters that are phonologically legal word-finally such as [mp] (compare to illegal word-final [tp]). This division was done for the sake of simplicity; we recognize that some consonants clusters are accepted as better word-initial clusters than others, as experimental data have shown (Hayes & Wilson, 2008).

In order to examine orthographic effects, we distinguished cases in which the medial spelling sequence that included the consonant cluster was legal in word-initial position (e.g. *abridge, quisling*), legal in word-final position (e.g. *fountain, gambol*), illegal initially (e.g. *conjure, truckload*), or illegal finally (e.g. *acquire, asleep*). In some cases (e.g. *colonel, likely*, and *every*), orthographic vowels interrupt the consonant cluster. Since these words behaved like the illegal clusters, they were merged with illegal clusters for the analyses. Some words were classified similarly on the phonological legality and orthographic legality variables, but many were not. For example, the [zl] cluster of *quisling* is illegal phonologically in word-initial position but the *sl* cluster is orthographically legal in this position.

The type of morpheme boundary (transparent, opaque, compound) was not independent from a number of other variables, which meant that this variable had to be modified. Therefore, we coded only whether there was a morpheme boundary of any type and whether it coincided with the syllabification being analysed or not. Monomorphemic words appeared in a separate category.

2.1 Results for Words with Two Medial Consonants

Results of the regression analyses appear in Table 1. This and subsequent tables do not include the number of responses in each category in order to reduce the size of the tables and make them more readable. Dashes in the table indicate that a variable was not included in an analysis. This occurs in the case of the CC. analysis, for example, where the legality of a consonant cluster word-initially is not relevant, just as the legality of a word-final cluster is not relevant in the .CC analysis. An “ns” in this and subsequent tables indicates a variable that was not found to improve the model in the stepping up and down runs to a statistically significant level ($p < 0.05$).

One of the strongest is the phonological legality of the consonant cluster in word-initial and final position. Table 1 shows that consonant clusters that are legal word-initially tend to be placed together in the onset of the second syllable. In like manner, clusters that are legal word-finally favour placement into the coda of the first syllable. This contrasts with Treiman & Zukowski (1990) who observed legal word-final clusters being divided. Clusters tend to be divided between syllables when they are illegal word-initially, and legal word-initial clusters strongly disfavour being divided between syllables.

Morphology also exerts a robust influence on syllabification. In the analysis of variables that favour .CC syllabification, for example, the high log odds of 1.76 indicates that a morpheme boundary before the consonant cluster (e.g. *tea* + *spoon*) is strongly associated with placement of the syllable boundary before the cluster. When a morpheme boundary occurs elsewhere in the word (e.g. *part* + *ing*) .CC is strongly disfavoured, as indicated by the log odds of −1.42. Monomorphemic words have log odds closer to zero for each syllabification analysed.

Table 1. Logistic regression analysis of factors that influence the syllabification of words with two medial consonants.

	.CC*		C.C		CC.	
	Log odds	%	Log odds	%	Log odds	%
Phonological legality of consonant cluster (.CC p = 4.96e-287; C.C p = 8.9e-238; CC. p = 3.42e-88)						
Legal word-initially	1.44	51	−1.24	46	–	
Illegal word-initially	−1.44	4	1.24	91	–	
Legal word-finally	–		–		0.61	8
Illegal word-finally	–		–		−0.61	2
Morphological boundary (.CC p = 4.81e-38; C.C p = 7.72e-78; CC. p = 6.85e-49)						
Coincides with syllabification analysed	1.76	91	1.51	94	1.17	16
Monomorphemic	−0.34	17	−0.20	77	−0.08	5
Does not coincide with syllabification analysed	−1.42	6	−1.31	60	−1.09	2
Quality of first consonant of cluster (.CC p = 4.15e-47; C.C p = 8.96e-26; CC. ns)						
Obstruent	0.90	31	−0.51	65	ns	4
Sonorant	−0.90	3	0.51	91	ns	6
Legality of first vowel word-finally (.CC p = 3.95e-10; C.C p = 7.57e-09; CC. ns)						
Legal	0.27	19	−0.20	77	ns	5
Illegal	−0.27	12	0.20	82	ns	6
Stress (.CC p = 0.00101; C.C p = 0.0105; CC. p = 4.3e-09)						
Initial	−0.14	11	0.11	83	0.28	6
Final	0.14	31	−0.11	67	−0.28	3
Orthographic legality (.CC ns; C.C ns; CC. p = 0.000291)						
Legal word-initially	ns	34	ns	62	–	
Illegal word-initially	ns	6	ns	89	–	
Legal word-finally	–		–		0.26	7
Illegal word-finally	–		–		−0.26	3
Quality of the second vowel (.CC p = 0.0247; C.C ns; CC. p = 0.0102)						
Tense	0.10	20	ns	76	−0.12	4
Lax	−0.10	13	ns	82	0.12	6

Notes: *The .CC analysis was run only with subjects having random intercepts, but not with the test word, since in the latter case the model failed to converge. %, percent of words in a category with a particular syllabification.

Table 2. Percentage of responses given to polymorphemic words when a particular type of morpheme boundary appears in a particular syllable position.

	.CC	C.C	CC.
Compound	91	94	47
Transparent	91	91	16
Opaque	82	86	6

Although type of morpheme boundary could not be included in the statistical analysis for reasons mentioned above, inspection of the results for polymorphemic words of various types, as shown in Table 2, suggests that morpheme boundaries in compound words are highly influential. This is expected under the hypothesis that syllables are made to be as word-like as possible, since this division results in actual words. Boundaries in compound words are followed in turn by transparent and opaque boundaries. The data indicate, for example, that when a compound morpheme boundary appears in the C.C position, 94% of the responses divided the word between the two consonants. Words with a transparent morpheme boundary between the consonants garnered 91% C.C responses, in comparison with words with an opaque boundary, which yielded 86% C.C responses.

2.2 Results for [s]C clusters

Of special interest are [s]C clusters. Although they exist word-initially (e.g. *spike*, *slide*), the combination of [s] followed by a stop violates the ideal sonority profile (Selkirk, 1982; Clements, 1990; Giegerich, 1992). In contrast [sl, sj, sw, sm, sn] have a typical sonority profile. While the sonority interpretation is most widely accepted, Wright (2004) argues that what makes [s]C clusters different from other clusters has nothing to do with sonority. Instead, he argues that [s] has enough perceptual acoustic cues by itself that it does not depend on formant transitions in the following vowel in order to be perceived. This property allows it to be separated from a vowel by an intervening (low sonority) consonant, yet be part of the same syllable.

In any event, whether [s]C clusters behave differently from other legal clusters may be tested. In order to do so, we examined all morphologically simple words that have clusters that are both orthotactically and phonotactically legal word-initially, and in this way controlled for these variables. We contrasted the [s]C clusters with other legal clusters. The results of the analysis appear in Table 3. The number of CC. responses was too small to be statistically analysed. Of the variables that were not factored out, the quality of the second vowel was not a significant factor, but stress and the legality of the first vowel word-finally were. The most important finding is that [s]C clusters strongly favour being divided between syllables. This contrasts with other consonant clusters that are legal word-initially, which demonstrate a strong tendency to be placed together in the onset of the second syllable. The difference in syllabification between [s]C clusters and other CC clusters confirms the previous experimental findings of Redford & Randall (2005) and Treiman, Gross, & Cwikiel-Glavin (1992).

The idea that [s]C clusters are generally divided between syllables in order to provide an ideal sonority contour could be tested further by comparing [s]C clusters that are [s]+obstruent against [s]+sonorant clusters. The latter do not violate sonority and would be expected to have a .CC syllabification. Unfortunately, such an analysis is impossible because there is only one word containing an [s] + sonorant cluster once words containing illegal word-initial onsets such as [sɹ], and morpheme boundaries (e.g. *dislike*) are eliminated.

Table 3. Logistic regression analysis that compares [s]C and other consonant clusters that are legal word-initially.

	.CC		C.C	
	Log odds	%	Log odds	%
Type of consonant cluster (.CC p = 2.93e-37; C.C p = 5.01e-23)				
sC-	−1.17	33	0.92	60
Other	1.17	79	−0.92	21
Stress (.CC p = 8.16e-13; C.C p = 2.77e-14)				
Initial	−0.73	50	0.65	46
Final	0.73	75	−0.65	23
Legality of first vowel word-finally (.CC p = 4.61e-08; C.C p = 1.37e-07)				
Legal	0.53	77	−0.47	21
Illegal	−0.53	47	0.47	49
Quality of the second vowel (.CC ns; C.C ns)				
Lax	ns	52	ns	44
Tense	ns	69	ns	28

Note: %, percent of those words with a particular syllabification.

Returning to the data in Table 1, we see that sonority played a significant role. As far as the first consonant of the cluster is concerned, the sonority sequencing principle (Clements, 1990) was followed fairly closely. Obstruents were favoured as the first consonant in .CC syllabifications over sonorants. Although we could not determine the influence of individual sonorants, the raw data reveal some tendencies. The percentage of .CC responses was 1% for the 362 words with [ɹ] as the first consonant of the cluster, 3% for the 125 words with [l], 3% for the 607 words with nasals, and 31% for the 847 words with obstruents. In contrast, sonorants often appeared as the first consonant in C.C syllabifications: 91% with [ɹ] as the first consonant of the cluster, 90% with [l], 90% with nasals, and 65% with obstruents. The quality of the first consonant was not statistically significant in the CC. analysis.

The legality of consonant clusters is one phonotactic influence that is manifest in the data. The legality of the first vowel word-finally is another. Vowels that are attested word-finally favour .CC presumably because they do not need a consonant in the coda to be phonotactically legal. Those that are not attested word-finally, on the other hand, pull a consonant into their coda and favour C.C syllabification. This appears to be another instance of syllables being formed in accordance with word-level phonotactics. Primary word stress appears to attract consonants into its syllable as well. Initially stressed words favour C.C and CC. in contrast to words with final stress that favour .CC.

Although orthographic and phonological legality are identical in many cases, there is enough difference that their effects may be evaluated separately. Phonological legality is highly significant in all three analyses, judging by the *p* values. Orthographic legality, on the other hand, is only significant in the CC. analysis where letter clusters that are legal word-initially favour .CC syllabification. Finally, the vowel quality of the nucleus of the second syllable affects syllabification to a small degree; tense second vowels favour .CC syllabifications while disfavouring CC.

2.3 Discussion of Words with Two Medial Consonants

The majority of research on English syllabification has centred on single medial consonants. Our analysis of words with two consonants expands on previous research and is revealing in a number of ways. The evidence points to a strong tendency to parse syllables so that they are as word-like as possible, in the sense that they are either identical to words or are plausible words. Syllabification is carried out with a strong influence of the phonotactics of English words. In some cases, orthotactic influence is also registered. Syllables containing vowels that are not legal word-finally are closed by consonants, rendering them more wordlike. In like manner, when words and morphemes exist within a word (e.g. *tank+er, rhythm+ic*), syllable boundaries are made that identify those words and morphemes.

Syllabification is also influenced by sonority. It often appears to be carried out so that an ideal sonority contour is maintained throughout the syllable. This is seen in the tendency for the first consonant in a cluster to be placed in the coda of the first syllable if it is a sonorant, but not if it is an obstruent. There is some interplay between legality and maintenance of an ideal sonority contour. For example, [s]C clusters are more often divided between syllables than other clusters that are attested word-initially, which may be done to maintain the ideal sonority profile. On the other hand, [s]C clusters are legal word-initially, a fact that may influence them to remain together in the onset.

Our findings concur with that of previous research regarding the pull of stress on word-medial consonants. The fact that the quality of the vowel in the second syllable exerts a small influence, although difficult to interpret, suggests that previously unrecognized factors may play a part in syllabification.

3. SYLLABIFICATION OF WORDS WITH THREE MEDIAL CONSONANTS

We analysed data from 447 words with medial clusters consisting of three consonants. From the 9895 responses we removed the 44 *I'm not sure* responses, and the 43 responses to the three words with an ambimorphemic consonant prior to analysis. Responses in which all three consonants were placed at the beginning of the second syllable, or .CCC syllabifications, constituted 6% of all responses to these words. C.CC syllabifications made up 57% of the total, and CC.C syllabifications 37%. Only 63 responses placed all three consonants at the end of the first syllable, too small a sample to be informative. Of these 447 words, 69% were syllabified in the same way by 80% or more of the subjects. Three mixed effects logistic regression analyses were performed each with a different dependent variable: .CCC syllabifications, C.CC syllabifications, and CC.C syllabifications.

A number of variables were non-independent, and finding a parsimonious statistical model was challenging. In part, this was achieved by considering orthographic legality and phonological legality in separate analyses. Stress was not completely independent from a number of other variables, and for this reason was removed from consideration in the .CCC analysis. This step was justified by an inspection of the raw data that suggested that stress was not a highly relevant factor there. As in the analysis of words with two medial consonants, we included the legality of the first vowel, the quality of the second vowel, and the existence and placement of morpheme boundaries.

In order to examine the effect of the phonological legality of the medial consonant clusters, the analysis for each type of response included a cluster legality variable. For the .CCC analysis, this indicated whether or not the three phoneme sequence was legal word-initially. Of course, all legal combinations consist of /s/ + obstruent + sonorant clusters. For the analysis of C.CC syllabifications, we coded whether the second and third consonants of the medial sequence formed a legal word-initial cluster (C CC). For example, the cluster in *central* was considered legal because [tɹ] is legal word-initially, while the [ktʃ] cluster in *tincture* was considered illegal. For the analysis of CC.C responses, the cluster legality variable coded whether the first and second consonants of the medial sequence (CC C) could appear in word-final position. The cluster in *central* was coded as legal because words may end with [nt]. The [tg] in *outgrow*, on the other hand, was coded as illegal.

To examine the possible effects of spelling, an orthographic legality variable was included for each dependent variable. A word's spelling was considered legal if, when divided in the manner under consideration, the letters and letter groups were legal in the relevant positions of the spelling. In the .CCC analysis, *str* in *nostril* was considered legal and *rnk* in *turnkey* as illegal. For the analysis of C.CC responses, the consonants in *warbler* and *courthouse* were coded as legal because words may begin with *bl* and *th*. This appears as C CC in the tables. The *dl* sequence in *groundling* was coded as illegal since words may not begin with *dl*. The orthographic legality of the first two consonants in word-final position (CC C) was considered in the CC.C analysis. Many words were classified the same way on the spelling legality and phonological cluster legality variables, but not all. For example, the [th] of *courthouse* is not a legal phonological cluster at the beginning of a word but *th* is a legal spelling in this position.

3.1 Results for Words with Three Medial Consonants

Results of the analyses appear in Tables 4 and 5. Due to a lack of independence among variables, Table 4 includes orthographic legality and excludes phonological legality. Table 5 includes phonological legality, but excludes orthographic legality. As evidenced by the *p* values in Tables 4 and 5, morphology exerted a strong influence on syllabification. Syllabification at morpheme boundaries is strongly favoured in the all the analyses. When morpheme boundaries do not coincide with the place of syllabification analysed, syllabification there is highly disfavoured. Although we could not

Table 4. Logistic regression analysis of factors that influence the syllabification of words with three medial consonants. Phonological legality not included.

	.CCC		C.CC		CC.C	
	Log odds	%	Log odds	%	Log odds	%
Morphological boundary (.CCC p = 2.99e-12; C.CC p = 1.49e-70; CC.C p = 2.07e-88)						
Coincides with syllabification analysed	4.57	87	2.07	93	3.44	79
Monomorphemic	−1.85	8	0.19	68	−0.66	23
Does not coincide with syllabification analysed	−2.72	2	−2.27	24	−2.77	6
Orthographic legality (.CCC p = 2.99e-12; C.CC p = 5e-17; CC.C ns)						
CCC legal word-initially	1.79	30	–		–	
CCC illegal word-initially	−1.79	3	–		–	
C CC legal word-initially	–		0.78	75	–	
C CC illegal word-initially	–		−0.78	34	–	
CC C legal word-finally	–		–		ns	38
CC C illegal word-finally	–		–		ns	36
Legality of first vowel word-finally (.CCC ns; C.CC ns; CC.C ns)						
Legal	ns	7	ns	51	ns	42
Illegal	ns	4	ns	62	ns	33
Quality of the second vowel (.CCC ns; C.CC ns; CC.C p = 0.000659)						
Tense	ns	7	ns	61	−0.34	32
Lax	ns	5	ns	55	0.34	40
Stress (C.CC ns; CC.C ns)						
Initial	–		ns	54	ns	41
Final	–		ns	69	ns	21

Note: %, percent of those words with a particular syllabification.

evaluate the type of morphological boundary in the statistical model, the raw percentages in Table 6 suggest a slight tendency for syllabification to be made to coincide with a morpheme boundary in compound words, followed by words with transparent boundaries (in the .CCC and C.CC analyses), then words with opaque boundaries.

Phonological legality and orthographic legality were also strong influences on syllabification. Legal word-initial clusters favoured .CCC syllabifications. When the last two members of the cluster comprise a legal word-initial phonological sequence, C.CC syllabifications were preferred.

Table 5. Logistic regression analysis of factors that influence the syllabification of words with three medial consonants. Orthographic legality not included.

	.CCC		C.CC		CC.C	
	log odds	%	log odds	%	log odds	%
Phonological legality of consonant cluster (.CCC p = 5.10e-05; C.CC p = 1.51e-71; CC.C p = 9.14e-07)						
CCC legal word-initially	2.33	42			–	
CCC illegal word-initially	−2.33	2			–	
C *CC* legal word-initially	–		1.27	76	–	
C *CC* illegal word-initially	–		−1.27	16	–	
CC C legal word-finally	–		–		0.79	43
CC C illegal word-finally	–		–		−0.79	4
Morphological boundary (.CCC p = 0.0125; C.CC p = 8.03e-29; CC.C p = 2.07e-88)						
Coincides with syllabification analyzed	2.40	87	1.65	93	3.16	79
Monomorphemic	−1.01	8	−0.07	68	−0.72	23
Does not coincide with syllabification analyzed	−1.38	2	−1.58	24	−2.44	6
Legality of first vowel word-finally (.CCC p = 5.19e-05; C.CC ns; CC.C ns)						
Legal	0.28	7	ns	51	ns	42
Illegal	−0.28	4	ns	62	ns	33
Quality of the second vowel (.CCC ns; C.CC ns; CC.C p = 0.000423)						
Tense	ns	7	ns	61	−0.35	32
Lax	ns	5	ns	55	0.35	40
Stress (C.CC ns; CC.C ns)						
Initial	–		ns	54	ns	41
Final	–		ns	69	ns	21

Note: %, percent of those words with a particular syllabification.

Table 6. Percentage of responses given to polymorphemic words when a particular type of morpheme boundary appears in a particular syllable position.

	.CCC	C.CC	CC.C	CCC.
Compound	90	93	83	No cases
Transparent	86	92	86	4
Opaque	No cases	No cases	47	3

CC.C syllabifications were favoured in cases in which the first two consonants in the cluster are legal word-finally. Spelling legality was also influential. Sequences of three letters that appear word-initially result in .CCC syllabifications. Sequences of two letters that are legal at the beginning of English words are preferred in C.CC syllabifications. However, the orthographic legality of the first two consonants of the cluster word-finally was not a significant determinant of CC.C syllabifications.

Only small effects were exerted by the vowels. Vowels that are legal word-finally were associated with more .CCC in one of the analyses (Table 5), while a lax vowel in the second syllable favoured CC.C syllabifications. The role of the second vowel in syllabification is not something that previous work on syllabification has addressed, but future research should include.

3.2 Discussion of Words with Three Medial Consonants

The results of the analysis of words with three medial consonants indicates no role of stress and little influence of the vowels surrounding the three medial consonants. Morphological boundaries, the phonological legality of the consonant clusters, and the orthographic legality of the consonant clusters are the most robust variables. Once again, when taken together these suggest a model in which people syllabify words to make the resulting syllables look like English words. Syllables are most highly favoured when they are actual words, as with compounds. They are also favoured when they are phonotactically and orthographically plausible, but not necessarily actual words.

4. SYLLABIFICATION OF WORDS WITH FOUR MEDIAL CONSONANTS

We had data for only 56 items containing four medial consonants. There were 1,244 responses to these items, of which only 5 were *not sure* responses. Responses in which all four consonants were placed at the beginning of the second syllable constituted less than 2% of all responses. C.CCC syllabifications made up 42% of the total, CC.CC syllabifications 37%, and CCC.C syllabifications 19%. No response placed all four consonants in the coda of the first syllable. Forty-five percent of the words were syllabified in the same way by at least 80% of the subjects. The small number of items results in a situation in which many of the values of one variable do not co-occur with those of another, making logistic regression analysis impossible. However, inspection of the data suggests that morpheme boundaries were important for compound words. For the 27 compound words, 76% of the syllabifications were at the location of the morpheme boundary. The figure was lower, 51%, for the seven morphologically complex words that were not compounds.

5. GENERAL DISCUSSION AND CONCLUSIONS

The present experiment yields a great deal of information about how people prefer to syllabify English words in a meta-linguistic task. Of course, these data relate only to syllabification in English and not the question of which syllabification strategies may be universal. English syllable boundaries are often ambiguous. Different people may syllabify the same word differently, and pronunciation dictionaries sometimes also disagree. This ambiguity regarding syllable boundary placement may help explain the lack of syllabic effects in some psycholinguistic experiments with English (Cutler, Mehler, Norris, & Seguí, 1986; Bradley, Sánchez-Casas, & García-Albea, 1993; Schiller, 1999; Schiller & Costa, 2006). This contrasts with the situation in other languages. For example, syllable boundaries are less ambiguous in French, and the syllable has been shown to play an important role in language processing in some studies (Mehler, Dommergues, Frauenfelder, & Seguí, 1981; Ferrand, Seguí, & Grainger, 1996), although not in others (Schiller, Costa, & Colomé, 2002).

Previous work on English syllabification, although it provides some insight into the variables that influence performance, has involved relatively small numbers of items. Studies have often considered only one or two variables at a time, and they have included only words containing one or two medial consonants. These are issues when attempting to make broad generalizations about English syllabification. We went beyond previous work by collecting data on 4990 bi-syllabic words containing from one to four medial consonants and by using mixed effects logistic regression analysis, which allows the influence of many different variables to be ascertained. Over 800 subjects participated in our study, many more than in previous studies. We were able to carry out a project of this size by conducting it on the Internet rather than by collecting data from one participant at a time, as most past researchers have done. Although the large size of the study meant that certain issues could not be examined, it allowed us to go beyond previous work that has used smaller number of participants and smaller numbers of words.

Our subjects disagreed on the syllabifications of a number of words, but the variability across subjects was less than might have been anticipated using this sort of experimental method. For 63% of the words, 80% or more of the subjects agreed on the syllabification. One would have expected a higher degree of agreement if all speakers applied a uniform syllabification algorithm. On the other hand, a lower degree of agreement would be expected if English syllabification were as unsystematic as some assert (e.g. Titone & Connine, 1997, p. 251).

The variability that we found was to some extent related to the number of response options that subjects were provided. Words with more medial consonants had more response options than those with fewer medial consonants, and on this basis one would expect a smaller degree of agreement among the subjects on words with more medial consonants. This was borne out to a degree: 80% or more of the subjects agreed on the syllabification of 45% of the items with four medial consonants, 69% of the items with three consonants, and 80% of the words with two consonants. What is surprising is that this number drops to 50% for words with a single medial consonant in spite of the fact that only two syllabification responses are possible (.C and C.).

One thing that may explain the differences in agreement rates is ambisyllabicity. The clearest cases of ambisyllabicity involve a single medial consonant that belongs to the previous and following syllable (Anderson & Jones, 1974; Giegerich, 1992; Hockett, 1955; Jones, 1976; Kahn, 1980). For this reason, it is possible that a good number of the test items with a single consonant are ambisyllabic. No response option encoded ambisyllabicity, which means that subjects were forced to choose either .C or C. Therefore, the unexpectedly low degree of agreement among the subjects on these words could be interpreted as evidence for ambisyllabic medial consonants.

Our findings suggest that language-specific knowledge of words plays the most crucial role in the syllabification of English. English speakers, or at least literate adults as in our study, make syllables as word-like as possible. This is not surprising because words, both written and spoken, are more salient to literate adults than isolated syllables. For instance, words appear between spaces in writing, not syllables. Indeed, some linguists (e.g. Bybee, 1985; 2001) suggest that the word is the most natural unit for understanding phonology and morphology. This may be especially true for English where affixes are generally added to a word which serves as the base rather than to a root that is a bound morpheme. What is more, it is often easy to identify base words within complex words in English. Our results suggest that people tend to syllabify by looking for individual words (or units that follow English phonotactics and orthotactics) within a word. This explains why we found such strong effects for morphology, especially in compound words, where both syllables can form words on their own. Aligning syllable and morpheme boundaries in non-compound words also identifies the part of a complex word that can often stand on its own as a separate word.

Our data support the claim of several linguists that syllabification respects morpheme boundaries (Borowsky, 1986; Derwing, 1992; Wells, 1990), while speaking against the claim of Pulgram (1970) that syllables must be defined without reference to morphology. However, the null results of Smith & Pitt (1999) with respect to morphology, which were mentioned previously, contrast with our findings and deserve to be explored in more detail in future research.

Not all syllable divisions, of course, yield words. However, our data suggest that people use their familiarity with words in the syllabification of even monomorphemic items by breaking them into syllables which – where even though they may not be real words – are similar to words. For example, we found that people tend to place syllable boundaries so that the resulting consonants are legal word-initial and word-final consonants and consonant clusters. Similarly, people tend to avoid stranding certain vowels in syllable-final position because they are not legal word-finally. This finding has been reported previously (e.g. Meyers, 1987; Treiman & Zukowski, 1990). Our study contradicts the idea that the phonotactics of the word do not influence syllabification (e.g. Harris, 1994; Kaye, Lowenstamm, & Vergnaud, 1990). Instead, the results confirm other evidence that word phonotactics are important in syllabification (e.g. Anderson & Jones, 1974; Martens et. al., 2002; Redford & Randall, 2005; Smith & Pitt, 1999; Steriade, 1999; Wells, 1990).

The tendency to produce word-like syllables involves not only phonological legality but also orthographic legality. Spelling has largely been ignored in most previous studies of oral syllabification and of spoken language processing in general. Nevertheless, our data show that spelling affects syllabifications made by literate adults even when their attention was focused on phonetic representations. Syllabification tended to be carried out so that graphemes that are unusual or unattested word-initially or word-finally were avoided syllable-initially and finally as well. The influence of geminate spellings has been noted in some previous experiments (Treiman, Bowey, & Bourassa, 2002; Treiman & Danis, 1988), but the influence of other kinds of spelling on syllabification is a novel finding of our study. Our results are consistent with the emerging view that literate people's knowledge of spelling affects their performance in a wide range of tasks that nominally involve only spoken forms (e.g. Rastle, McCormick, Bayless & Colin, 2011; Taft, Castles, Davis, Lazendic, & Nguyen-Hoan, 2008; Ziegler & Ferrand, 1998).

In addition to the effects of word-likeness, two other influences were manifest. The first involves sonority. The more sonorous the medial consonant or first consonant in a CC cluster, the more likely it is to be placed into the coda of the first syllable. This trend has been observed by a number of previous researchers (e.g. Treiman & Danis, 1988; Derwing & Neary, 1991; Derwing, 1992; Treiman, Staub, & Lavery, 1994; Treiman et al., 1992; Zamuner & Ohala, 1999). A novel finding is that the tense vowels in the second syllable tend to draw consonants into their onsets, although the effect is small. We examined a number of possibilities that could be related to this variable (second vowel length by stress, relationship between the vowel quality in the first and in the second syllable, tendency to syllabify so that syllables are of more equal length), but none were significant.

Differences in how people syllabify the same word may reflect how they weight various factors. Some of the differences may also reflect characteristics of an individual's mental lexicon, such that some people may perceive a particular word to be monomorphemic and others may view it as consisting of several parts. Variation in the size and content of an individual's mental lexicon may also affect which consonant clusters they accept at the margins of words and syllables. Our study focused on explaining variation across words rather than variation across people. Not all of the test words were syllabified by every subject, and further research is needed to examine across-subject variability.

The present findings have implications for a number of areas. The data for individual words that we have made available [linguistics.byu.edu/faculty/eddingtond/BYU Syllabification Survey.xls] should be useful in future linguistic studies of English syllabification and in future editions of pronunciation dictionaries. Our findings should also be of interest to psycholinguists who examine the role of syllables in the production and perception of spoken and written language. Our study also demonstrates the value of linguistic mega-studies and the feasibility of carrying them out online, allowing researchers to generate a great deal of data in a short period of time and at a low cost.

ACKNOWLEDGEMENTS

We thank Lindsey Clasen, David Balota, Brett Kessler, Jennifer Russ, and Martha Storandt for their assistance and comments. We are particularly grateful to Mark Davies for his help with the online survey.

REFERENCES

Anderson, J. M., & Jones, C. (1974). Three theses concerning phonological representations. *Journal of Linguistics, 10*, 1–26.

Borowsky, T. (1986). *Topics in the Lexical Phonology of English*. PhD Dissertation: University of Massachusetts, Amherst.

Bradley, D. C., Sánchez-Casas, R. M., & García-Albea, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes, 8*, 197–233.

Bybee, J. (1985). *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: John Benjamins.

Bybee, J. (2001). *Phonology and Language Use*. Cambridge: Cambridge University Press.

Clements, G. N. (1990). The sonority cycle and syllable organization. In W. U. Dressler, H. C. Luschützky, O. E. Pfeiffer, & J. R. Rennison(Eds), *Phonologica 1988* (pp. 63–76). Cambridge: Cambridge University Press.

Cutler, A., Mehler, J., Norris, D., & Seguí, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385–400.

Derwing, B. L. (1992). A 'pause-break' task for eliciting syllable boundary judgments from literate and illiterate speakers: Preliminary results for five diverse languages. *Language and Speech, 35*, 219–235.

Derwing, B. L., & Neary, T. M. (1991). The "vowel-stickiness" phenomenon: Three experimental sources of evidence. *Proceedings of the 12th International Congress of Phonetic Sciences*, 3, 210–213.

Eddington, D., Treiman, R., & Elzinga, D. (2013). Syllabification of American English: Evidence from a Large-scale Experiment. Part I. *Journal of Quantitative Linguistics, 20* (1), 45–67.

Ferrand, L., Seguí, J., & Grainger, J. (1996). Masked priming of word and picture naming: The role of syllabic units. *Journal of Memory and Language, 35*, 708–723.
Giegerich, H. (1992). *English Phonology.* Cambridge: Cambridge University Press.
Harris, J. (1994). *English Sound Structure*. Oxford: Blackwell Publishers.
Hockett, C. (1955). *A Manual of Phonology.* Bloomington IN: Indiana University Publications in Anthropology and Linguistics.
Jones, C. (1976). Some constraints on medial consonant clusters. *Language, 52*, 121–30.
Kahn, D. (1980). *Syllable-based Generalizations in English Phonology.* New York: Garland.
Kaye, J., Lowenstamm, J., & Vergnaud, J. -R. (1990). Constituent structure and government in phonology. *Phonology Yearbook, 7*, 193–231.
Martens, E., Daelemans, W., Gillis, S., & Taelman, H. (2002). Where do syllables come from? In W. Gray& C. Schunn (Eds), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 657–664). Fairfax, Virginia: George Mason University.
Mehler, J., Dommergues, J. Y., Frauenfelder, U., & Seguí, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298–305.
Myers, S. (1987). Vowel shortening in English. *Natural Language and Linguistic Theory, 5*, 485–518.
Pulgram, E. (1970). *Syllable, Word, Nexus, Cursus*. The Hague: Mouton.
Rastle, K., McCormick, S. F., Bayliss, L., & Davis, C. J. (2011). Orthography influences the perception and production of speech. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1588–1594.
Redford, M. A., & Randall, P. (2005). The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics, 33*, 27–46.
Schiller, N. O. (1999). Masked syllable priming of English nouns. *Brain and Language, 68*, 300–305.
Schiller, N. O., & Costa, A. (2006). Activation of segments, not syllables, during phonological encoding in speech production. *The Mental Lexicon, 1*, 231–250.
Schiller, N. O., Costa, A., & Colomé, A. (2002). Phonological encoding of single words: In search of the lost syllable. In C. Gussenhoven& N. Warner(Eds), *Laboratory Phonology 7* (pp. 35–59). Berlin: Mouton de Gruyter.
Selkirk, E. O. (1982). The syllable. In H. van der Hulst& N. Smith(Eds), *The Structure of Phonological Representations II* (pp. 337–383). Dordrecht: Foris.
Smith, K. L., & Pitt, M. A. (1999). Phonological and morphological influences in the syllabification of spoken words. *Journal of Memory and Language, 41*, 199–222.
Steriade, D. (1999). Alternatives to syllable-based accounts of consonantal phonotactics. In O. Fujimura, B. Joseph, & B. Palek(Eds), *Proceedings of the 1998 Linguistics and Phonetics Conference* (pp. 205–242). Prague: The Karolinum Press.
Taft, M., Castles, A., Davis, C., Lazendic, G., & Nguyen-Hoan, M. (2008). Automatic activation of orthography in spoken word recognition: Pseudohomograph priming. *Journal of Memory and Language, 58*, 366–379.
Titone, D., & Connine, C. M. (1997). Syllabification strategies in spoken word processing: Evidence from a new methodology. *Psychological Research, 60*, 251–263.
Treiman, R., & Zukowski, A. (1990). Toward an understanding of English syllabification. *Journal of Memory and Language, 29*, 66–85.
Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language, 27*, 87–104.
Treiman, R., Gross, J., & Cwikiel-Glavin, A. (1992). The syllabification of /s/ clusters in English. *Journal of Phonetics, 20*, 383–402.

Treiman, R., Bowey, J. A., & Bourassa, D. (2002). Segmentation of spoken words into syllables by English-speaking children. *Journal of Experimental Child Psychology, 83*, 213–238.

Treiman, R., Staub, K., & Lavery, P. (1994). Syllabification of bisyllabic nonwords: Evidence from short-term memory errors. *Language and Speech, 37*, 45–60.

Wells, J. C. (1982). *Accents of English*, Vol. 1. Cambridge: Cambridge University Press.

Wright, R. (2004). A review of perceptual cues and cue robustness. In B. Hayes, R. Kirchner, & D. Steriade(Eds), *Phonetically Based Phonology* (pp. 34–57). Cambridge: Cambridge University Press.

Zamuner, T. S., & Ohala, D. K. (1999). Preliterate children's syllabification of intervocalic consonants. In A. Greenhill, H. Littlefield, & C. Tano(Eds), *Proceedings of the 23rd Annual Boston Conference on Language Development* (pp. 753–763). Somerville MA: Cascadilla Press.

Ziegler, J. C., & Ferrand, L. (1998). Orthography shapes the perception of speech: The consistency effect in auditory word recognition. *Psychonomic Bulletin and Review, 5*, 683–689.