Journal of Research Design and Statistics in Linguistics and Communication Science

JRDS (PRINT) ISSN 2052-417X
JRDS (ONLINE) ISSN 2052-4188

Article

The Influence of Experimental Method on English Syllabification

David Eddington and Ross J. Cairns

Abstract

A number of experimental methods have been used to elicit metalinguistic judgments about syllable division, a good deal of which deals with the syllabification of English words. However the syllabification literature is largely silent on the issue of intratask reliability, that is, whether the tasks all yield the same kinds of intuitions from speakers. Côté and Kharlamov (2011) gathered data from Russian speakers who syllabified nonce words in four different experimental conditions. When the results were compared they observed widely different results in many instances. This suggests that syllabification preferences are highly influenced by the particular task used to elicit them, which in turn casts doubt on the intratask reliability of syllabification studies.

In order to test the reliability of different experimental methods in English, syllable divisions of 120 English words were elicited with eight different experimental tasks. In a mixed-effects logistic regression, no main effect of experimental method was found, although the method showed some interaction with stress and the legality of the consonant cluster word-initially and word-finally. Reasons why these results differ from those of Côté and Kharlamov are discussed, some of which are due to methodological flaws in their analysis.

KEYWORDS: ENGLISH; EXPERIMENT; INTRATASK RELIABILITY; SYLLABIFICATION

Affiliation

Brigham Young University, Provo, UT, USA.
email: eddington@byu.edu

JRDS VOL 2.1 2015 37–52

doi : 10.1558/jrds/28518

Introduction

Since the last part of the twentieth century, linguists have shown a renewed interest in the syllable, especially in its role in phonological processes. One topic area that psycholinguists have taken up is investigating how speakers divide words into syllables, along with the factors that influence syllabification. To this end, a variety of different experimental techniques have been devised. One unspoken assumption that exists in the literature on syllabification is that all of the methods of elicitation are equally valid; they all tap into the same speaker intuitions about syllabification and should yield the similar results.

Côté and Kharlamov (2011) have recently challenged the assumption that different tasks produce similar results. They compared the results of four syllabification studies of Russian that used the same test items, but that employed different experimental tasks. They observed a lack of significant correlations between some of the outcomes of their four experiments, which is a serious cause for concern. If different tasks produce different results, then it is not only not valid to compare the results of syllabification studies that use different tasks, but it raises the question of which task, if any, reflects actual syllabification strategies, or if different tasks tap into different strategies.

A good deal of what we know about how English speakers syllabify words is based on experimental research carried out in the past 30 years. The validity of these studies could be put in serious jeopardy if it is found that different tasks produce radically different results. The present paper addresses this issue directly by comparing the syllabification of extant English words using eight different experimental tasks similar to those used by Côté and Kharlamov. The results of the present study are contrasted and compared with those of Côté and Kharlamov, and a number of difficulties with their study are highlighted. The problematic aspect of their study are rectified in the present study by analyzing the data using a mixed-effects analysis.

Review of experimental methods

Quite a few different experimental tasks have been use to determine syllabification. Some of these are designed to elicit syllabifications subconsciously (e.g. Prinzmetal, Treiman and Rho 1986; Rapp 1992; Smith and Pitt 1999; Treiman, Straub, and Lavery 1994). However, the focus of the present paper is on methods used to test metalinguistic judgments of syllable division. The majority of work in this area entails variations of three methods. Generally, the instructions for these tasks do not mention syllables, but merely dividing a word or identifying its parts.

In division experiments, subjects are presented an entire word and asked to divide it. With written stimuli this is done by inserting a slash or space in the word or choosing from among responses with different slash or space divisions (Eddington, Treiman, and Elzinga 2013; McCrary 2004; Redford and Randall 2005; Treiman and Danis 1988; Treiman, Gross, and Cwikiel-Glavin 1992; Treiman and Zukowski 1990). For example, the responses for *gasket* could be *ga / sket, gas / ket,* or *gask / et.* Other versions of this task involve producing a word with a pause between halves or choosing from among auditory options containing pauses in different places (Barry, Klein, and Köser 1999; Côté and Kharlamov 2011; Derwing 1992; Fallows 1981; Gillis and Sandra 1998; Goslin and Floccia 2007; Ishikawa 2002; McCrary 2004; Schiller, Meyer, and Levelt 1997; Zamuner and Ohala 1999). Another variant has subjects insert other material between the parts of the word (Content, Kearns, and Frauenfelder 2001.)

In contrast to division experiments where subjects deal with an entire word, identification studies ask subjects to say (or choose from among written options) what the first or last part of a word is (Cebrian 2002; Content, Kearns, and Frauenfelder 2001; Côté and Kharlamov 2011; Fallows 1981; Goslin 2002; Goslin and Frauenfelder 2000; Treiman, Bowey, and Bourassa 2002). When asked what the last part of *gasket* is, subjects could respond *sket, ket,* or *et.* Equivalently, reduplicating the first or last part of a word is another way of inducing subjects to identify their intuitions about word division (Berg 2001; Berg and Niemi 2000; Bertinetto, Carboara, Gaeta, and Agonigi 1994; Bertinetto, Scheuer, Dziubalsca-Kolaczyk, and Agonigi 2007; Fallows 1981; Treiman, Gross, and Cwikiel-Glavin 1992; Treiman and Zukowski 1990). The first part of *gasket* would be *gas* if the reduplicated form were *gasgasket,* while it would be *gask* if *gaskgasket* were produced.

In reversal tasks, word divisions are tested by prompting subjects to switch the first and last parts of the word (Barry, Klein, and Köser 1999; Berg 2001; Berg and Niemi 2000; Bertinetto, Scheuer, Dziubalsca-Kolaczyk, and Agonigi 2007; Cebrian 2002; Content, Kearns, and Frauenfelder 2001; Côté and Kharlamov 2011; Schiller, Meyer, and Levelt 1997; Treiman and Danis 1988; Treiman, Gross, and Cwikiel-Glavin 1992). Accordingly, *gasket* could be rendered as *etgask, ketgas,* or *sketga.* Ambisyllabicity is evidence by responses such as *ketgask.*

Experimental method

The purpose of the present experiment is to elicit syllabifications from native English speakers using a variety of tasks. Four written response and four oral response tasks were used. The four written tasks were: reversal, division, first part, and second part. Using *agree* as an example, in the syllable reversal task subjects were asked: 'How would you reverse the parts of *agree*?' They then chose either *gree-a, ree-ag,* or *ee-agr.* For the syllable division task they were asked: 'How would you divide *agree*?' They chose between *a / gree, ag / ree,* and *agr / ee.* For the first part task they were asked: 'What is the first part of *agree*?' and had *a, ag,* and *agr* as options. The second part task asked: 'What is the second part of *agree*?' and response options were *ee, ree,* and *gree.* The oral response tasks included the same test items and the same four tasks, but differed in that the subjects gave their responses out loud rather than choosing from among prefabricated written responses.

Test items

Following Côté and Kharlamov (2011), all 120 test items were bisyllabic words containing a variety of two-consonant clusters, along with a variety of vowel types and stress patterns. However, in contrast to Côté and Kharlamov, who used nonce words, all of the test items in the present study were extant English words (see Appendix). This was done to more closely mirror previous studies of English syllabification that were carried out with actual words. In order to control for the influence of morphology, all of the test words were monomorphemic. Four different tests were made from the 120 test items. Each test contained all 120 items, but differed as to which items were presented in which experimental task. In each test, 30 test words were presented in each of the four tasks.

Procedure

The written response tests were carried out online using Qualtrics[1] questionnaire software. Subjects logged onto the site and first responded to an informed consent form and provided biographical information. They were randomly assigned one of the four tests and then were ask to respond to all 120 items. Each subject saw 30 of the items in each of the four experimental tasks. The order of presentation of the four tasks, and the order of the items within each task was randomized. The test was designed so that each subject responded to each test item only once. No time limit was imposed, but the average time to complete the survey was about 15 minutes. In the oral response paradigm, the same four tests were used. The only difference was that the subjects did not choose from among written responses, but gave their responses out loud. The responses were recorded by the experimenter.

Subjects

Of the 148 native American English speakers who took the written survey, four completed it in under eight minutes which was considered inordinately fast, so their responses were eliminated. The remaining 144 subjects were comprised of 104 females and 40 males, the majority of whom ($n = 126$) were between 18 and 29 years of age. Only 18 were 30 or older. The respondents were fairly well educated in that 114 had some college experience, and 27 at least a four-year degree. The remaining three had no college experience. Residents of 29 different states of the US were represented. Seven of the subjects did not complete the entire survey, but the answers they did provide were included in the analyses.

A total of 44 subjects took the oral response survey: 28 males and 16 females, of which 33 were between the ages of 18 and 29, nine were in their 30s, and two were 50 or older. Two subjects indicated that they had a high school degree, 22 had some college education, 11 were college graduates, and nine had graduate degrees. Residents of 14 different US states were represented.

Results and discussion

As Table 1 indicates, splitting the medial cluster between the two parts of the test words was the most common division across all tasks. Parsing the words so that both medial consonants are placed with the second part of the word was the second most common strategy, followed by putting both consonants in the coda of the first syllable. This same hierarchy was found by Côté and Kharlamov (2011). The major difference is that a much higher percent of their nonce word test items (about 95%) demonstrated a split between medial consonants, while keeping the cluster together in either part of the word was uncommon.

Table 1: Percentage of responses in each experimental task.[2]

	.CC	C.C	CC.
Division-Written	17	76	7
Reversal-Written	16	76	8
First Part-Written	13	77	10
Second Part-Written	16	74	10
Division-Oral	18	81	0
Reversal-Oral	17	80	1
First Part-Oral	16	79	5
Second Part-Oral	18	79	3

Table 2 gives some percentages of each syllabification for three words. Most of the responses for *thermal* divided the medial cluster (i.e. *ther.mal*), which was followed by placing both consonants in the first syllable (e.g. *therm.al*). None placed both in the onset of the second syllable (i.e. *the.rmal*). In like manner, the most common syllabification strategy for *gasket* was *gas.ket* while *gask.et* and *ga.sket* were about equally likely. On the other hand, *matrix* elicited mostly syllabifications that placed both consonants in the onset of the second syllable (i.e. *ma.trix*) with only occasional divisions such as *mat.rix* and *matr.ix*.

However, the most crucial part of the present study is not the exact way subjects divided words, but whether the experimental tasks would elicit similar or differing divisions. Côté and Kharlamov compared the results of their experiments by performing a Pearson correlation on the average number of consonants placed in the first syllable for each combination of experimental task. We followed suit and our results appear in Table 3 where all the correlations are significant at $p < 0.0005$. The extremely high correlations between all eight experimental tasks appears to be quite telling; in spite of the differences in task, subjects divided the test words in much the same way.

Table 2: Examples of mean responses averaged by test item and task.

	thermal			*gasket*			*matrix*		
	CC.	C.C	.CC	CC.	C.C	.CC	CC.	C.C	.CC
Division-Written	24	76	0	5	70	25	4	4	92
Reversal-Written	25	75	0	20	60	20	11	16	73
First Part-Written	33	67	0	19	74	7	0	25	75
Second Part-Written	23	77	0	21	60	19	0	11	89
Division-Oral	0	100	0	0	73	27	0	9	91
Reversal-Oral	9	91	0	0	100	0	0	9	91
First Part-Oral	10	90	0	22	78	0	0	0	100
Second Part-Oral	0	100	0	0	91	9	0	10	90

Table 3: Pearson correlations between the four experimental tasks when compared by word mean.

	Reversal -Written	**First Part -Written**	**Second Part -Written**	**Division -Oral**	**Reverse -Oral**	**First Part-Oral**	**Second Part-Oral**
Division-Written	0.95	0.97	0.94	0.95	0.94	0.87	0.95
Reversal-Written		0.95	0.97	0.93	0.95	0.89	0.92
First Part-Written			0.92	0.93	0.89	0.88	0.92
Second Part-Written				0.92	0.93	0.88	0.91
Division-Oral					0.93	0.87	0.95
Reversal-Oral						0.87	0.94
First Part-Oral							0.88

It would be tempting to finish the study at this point and conclude that the high correlations in the present study indicate that the particular experimental task exerted little influence on the subject's division preferences. This would contrast with Côté and Kharlamov's study in which a number of the correlations they obtained between experimental tasks were much lower, and could be interpreted to mean that the tasks produced different outcomes. However, coming to such conclusions based on the results of a correlation is problematic. In the first place, performing a correlation on these data is statistically unsound because it violates the assumption the data are independent. In a correlation, a subject can only provide one data point on the X axis and one on the Y axis (Grissom and Kim 2012: 109). In contrast, both our subjects and those of Côté and Kharlamov responded to each of the 120 test items, and then the responses for each test item were averaged. In this way, each subject contributed to the average of each test item.[3]

Statistical analysis of the data

Given the repeated measures in the study, an appropriate way to analyze the results is with a mixed-model analysis. Instead of using correlation to measure the degree of similarity among the results from the different experimental tasks, this approach asks whether experimental task significantly influences syllabification preferences. It controls for the existence of the repeated measures by allowing each test subject to have an individual random slope over experimental task. Test items are allowed random intercepts.

The experiments yielded a total of 21,840 responses. The 20 ambisyllabic responses and one uncategorizable response were eliminated prior to statistical analysis. However, after collecting the responses, we observed a number of test items (i.e. *segment, perfect, insult, insert,* and *combat*) that may be stressed on either syllable. This could potentially muddle the influence of stress on the results, so the 911 responses to these words were eliminated as well. We then attempted a number of multinomial logistic regressions, but the models in each case either failed to converge or produced Hessian matrices that were not positive definite. This appears to be due to the fact that there were so few CC. responses (e.g. *therm.al*), which in any event, make up only 7% of the data. For this reason we removed them from the analysis and concentrated on C.C versus .CC divisions in a binomial regression.

Although the focus of the study is on the influence of experimental task on syllabification, a number of other variables are known to influence syllabification and need to be included as control variables (c.f. Eddington *et al.* 2013). These are whether the word is stressed on the first or last syllable, whether the vowel in the first syllable is tense or lax, whether the word-medial cluster is legal word-initially (e.g. [*tr-, fl-*]), word-finally (e.g. [*-rk -ns*]) both initially and finally (e.g. [*st, sk*]) or in neither position (e.g. [*gm, mb*]). We also included all interactions between these variables, but only report the results of the model that includes the significant variables and interactions.

The most parsimonious model correctly predicts 95.3% of the remaining syllabifications which is much higher than the 70.4% by-chance accuracy rate.[4] In Table 4 it is evident was a significant main effect for the control variables that encoded the quality of the first vowel and the legality of the word-medial consonant cluster, while stress was not significant. In like manner, test method was not significant either. However, these main effects are moderated by the fact that they participate in a number of interactions. Stress by cluster legality, stress by test method, and cluster legality by test method are also significant. We will not bother investigating the stress by cluster legality interaction since both of those variables were included as controls and are not of interest to the hypothesis we are investigating. The two interactions that involve test method, on the other hand, are crucial since the influence of test method is crucial to our hypothesis.

Table 4: Result for the main and interaction effects.

Variable	*F*	*df1*[5]	*df2*	*p*
Corrected Model	19.93	43	589	0.0005
Stress	2.54	1	94	0.114
Quality of First Vowel	22.50	1	62	0.0005
Cluster Legality	157.07	3	99	0.0005
Test Method	0.39	7	705	0.907
Stress by Cluster Legality	11.04	3	87	0.0005
Stress by Test Method	2.39	7	19,318	0.020
Cluster Legality by Test Method	3.28	21	19,318	0.0005

The first interaction of interest is test method by consonant cluster legality which is depicted graphically in Figure 1. Consonant clusters that are legal in word-initial position (e.g. [*tr-*, *bl-*] are the most likely to be syllabified as .CC, and placed in the onset of the first syllable. These are followed by clusters such as [*sp*, *sk*] that are attested both word-initially and finally which prefer .CC to a high degree as well. However, the lines representing the estimated means for consonant clusters that are legal word finally and those that are not legal either in word-initial or word-final position appear at the bottom of the graph. Words with those kinds of clusters are least likely to receive a .CC syllabification, which means that they are most likely to have the consonants split between syllables (e.g. *mem.ber*, *ten.der*). It is not surprising that the final and neither legalities are not statistically significant from each other. The lines representing them are entangled with each other, and are impossible to distinguish.

Table 5 shows that with the exception of consonant clusters that are legal word-finally and in neither position, all other paired comparisons between consonant cluster legalities differ significantly from each other. The point of exploring this interaction is that if different experimental tasks yield different results in words with consonant clusters of different legalities, that would suggest that different tasks result in different syllabifications. In this case, however, the final and neither legality types were essentially identical across all of the experimental tasks, while the remainder of the legality types demonstrated significant differences across all of the experimental tasks. In other words, consonant cluster legality interacted with task, but it did so in the exact same way in each task. Therefore, this interaction supports the hypothesis that different tasks do not produce different outcomes.

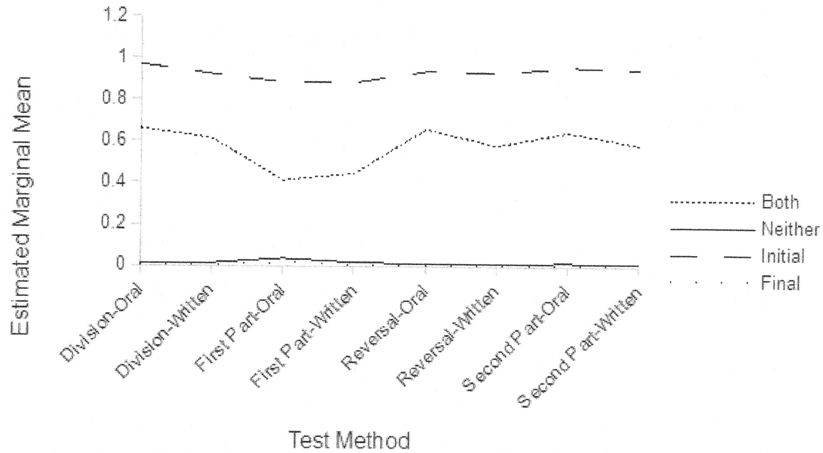

Figure 1: Estimated marginal means for consonant cluster legality by test method.

The second significant interaction of interest occurs between experimental task and word stress. As Table 6 indicates, stress does not interact significantly with any of the experimental tasks except one, the written first part task. Evidently in this task words with final stress (e.g. *estate, afraid*) were syllabified .CC more often that words with initial stress such as *cluster* and *aspen*. Although the difference is significant, the interaction is not readily apparent in the graph of the estimated marginal means (Figure 2).

Comparison with Côté and Kharlamov

In the present study it is safe to conclude that different experimental tasks produced largely the same syllabifications. The one case in which stress affects it is a far cry from the widely differing results reported by Côté and Kharlamov. The question that naturally arises is why our results differ so much from theirs. Only four of their ten comparisons were significantly correlated. It must be emphasized that while the two studies have the goal of investigat-

Table 5: Interaction between test task and the legality of the consonant cluster in various positions.

Test Method	Cluster Legality	Contrast Estimate	Adj. *p*	Test Method	Cluster Legality	Contrast Estimate	Adj. *p*
Division-Oral	Both vs. Final	0.659	0.0005	**Reversal-Oral**	Both vs. Final	0.656	0.0005
	Both vs. Initial	−0.308	0.002		Both vs. Initial	−0.279	0.004
	Both vs. Neither	0.665	0.0005		Both vs. Neither	0.648	0.0005
	Final vs. Initial	−0.967	0.0005		Final vs. Initial	−0.938	0.0005
	Final vs. Neither	−0.004	0.392		Final vs. Neither	−0.008	0.191
	Neither vs. Initial	−0.963	0.0005		Neither vs. Initial	−0.927	0.0005
Division-Written	Both vs. Final	0.611	0.0005	**Reversal-Written**	Both vs. Final	0.576	0.0005
	Both vs. Initial	−0.310	0.0005		Both vs. Initial	−0.349	0.0005
	Both vs. Neither	0.601	0.0005		Both vs. Neither	0.569	0.0005
	Final vs. Initial	−0.921	0.0005		Final vs. Initial	−0.925	0.0005
	Final vs. Neither	−0.011	0.057		Final vs. Neither	−0.008	0.138
	Neither vs. Initial	−0.910	0.0005		Neither vs. Initial	−0.917	0.0005
First Part-Oral	Both vs. Final	0.396	0.0005	**Second Part-Oral**	Both vs. Final	0.640	0.0005
	Both vs. Initial	−0.472	0.0005		Both vs. Initial	−0.314	0.002
	Both vs. Neither	0.377	0.0005		Both vs. Neither	0.631	0.0005
	Final vs. Initial	−0.868	0.0005		Final vs. Initial	−0.954	0.0005
	Final vs. Neither	−0.018	0.274		Final vs. Neither	−0.010	0.238
	Neither vs. Initial	−0.849	0.0005		Neither vs. Initial	−0.944	0.0005
First Part-Written	Both vs. Final	0.440	0.0005	**Second Part-Written**	Both vs. Final	0.580	0.0005
	Both vs. Initial	−0.438	0.0005		Both vs. Initial	−0.365	0.0005
	Both vs. Neither	0.429	0.0005		Both vs. Neither	0.575	0.0005
	Final vs. Initial	−0.878	0.0005		Final vs. Initial	−0.945	0.0005
	Final vs. Neither	−0.011	0.120		Final vs. Neither	−0.005	0.160
	Neither vs. Initial	−0.867	0.0005		Neither vs. Initial	−0.939	0.0005

ing the influence of experimental paradigm, they vary in many significant regards. First, the present study contrasted eight tasks that used real English words. In contrast, Côté and Kharlamov's study contained four tasks in which Russian nonce words served as test items. Second, the Russian study averaged responses to all test items that contained the same medial consonant clusters. In our study, we kept words with the same medial cluster distinct.

Table 6: Interaction between test method and word stress.

Test Method	Stress	Contrast Estimate	Adj. *p*
Division-Oral	Final vs. Initial	0.155	0.079
Division-Written	Final vs. Initial	0.055	0.179
First Part-Oral	Final vs. Initial	0.009	0.880
First Part-Written	Final vs. Initial	0.093	0.043
Reversal-Oral	Final vs. Initial	−0.072	0.182
Reversal-Written	Final vs. Initial	0.090	0.060
Second Part-Oral	Final vs. Initial	0.053	0.450
Second Part-Written	Final vs. Initial	0.060	0.177

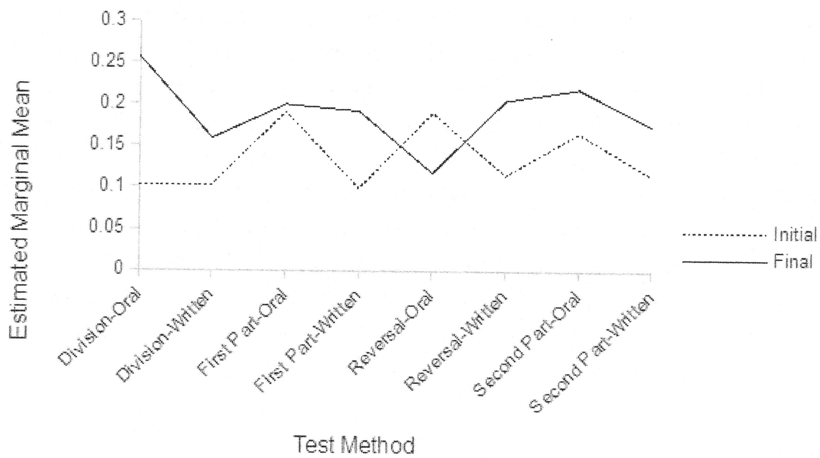

Figure 2: Estimated marginal means for stress by test method.

Third, Côté and Kharlamov performed a correlation on data that were averaged across words containing the same consonant cluster, and across different subjects' responses to those words. The fact that each subject provided multiple responses that were all included in the correlation violates the assumption of independence that tests like correlation, *t*-tests, and regression require.

To account for the repeated measures of the subjects we used a mixed-effects model with a random effect of subject over experimental method, and also allowed each test word it's own random intercept. All of these differences in method could be responsible for the different outcomes in the two studies.

Conclusions

Côté and Kharlamov raise serious concerns about intratask reliability in syllabification studies since they report only sporadic correlations between the different syllabification tasks they utilized. However, their experiment appears to be the only empirical study of Russian syllabification. In contrast, English syllabification has been examined experimentally for several decades using a variety of tasks, which has given us some sense of what factors influence seem to influence metalinguistic judgments of English syllabification. Nevertheless, the question of whether results with different tasks are comparable has not been investigated. If the same lack of intratask reliability that Côté and Kharlamov observed were found in English as well, that would make comparing the findings of the numerous studies of English syllabification problematic and the significance of individual studies difficult to assess.

The present study addressed this issue by examining the results of four written and four oral response tasks. With only one small caveat, the experimental tasks yielded highly similar results. This suggests that the different experimental methods that have been used over the years appear to tap into the same kinds of speaker intuitions. However, a number of additional questions remain to be answered. The present study is limited to words containing two medial consonants. Whether similarly high correlations would be obtained with words containing one, three, or four consonants clusters has yet to be determined. In like manner, metalinguistic judgments need to be compared with those obtained in a more indirect manner, and the particular languages that are tested needs to be expanded. In short, there is ample room for further research into syllabification as well as the methods used to investigate it.

Notes

1. www.qualtrics.com
2. In some cases, the percentages don't add up to 100. This is due to rounding and in the oral response tasks it is also due to the existence of 20 ambisyllabic responses, and one uncategorizable response.
3. Côté and Kharlamov performed *t*-tests and regressions on their subject-averaged data which violates the independence assumption of those tests as well.
4. 3,475 .CC responses / 19,362 total responses = 0.179. 15,887 C.C responses / 19,362 total responses = 0.820. $0.179^2 + 0.820^2 = 0.704$.
5. The Satterthwaite approximation was applied to the degrees of freedom.

About the authors

David Eddington is presently a professor in the Department of Linguistics and English Language at Brigham Young University. He specializes in data-oriented approaches to phonology, morphology, and the Spanish language.

Ross Cairns is an International Account Manager for Spark Innovation. He has an MA in Hispanic Studies from The University of Glasgow and recently graduated from Brigham Young University with an MA in Spanish Linguistics.

References

Barry, W., Klein, C. and Köser, S. (1999). Speech production evidence for ambisyllabicity in German. *Phonus* 4: 87–102.

Berg, T. (2001). An experimental study of syllabification in Icelandic. *Nordic Journal of Linguistics* 24 (1): 71–106. http://dx.doi.org/10.1080/03325860120533

Berg, T. and Niemi, J. (2000). Syllabification in Finnish and German: Onset filling vs. Onset maximazation. *Journal of Phonetics* 28 (2):187–216. http://dx.doi.org/10.1006/jpho.2000.0112

Bertinetto, P. M., Scheuer, S., Dziubalska-Kolaczyk, K. and Agonigi, M. (2007). Intersegmental cohesion and syllable division in Polish. *Proceedings of the 16th International Congress of Phonetic Sciences, 1953-1956.* University of Saarbrücken.

Bertinetto P. M., Caboara, M., Gaeta, L. and Agonigi, M. (1994). Syllabic division and intersegmental cohesion in Italian. In W. U. Dressler, M. Prinzhorn and J. R. Rennison (Eds), *Phonologica 1992: Proceedings of the 7th International Phonology Meeting,* 19–33. Torino: Rosenberg and Sellier.

Cebrian, J. (2002). *Phonetic similarity, syllabification and phonotactic constraints in the acquisition of a second language contrast.* Dissertation, University of Toronto.

Content, A., Kearns, R. K. and Frauenfelder, U. (2001). Boundaries versus onsets in syllabic segmentation. *Journal of Memory and Language* 45 (2): 177–199. http://dx.doi.org/10.1006/jmla.2000.2775

Côté, M.-H. and Kharlamov, V. (2011). The impact of experimental tasks on syllabification judgments: A case study of Russian. In C. E. Cairns and E. Raimy (Eds) *Handbook of the Syllable*, 273–294. Leiden and Boston, MA: Brill.

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 (1–2): 219–235.

Eddington, D., Treiman, R. and Elzinga, D. (2013). Syllabification of American English: Evidence from a Large-scale Experiment. Part II. *Journal of Quantitative Linguistics* 20 (2): 45–67. http://dx.doi.org/10.1080/09296174.2012.754601

Fallows, D. (1981). Experimental evidence for English syllabification and syllable structure. *Journal of Linguistics* 17 (2): 309–317. http://dx.doi.org/10.1017/S0022226700007027

Gillis, S. and Sandra, D. (1998). Children's and Adults' syllabification: The influence of spelling. In A. Aksu-Koç, E. Erguvanli-Taylan, A. Sumru Özsoy and A. Kuntay (Eds) *Perspectives on Language Acquisition*, 336–354. Istanbul: Bogazici University Press.

Goslin, J. (2002) *A comparison of theoretical and human syllabification*. Dissertation, University of Sheffield.

Goslin, J. and Floccia, C. (2007). Comparing French syllabification in preliterate children and adults. *Applied Psycholinguistics* 28 (2): 341–367. http://dx.doi.org/10.1017/S0142716407070178

Grissom, R. J. and Kim, J. J. (2012). *Effect Sizes for Research: Univariate and Multivariate Applications*, 2nd edition. New York: Routledge.

Ishikawa, K. (2002). Syllabification of intervocalic consonants by English and Japanese speakers. *Language and Speech* 45 (4): 355–385. http://dx.doi.org/10.1177/00238309020450040301

McCrary, K. M. (2004). *Reassessing the role of the syllable in Italian phonology: An experimental study of consonant cluster syllabification, definite article allomorphy and segment duration*. PhD Dissertation, UCLA.

Prinzmetal, W., Treiman, R. and Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language* 25 (4): 461–475. http://dx.doi.org/10.1016/0749-596X(86)90038-0

Rapp, B. C. (1992). The nature of sublexical orthographic organization: The bigram trough hypothesis examined. *Journal of Memory and Language* 31 (1): 33–53. http://dx.doi.org/10.1016/0749-596X(92)90004-H

Redford, M. A. and Randall, P. (2005). The role of juncture cues and phonological knowledge in English syllabification judgments. *Journal of Phonetics* 33 (1): 27–46. http://dx.doi.org/10.1016/j.wocn.2004.05.003

Schiller, N. O., Meyer, A. S. and Levelt, W. J. M. (1997). The syllabic structure of spoken words: Evidence from syllabification of intervocalic consonants. *Language and Speech* 40 (2): 103–140.

Smith, K. L. and Pitt, M. A. (1999). Phonological and morphological influences in the syllabification of spoken words. *Journal of Memory and Language* 41 (2): 199–222. http://dx.doi.org/10.1006/jmla.1999.2646

Treiman, R. and Zukowski, A. (1990). Toward and understanding of English syllabification. *Journal of Memory and Language* 29 (1): 66–85. http://dx.doi.org/10.1016/0749-596X(90)90010-W

Treiman, R. and Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language* 27 (1): 87–104. http://dx.doi.org/10.1016/0749-596X(88)90050-2

Treiman, R., Gross, J. and Cwikiel-Glavin, A. (1992). The syllabification of /s/ clusters in English. *Journal of Phonetics* 20 (3): 383–402.

Treiman, R., Bowey, J. A. and Bourassa, D. (2002). Segmentation of spoken words into syllables by English-speaking children. *Journal of Experimental Child Psychology* 83 (3): 213–238. http://dx.doi.org/10.1016/S0022-0965(02)00134-0

Treiman, R., Staub, K. and Lavery, P. (1994). Syllabification of bisyllabic nonwords: Evidence from short-term memory. *Language and Speech* 37 (1): 45–60.

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