The Effect of Sonority on Word Segmentation: Evidence for the Use of a Phonological Universal

Marc Ettlinger,[a] Amy S. Finn,[b] Carla L. Hudson Kam[c]

[a]*Research Service, Veterans Affairs, Northern California Health Care System*
[b]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology*
[c]*Department of Linguistics, University of British Columbia*

Abstract

It has been well documented how language-specific cues may be used for word segmentation. Here, we investigate what role a language-independent phonological universal, the sonority sequencing principle (SSP), may also play. Participants were presented with an unsegmented speech stream with non-English word onsets that juxtaposed adherence to the SSP with transitional probabilities. Participants favored using the SSP in assessing word-hood, suggesting that the SSP represents a potentially powerful cue for word segmentation. To ensure the SSP influenced the segmentation process (i.e., during learning), we presented two additional groups of participants with either (a) no exposure to the stimuli prior to testing or (b) the same stimuli with pauses marking word breaks. The SSP did not influence test performance in either case, suggesting that the SSP is important for word segmentation during the learning process itself. Moreover, the fact that SSP-independent segmentation of the stimulus occurred (in the latter control condition) suggests that universals are best understood as biases rather than immutable constraints on learning.

Keywords: Phonology; Word segmentation; Language universals; Sonority; Sonority sequencing principle

Language acquisition is a product of both the language-specific information learners are exposed to and the intrinsic biases that they bring to the task. Understanding how language is learned requires understanding how the two interact. A number of studies point to a set of very general biases that constitute some of the abilities of the language learner. These include a preference for human speech over acoustically matched non-speech (Vouloumanos & Werker, 2004) and an ability to perceive the phonetic contrasts used in speech (Levitt, Jusczyk, Murray, & Carden, 1988). On the other hand, myriad studies have shown

Correspondence should be sent to Marc Ettlinger, Research Service (mail code 151), Veterans Affairs Northern California Health Care System, 150 Muir Rd, Martinez, CA 94553. E-mail: ettlinger@gmail.com

how native-language input can affect language learning by impacting phone perception and categorization (Werker & Tees, 1992), word recognition (Church, 1987), and of particular interest for this study, word segmentation (Jusczyk, Houston, & Newsome, 1999). Here, we investigate the role a universal linguistic bias—the sonority sequencing principle (SSP)—may play in word segmentation. Adult learners were asked to segment a novel speech stream in which SSP-based segmentation was pitted against transitional probabilities (TPs) or other cues. Results point to a strong role for the SSP in word segmentation, suggesting that it behaves as a universal bias, but importantly, one that can be overcome given the right input.

1.1. Background

The segmentation of the speech stream into words is a non-trivial task as there are no obvious acoustic signals that consistently mark word boundaries (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). At the same time, it is a necessary task in language learning and there is evidence that infants as young as 7.5 months of age can extract at least some words from running speech (Jusczyk & Aslin, 1995).

Despite the lack of consistent acoustic signals, there are various cues for likely word boundaries and studies have shown that learners can make use of many of them. These cues include stress patterns (Jusczyk, Houston, et al., 1999), relative frequency of certain consonant + consonant patterns within and between words (Mattys, Jusczyk, Luce, & Morgan, 1999), and subphonemic co-articulatory cues (Jusczyk, Hohne, & Bauman, 1999). Importantly, TPs can also influence segmentation. In a series of studies, Saffran, Aslin, and Newport (1996) and Saffran, Newport, and Aslin (1996) exposed infants, children, and adults to a continuous stream of different CV-syllables, devoid of pauses or stress. The syllables were grouped into word-like units such that within a word, the syllable TP was 1.0 (the same three syllables appeared together in a word 100% of the time), but across word boundaries was .33. Tests showed that participants of all ages extracted the words from the speech stream, suggesting that learners can use TPs to segment speech.[1]

However, learners do not always follow TPs when segmenting words. Finn and Hudson Kam (2008) found that native-language phonotactics had a significant effect on performance for adults segmenting a novel language. Participants were more likely to segment words that adhered to licit native-language phonotactics over those that were illicit even when TP dictated otherwise. Similarly, Thiessen and Saffran (2003) found that by 9 months, infants use the stress patterns of their native language over TPs to extract words from running speech, while at 6 months they do not (they go with TPs).

Importantly, in all the above cases, the cues examined are language specific and therefore require experience to be useful. In the studies just mentioned, the stress and phonotactic cues were the stress and phonotactic patterns of English, presumably extracted via exposure to English; people exposed to other languages acquire and use the cues relevant to the patterns they are exposed to (Tyler & Cutler, 2009). TP is likewise experience dependent; individual experiments assess participants' ability to extract information from a (artificial)

language during an experimental session and while the ability to use TPs may be universal, the TPs of any one language are specific to that language.

However, typological data suggest that there may be certain universal tendencies of word and syllable formation that restrict the structure of words. These tendencies are hypothetically available to all learners and could guide learners' attempts at word segmentation prior to having acquired any of the language-specific cues.

These typological tendencies of word and syllable formation include a preference for syllables and words to have onsets but not codas (Jakobson, 1962), for syllables to have a greater variety of consonants in onset position as compared to coda position (Beckman, 1997), and—relevant for the purposes of this study—for syllables to adhere to the SSP (Clements, 1990). Sonority is a measure of intensity (Parker, 2008; cf. Ohala, 1990) and the SSP describes the generalization that syllables generally rise in sonority through the onset to a peak at the nucleus. Put another way, more sonorous segments are closer to the vowel in a syllable than less sonorous segments. Individual speech sounds can be classified as more or less sonorous primarily according to their manner of articulation, with vowels being the most sonorous and plosives (sounds like *p* and *b*) the least. Although there is some debate as to whether the SSP is a phonology-specific principle (Clements, 2006) or a consequence of constraints on perception (Steriade, 1999) or production (MacNeilage & Davis, 2000), the fact remains that some sound combinations are more easily articulated and heard than others and that this could help learners find word boundaries.

In particular, sonority information might bias the learner to identify word breaks between speech sounds that violate the SSP. In Fig. 1, we order speech sound types by sonority and show examples demonstrating why some particular syllables are bad or good according to the SSP. In particular, we show why *bl* tends to be a valid complex onset cross-linguistically, as opposed to *lb*. If the learner were to hear the unsegmented string *feelbad*, the SSP would direct her to (correctly) segment the string as *feel* and *bad*, both of which adhere to the SSP, as opposed to *fee* and *lbad* where the onset of *lbad* violates the SSP. Thus, for any language that allows clusters and codas, the SSP can be a very useful word-segmentation cue.

Several studies have explored people's knowledge and use of the SSP in various tasks, and the results generally support the notion that it affects speech perception and production. Redford (2008) showed that novel onset clusters that rose in sonority were more easily
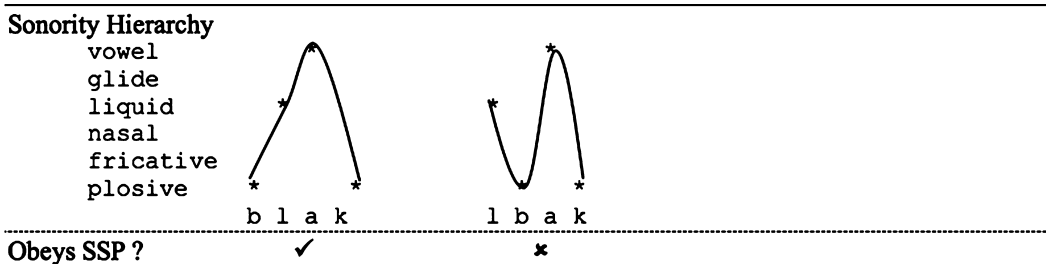

Fig. 1. Words that adhere to and violate the sonority sequencing principle (SSP).

learned and extrapolated for use in syllabification than those that had a sonority plateau. When exposed to words like *tlevat*, with rising sonority in the onset, and *bdevat*, with a sonority plateau, English speakers were more likely to generalize to the syllabification of *vatlet* as *va.tlet* over *vabdet* as *va.bdet*, showing that English speakers have some knowledge of how the SSP affects syllables they have not heard. Furthermore, Berent, Lennertz, Jun, Moreno, and Smolensky (2008) and Berent, Steriade, Lennertz, and Vaknin (2007) showed that the SSP was used by adults in syllable counting and discrimination tasks. Evidence for the influence of the SSP is not limited to adults, consistent with the idea that it is an early-present bias: Children's speech is also affected by the SSP at the earliest stages of language learning. Gierut (1999) showed that the greater the sonority difference between the first and second consonants in an onset, the better performance in onset production. Also, Ohala (1996) showed that children's production errors tended to result in words that reflect greater adherence to the SSP.

So, there is a growing body of evidence suggesting that English speakers show the effects of the SSP on their perception and production of language. Importantly, English (the participants' native language) provides little to no direct information about the syllable structures that were tested in these studies, suggesting that the underlying knowledge participants relied on is not derived from experience with a particular language (cf. Peperkamp, 2007). Instead of looking at cluster perception and production or syllable counting, however, we are interested in how knowledge of the SSP may contribute to language learning, and so we start with an early necessary step: word segmentation. More specifically, we investigate the contribution of the SSP to learners' perception of word boundaries in a novel speech stream.

2. Experiment 1: The impact of SSP on word segmentation

In this first experiment, participants were exposed to a language comprised of words with complex onsets that adhere to the SSP to varying degrees, then tested on their knowledge of the words contained in the speech stream. TPs cued the participants to segment all words with complex (CC) onsets, whereas the SSP cued CC-onsets for some words and simple C-onsets for others. If learners use TPs alone, there should be no difference in participants' ability to segment words that adhere to the SSP versus those that violate it. However, if participants segment words that adhere to the SSP according to TPs but words that violate the SSP in conflict with TPs, it would suggest that the SSP plays a role in language learning.

2.1. Materials and methods

2.1.1. Participants
Participants were 16 (9 females) adult native-English speaking University of California, Berkeley students with no known history of speech or hearing problems.

2.1.2. Stimuli

The stimulus consisted of an 18 min artificial language stream comprising six 2-syllable CCV-CV words, two of which accorded with, and four of which violated, the SSP. As sonority is an ordered cline (Clements, 1990) we included onset clusters with a range of SSP adherence. We gave each cluster a score based on the number of tiers between the two consonants (see Fig. 1). For example, plosives are two tiers below nasals, so *dn* has a score of 2; liquids are two tiers above fricatives, so *lz* has a score of −2. Positive scores represent good clusters; negative scores and zero represent bad clusters (Table 1).

To avoid interference effects, no English or English-like onset clusters were included. Because we are interested in the effect of the SSP (i.e., sonority), we defined English-like as sharing the same manner of articulation, or sonority, as valid English clusters. For example, *dl* and English *gl* consist of stop + liquid sequences, so *dl* was not included. Conversely, *mr* has no English analogs (no *nl*, *nr*, or *ml* onsets), so *mr* was included. All segments were voiced to avoid voiceless fricative + voiced plosive combinations (e.g., *sdop*), which are illegal in English. The clusters included both homorganic (same place of articulation) and heterorganic (different place of articulation) clusters.

The cluster was followed by a vowel: [ɪ], [ɛ], [æ], [ʌ], [ʊ] or [ə]; then a consonant: [p], [k], [b], [g], [m], [v] or [f]; then ended with a vowel that can appear word-finally in English: [i], [u], [o], [ɑ]. Each vowel and consonant occurred once in any position in the language. We used two languages to avoid effects of individual words. Each participant was exposed and tested on only one language.

Each word was repeated 240 times in pseudo-random order such that a word never followed itself and each word was equally likely to occur after another. There were no pauses or any other indication of word boundaries. A sample of the speech stream is shown in (1). Italics highlight words but are not indicative of any acoustic difference.

(1) dnɛku*gbævi*lnʊpo*dnɛku*rdəsu*lzʌfa*mrɪti*gbævi*lnʊpo*lzʌfa

Segmental and syllable TPs were higher within than across word boundaries. Each syllable is unique so the within-word syllabic TP for each word is 1.0; across word boundaries, it is .2 (one of five words follows a given word). The segmental TP within the word ranged from .5–1.0, while the between-word segmental TP was .2. For example, in *lzʌfa* the lowest

Table 1
SSP score of the words in the two input languages

SSP Score	Language	
	1	2
2	**dn**ɛku	**bm**ɪfei
1	**mr**ɪtei	**ml**æpi
0	**gb**ævi	**dg**ʊsɑ
−1	**ln**ʊpo	**rn**ɛko
−2	**lz**ʌfɑ	**rv**ʌtu
−3	**rd**əsai	**lb**ɪzo

Note. SSP, sonority sequencing principle.

within-word segmental TP is .5 since *l* is followed by *z* in *lzʌfa* and *n* in *lnʊpo*. Otherwise, within-word segmental TPs was 1.0 since *z*, *ʌ*, *f*, and *a* appear once. Across word boundaries, the segmental TP is .2 because each word (and its final segment) is followed by one of five other words (and its initial segment). Thus, in both syllabic and segmental terms, within-word TPs were substantially higher than across-word TPs.

We refer to these TP-based strings of segments as ''words'' and use the term ''item'' to refer more generically to any consecutive string of segments. ''Non-words'' or ''part-words,'' explained in more detail below, refer to types of items that are not words.

The stimuli and test items were generated with the text-to-speech program SoftVoice (Katz, 2005), which uses terminal analog format synthesis as opposed to pre-recorded di-phones. Natural speech and di-phones were not used because they potentially include word-segmentation cues from English, including vowel length, co-articulation, and phonotactic cues, all of which are known to affect segmentation (Johnson & Jusczyk, 2001; Rytting, Brew, & Fosler-Lussier, 2010; Tyler & Cutler, 2009). Formant synthesis also eliminates any cues that might make the stimulus more English-like in its properties and excludes cues that provide information on a segment's location in the syllable (e.g., release bursts, dark vs. light /l/). Vowels were 170 ms in length and consonants varied in length from 60 to 140 ms as generated by the SoftVoice synthesizer using an average speaking rate (Katz, 2005). Individual phones were the same length regardless of location within the word, so segment length could not be used as a segmentation cue.

2.1.3. Tests

For each of the two languages, the test consisted of 48 forced-choice trials of three different types. Each test trial consisted of two items, one that appeared as a word in the exposure stimulus and one that had not. All test trials were presented in a randomized order. A summary of the three question types is shown in Fig. 2. Italics are included only for clarity.

The first test type (six trials) was to insure that participants could track syllabic TPs for items with complex onsets. Participants compared a word that was in the stimulus (syllabic TP = 1.0) to a non-word consisting of the first syllable of one word followed by the second syllable from another (syllabic TP = 0). A test trial of this type (hereafter referred to as syllabic non-word) asked, for example, whether *lzʌfa* or *lzʌku* was a better example of a word. (The syllabic TP for *lzʌku* is 0 because *ku* never follows *lzʌ* in the stimulus.) The SSP was not a factor since both had the same complex onset.

The second test type (six trials) assessed participants' sensitivity to segmental TPs. Participants compared a word from the stimulus to a non-word made up of a word with its initial consonant transposed to coda position. For example, a test trial of this type (hereafter segmental non-word) asked whether *lzʌfa* or *zʌfal* was a word. The segmental TPs for segmental non-words are 0 since words cannot repeat themselves. If participants are sensitive to segmental TPs, they should consistently select the word. These items assess whether participants simply prefer items with simple onsets indicated by consistent selection of non-words.

The remaining 36 trials assessed whether segmentation is driven by TPs alone or is also sensitive to the SSP. Participants were asked to compare a word from the stimulus to a

Syllabic Non-Word (NW):

Stimulus: *mrɪteilzʌfɑdnɛku*

Test Items: lzʌfɑ vs. lzʌku
Syllabic TP   1.0   0

Segmental Non-Word (NW):

Stimulus: *mrɪteilzʌfɑdnɛku*

Test Item   lzʌfɑ vs. zʌfɑl
Min. segmental TP   .5   0

Segmental Part-Word (PW):

Stimulus: *mrɪteilzʌfɑdnEku*

Test Items: lzʌfɑ vs. zʌfɑd
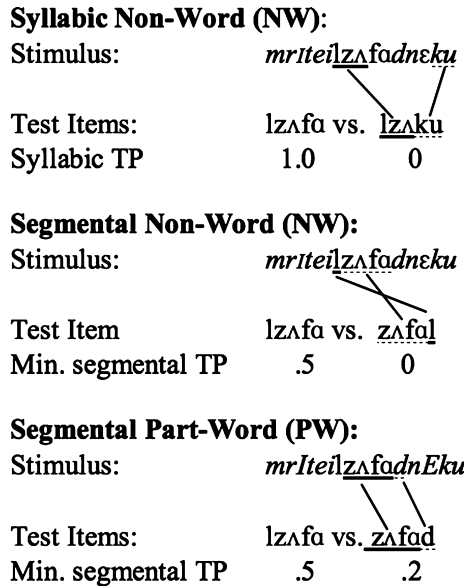Min. segmental TP   .5   .2

Fig. 2. Example test trials for the three test types in the experiments.

part-word item with a simple onset and a coda from the initial consonant of one of the five other words. Words had a minimum within-word segmental TP of .5 and part-words had a minimum segmental TP of .2 (see above). For example, a test trial of this type (hereafter segmental part-word) asked whether *lzʌfa* or *zʌfad* was a word. Both items occurred in the stimulus, but one more predictably than the other. Crucially, some of the words violate the SSP, while some do not. If segmentation is completely dictated by TPs, then participants should always prefer words to part-words (i.e., the more predictable option). If segmentation is sensitive to the SSP in addition to TPs, we expect participants to prefer words whenever they do not conflict with the SSP and part-words with lower TPs when the alternative violates the SSP. Each of the six words was tested against all six possible part-word foils to ensure that participants were not responding based on frequency of forms in the test.

2.1.4. Procedure

Participants were run individually in a quiet room. They were told they were listening to a new language and asked to listen to the stimulus but not to analyze or think too much about it. To encourage this, participants were invited to draw during exposure. Auditory stimulus was presented over headphones. Exposure lasted 18 min, after which the test was administered. For the test, participants heard the two tokens, separated by a 1 s interstimulus interval, and were asked to indicate which was more likely to be a word in the language they had just heard. There was no time limit for responding. Pairs were presented in random order using E-Prime (Schneider, Eschman, & Zuccolotto, 2002).

2.2. Results and Discussion

Recall that there were three different types of test trials, each designed to address a different question: Syllabic non-word trials assessed learning at a general level ensuring that participants were extracting information from the input; segmental non-word trials examined whether participants can track TPs over segments; and segmental part-word trials investigated the involvement of the SSP in word segmentation. We define correct as selection of the TP-defined word.

Participants performed well on syllabic non-word trials. A one-sample *t*-test indicates that participants choose the correct word significantly more often than chance (50%) (64% $t(15) = 3.31$, $p = .002$). Participants also chose the correct word significantly more often than chance for the segmental non-word test trials (64% $t(15) = 2.78$, $p = .007$), suggesting that they can track TPs at the level of individual segments. Importantly, this use of TPs outweighs any dispreference for complex onsets (or preference for codas—segmental non-word). There was no significant difference based on SSP for segmental non-word trials (SSP adhering: 66%; SSP violating: 63%; $t(15) = 0.46$, $p = .75$).

The test type of greatest interest for the question at hand is the segmental part-word. If participants were performing word segmentation on the basis of TPs alone, performance should be above chance. If performance were below chance, it would suggest that participants preferred simple onsets and codas despite this option always violating TPs. In contrast, if they are influenced by other factors, in particular adherence to the SSP of the onsets, we would not expect performance to differ from chance, as some of the words would be correctly segmented but others not. The analysis showed that for segmental part-word trials, performance (48%) did not differ significantly from chance ($t(15) = -0.57$, $p = .57$).

Overall chance performance on these items is not, by itself, meaningful; it could indicate that participants correctly extracted the words starting with clusters that obey the SSP but not those that violate it (per our hypothesis), or simply that they had difficulty extracting novel words. Therefore, we broke down performance according to the word's SSP status (Fig. 3). As predicted, performance on the SSP-adhering words (SSP score > 0) was significantly above chance (64%; $t(15) = 2.60$, $p = .010$), while performance on the SSP-violating words (SSP score ≤ 0) was significantly below chance (40%; $t(15) = -2.30$, $p = .017$). They are also significantly different from each other (paired *t*-test: $t(15) = 3.47$, $p = .002$).

We also assessed whether participants' learning was sensitive to degrees of adherence to the SSP. Fig. 4 shows the mean performance for the different onset clusters' SSP scores. A linear regression of SSP score on performance (clustered by subject) is significant ($F(1, 15) = 28.13$, $p < .0001$) with an adjusted R^2 of .18.

These results suggest that learners' word segmentation is affected by the SSP. Words beginning in SSP-adhering onset clusters were segmented according to TPs, whereas SSP-violating clusters were not. This was observed both when SSP adherence was determined in a binary fashion, with words and their onsets assessed as either SSP-adhering or SSP-violating, and when SSP adherence was measured along a cline. Crucially, when TPs were unequivocal, as in the segmental non-word test, it was TP that guided segmentation and there was no evidence that the SSP played a role.
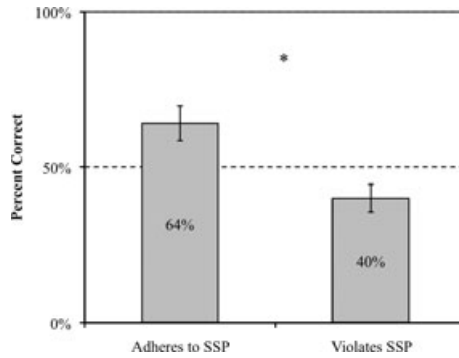
Fig. 3. Mean percent correct by onset cluster sonority sequencing principle (SSP) adherence for Experiment 1. In this graph and all others, error bars represent standard error. * indicates $p < .05$.

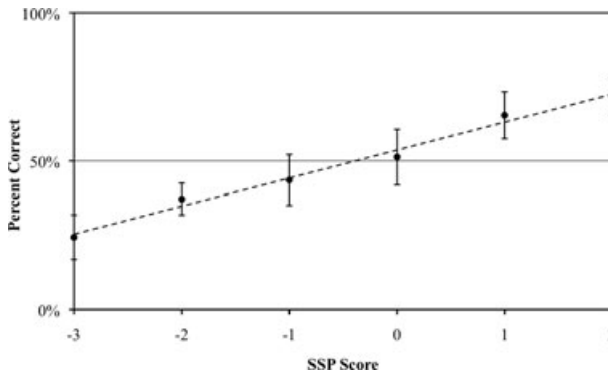

Fig. 4. Mean percent correct by sonority sequencing principle (SSP) score for Experiment 1.

3. Experiment 2

Despite the evidence from the syllabic and segmental non-word tests that the training stimulus itself did influence participants' performance, it is possible that performance on the segmental part-word test, which assessed the impact of the SSP, was due to other factors. In particular, it could be the case that perceptual biases or difficulties (Berent et al., 2007; Davidson, 2006) were directly impacting choice at test time, rather than impacting segmentation which then influenced participants' choices at test time, as we argue. That is, it is possible that participants answered on the basis of what sounded like a good or bad word more generally and not based on the effect of sonority on word segmentation.

To assess the possibility that the results of Experiment 1 were test effects rather than learning differences, we tested participants on the same items without any exposure to the artificial language. Testing without exposure assesses participants' perception and preference for the stimulus items, independent of any learning and segmentation process. If the results from Experiment 1 are due to test effects, then the findings from Experiment 1 should be replicated. That is, participants should still endorse SSP-adhering words, but not SSP-violating words. If, however, performance is independent of SSP status, we can con-

clude that the effects observed in Experiment 1 are reflective of the segmentation of running speech as guided by the SSP.

3.1. Materials and methods

3.1.1. Participants

Participants were 20 (14 females) adult native-English-speaking Northwestern University students with no known history of speech or hearing problems.

3.1.2. Tests

The segmental part-word test trials from Experiment 1, which showed the effects of the SSP, were used.

3.1.3. Procedure

Participants were tested in a similar manner as in Experiment 1, except they were instructed to select the item that ''sounds more like a word.''

3.2. Results and discussion

Participants selected the words only 44% of the time, a rate significantly below chance ($t(19) = 2.9$, $p = .010$) reflecting a preference for the items that were incorrect in Experiment 1.

One possible explanation for the below-chance performance is that participants preferred simple-onset clusters to complex clusters regardless of SSP adherence. This bias is interesting given the fact that English, our participants' native language, allows many different consonant clusters word initially, and so we might expect participants to be more accepting of complex onsets. Another possibility is that participants preferred words with codas, presumably because closed syllables are common in English. A third possibility has to do with the acceptability of the particular onsets and codas we used in their native language; none of the complex onsets from the stimulus occur in English, but all of the codas are attested.

Importantly, below-chance performance was observed for both SSP-adhering and SSP-violating words (43%, $t(19) = 2.2$, $p = .04$; 45%, $t(19) = 2.35$, $p = .03$ respectively), which showed no significant difference ($t(19) = 0.03$, $p = .97$). There was similarly little variation by SSP score (Fig. 5), and no significant correlation between participants' preference and SSP score ($F(1, 19) = 0.04$, $p = .85$). Participants therefore show no preference for the SSP-adhering words over the SSP-violating words when presented solely as segmented test items. This suggests that wordlikeness judgments are not subject to the SSP and that it must be the process of word segmentation that yielded the results in Experiment 1. That is, in Experiment 1, there are two ways in which biases may influence performance: during exposure to unsegmented speech or during testing. Performance in Experiment 2 rules out testing, leaving segmentation during exposure as the remaining option.
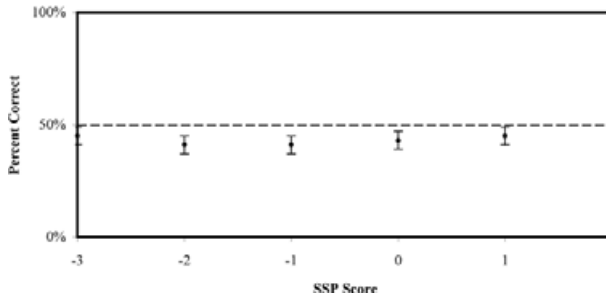
Fig. 5. Percent of complex-onset clusters judged acceptable by sonority sequencing principle (SSP) score for Experiment 2.

4. Experiment 3

Our results thus far point to a role for the SSP in word segmentation. However, there are languages that violate the SSP, including Russian (e.g. /rvat'/ "vomit"), Hebrew (/bga-dim/ "clothes") and, according to some analyses, some words in English (e.g., "stop"; see Vaux & Wolfe, 2009 for discussion).[2] Thus, while the SSP, as a universal, may serve as a cue for word segmentation for language learners, one must also account for the exceptions that must be segmented and learned. How might we account for the correct segmentation of words that violate the SSP? The performance of participants on segmental non-words in Experiment 1 foreshadows an answer. In instances when other cues to segmentation are unambiguous (in segmental non-words, TP = 0), people segment in violation of the phonological universal. In this experiment, we test this possibility more fully by presenting participants with a language with pauses as cues to word boundaries. The stimuli and test items are otherwise the same as in Experiment 1. If participants use the SSP to segment items despite having silence as a cue to intended (and statistically correct) word boundaries, then the question of how SSP-violating words are segmented from speech remains open. However, if participants select words with complex onsets that violate the SSP given this additional cue, then more information regarding word boundaries may result in violations of universal constraints.

4.1. Materials and methods

4.1.1. Participants

Participants were 20 (12 females) adult native-English-speaking Northwestern University students with no known history of speech or hearing problems.

4.1.2. Stimuli

The stimulus was the same as in Experiment 1, except that a 100-ms pause was inserted after each word, increasing the total exposure time to approximately 20 min.

4.1.3. Tests

The tests were the same as in Experiment 1.

4.1.4. Procedure

Participants were tested in the same manner as in Experiment 1.

4.2. Results and discussion

Correct indicates selection of words as defined by pauses and TP. As expected, participants' performance was above chance on syllabic and segmental non-word trials ($t(19) = 15.3$, $p < .0001$; $t(19) = 6.7$, $p < .0001$ respectively). Performance was also above chance on segmental part-words (73%; $t(19) = 12.6$, $p < .0001$), the test of main interest.

As before, we compared performance on segmental part-word trials where words adhered to versus violated the SSP. Unlike Experiment 1, there was no significant difference (Fig. 6; 75% vs. 72%; $t(19) = 0.96$, $p = .35$) and performance for both was significantly better than chance (SSP-adhering: $t(19) = 4.9$, $p < .001$; SSP-violating: $t(19) = 3.9$, $p < .001$). Similarly, performance was not sensitive to degree of SSP adherence; a regression of SSP score on performance is not significant (adjusted $R^2 = .008$; $F(1, 19) = 0.205$, $p = .84$). This contrasts with Experiment 1 where the effect of the SSP was evident, with correct segmentation for SSP-adhering words and incorrect segmentation for SSP-violating words. Thus, when learners are exposed to a segmented stimulus, they accept both SSP-adhering and SSP-violating words suggesting that the combination of TP and pauses cuing segmentation overrides the SSP bias.

A repeated measures ANOVA with exposure (continuous vs. segmented speech stream, or Experiment 1 vs. Experiment 3) as a between-subjects factor and adherence to the SSP (yes, no) as a within-subjects factor shows main effects of exposure ($F(1, 34) = 34.2$, $p < .0001$) and sonority adherence ($F(1, 34) = 11.5$, $p = .002$), with participants performing better with pause-segmented stimulus and with SSP-adhering words, overall. Crucially, there is a sig-
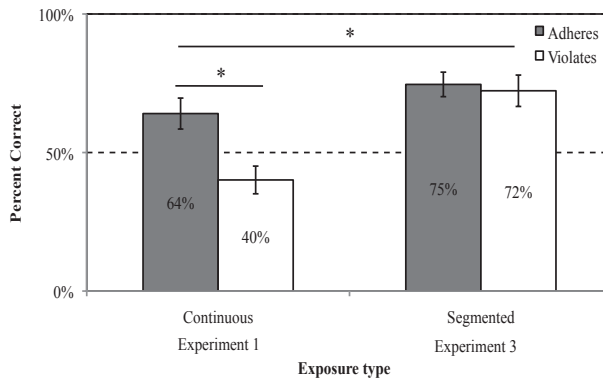

Fig. 6. Performance on sonority sequencing principle (SSP)-violating and SSP-adhering words for Experiments 1 and 3. * Indicates $p < .05$.

nificant interaction between SSP adherence and exposure ($F(1, 34) = 9.3$, $p = .004$), reflecting the fact that SSP was only a significant factor in the performance of participants exposed to the continuous speech stream (Experiment 1), not those exposed to the segmented speech stream (Fig. 6). These results suggest the word-segmentation bias evident in Experiment 1 may be over-ridden by other cues.

5. General discussion

Overall, our study supports the hypothesis that learners' word segmentation is affected by universal biases on syllable structure captured by the SSP. When a word violated the SSP and there was another way to segment the input that adhered to the SSP, participants selected the SSP-adhering alternative despite the fact that it occurred in their input less predictably than the TP-defined word (Experiment 1). This preference was not for the test items themselves, but rather reflected the effects of exposure to the stimulus through the process of segmentation (Experiment 2).

Furthermore, when there were more definitive cues for word segmentation in the stimulus, participants accepted words that violated the SSP, showing that it is not an insurmountable bias. This was demonstrated in Experiment 1 when the SSP-adhering alternative was not a word (i.e., segmental non-words), and in Experiment 3 when silence also marked word boundaries such that the resulting word violated the bias. This makes sense: If segmentation of complex onsets were based on only the SSP independent of any other factors, one could not account for the many languages that have words that violate the SSP, like Russian and Hebrew. It must be the case that with the right confluence of cues, violations of the SSP can be learned. An open question remains as to whether violations were learned because of multiple cues acting in concert (TPs plus silences) or because silence is a particularly strong cue.[3] What we intended to investigate here was merely whether the SSP could be overcome, and we found that it could.

5.1. Universal bias or English knowledge emerging?

In claiming that the SSP is a universal affecting segmentation, it is important to rule out the influence of participants' native-language experience, which has been shown to impact segmentation (Finn & Hudson Kam, 2008). In our study, the SSP knowledge demonstrated by learners was not linked directly to specific knowledge of their native language, as the clusters we tested bear no resemblance to English clusters. However, there is other language-specific information that could contribute to segmentation in this task, namely, the likelihood of a word boundary occurring between the two consonants in question. To rule out the possibility that this was guiding participants, rather than an SSP bias, we examined the frequency of our consonant pairs occurring with (e.g., *said no*) and without (e.g., *madness*) intervening word breaks in the 500,000-word Brent corpus (Brent & Siskind, 2001) of English child-directed speech. For this to explain our results, the SSP-violating pairs should occur across word boundaries more often than not, leading participants to

Table 2
Proportion of appearances of onset cluster that appear with an intervening word break in the Brent corpus (Brent & Siskind, 2001) of child-directed English speech

Language 1			Language 2			
Word	SSP	Proportion Word Break	Word	SSP	Proportion Word Break	Total
dn (ɛku)	2	0.55	bm (ɪfi)	2	1.00	0.78
mr (ɪtei)	1	0.40	ml (æpi)	1	0.69	0.55
gb (ævi)	0	1.00	dg (ʊsɑ)	0	1.00	1.00
ln (ʊpo)	−1	0.98	rn (ɛko)	−1	0.48	0.73
lz (ʌfa)	−2	0.01	rv (ʌtu)	−2	0.64	0.33
rd (əsu)	−3	0.32	lb (ɪzo)	−3	0.92	0.62
Average SSP-adhere		0.48			0.85	**0.66**
Average SSP-violate		0.58			0.76	**0.67**

Note. SSP, sonority sequencing principle.

assume a boundary in between the two consonants based on prior experience. Moreover, the opposite would have to hold for the non-SSP-violating consonants; they would have to occur within a word more often than not, leading participants to assign them to the same word based on prior experience. Table 2 presents the relevant data, in particular, the proportion of each pair containing a word break. There is strikingly little difference between the proportions for SSP-adhering versus SSP-violating clusters (.66 vs. .67) despite the wide variation from cluster to cluster. Furthermore, the absolute frequency of these sets of clusters (140 words-per-million SSP-violating vs. 170 words-per-million SSP-adhering; $t(11) = 0.5$, $p = .62$) suggests that neither set of clusters is more frequent nor more familiar. Thus, it does not appear that participants' experiences with English can explain their performance. Instead, these results support our hypothesis that the SSP (or an SSP-like bias) was responsible for word segmentation in Experiment 1.

Ultimately, it is still possible that there is something about English that makes the SSP particularly salient (e.g., Daland et al., 2011). Future research using speakers of languages other than English would be one way to investigate this. For instance, using speakers of languages without any complex onsets at all, such as Korean, would be a useful way to explore the universality of this bias. Another possibility would be to test infants with similar stimuli. This could also provide further evidence that this bias is universal and could also provide insight into whether this bias requires experience with language or language production to emerge.

5.2. Implications

This pattern of results has implications for thinking about perceptual repair in continuous versus segmented speech. Previous research points to vowel epenthesis as the main repair for English speakers for onset clusters that violate the SSP (Berent et al., 2007; Davidson, 2006). So Berent et al. (2007) found that English speakers judge *lbif* as having two syllables

(i.e., with an additional vowel between *l* and *b*). However, our results are the opposite of what one would predict if participants heard an extra syllable in SSP-violating words in Experiment 1. For example, if participants hear *lbizo* as *lebizo*, participants should select *lebizo* over *bizom* in the segmental part-word test trials because *lebizo* would be preferred by TP cues, but participants consistently select *bizom*.

Thus, the results of Experiment 1 appear not to reflect perceptual repair, and they also suggest that an alternate strategy to epenthesis is possible, and indeed likely, when words are perceived in continuous, natural speech. That is, participants may have resyllabified the initial consonant to the coda of the previous word or simply deleted it in their representations, options not available to participants in studies of the SSP for words in isolation. Because repair strategies were not the focus of the present study, our choices were not designed to distinguish between these possible repairs (resyllabification or deletion). However, the effect of the task on repair strategy, and particularly, how epenthesis might interact with, supersede, or be obviated by word resegmentation, merits further study, particularly since words are rarely heard in isolation in natural speech.

Our results also have a number of broader implications for our understanding of the SSP, language universals, and the process of word segmentation. As described above, there are several possibilities regarding the nature and source of the SSP: it may be domain-specific knowledge, or it may be an emergent property based on the auditory system; it may be present at birth, or it may depend on experience with sonority contours in acoustic stimuli or even in one's own vocal productions, and so learned in some sense (Berent et al., 2007; Clements, 1990; Ohala, 1990; Parker, 2008; Redford, 2008); it may represent an independent organizing principle of syllable structure (Clements, 2006); it may be epiphenomenal of more basic consonant–consonant interactions (Steriade, 1999). Indeed, on this last point the stimulus used in our experiment allows for the possibility that some of our findings are guided by a ban on obstruents as the second consonant in a cluster. Further studies are required to sort out these possibilities. Our data do suggest, however, that it is not a constraint on syllables (or words) per se—the SSP did not guide choices when people were simply asked to choose between possible words absent experience. Rather, it constrains the way people actively segment what they hear. This suggests that the SSP is not, in and of itself, a component of language; knowing a language does not require knowledge of the SSP. Contrast this with the notion of ''principles,'' as part of the Principles and Parameters framework (Chomsky, 1981), where actual components of linguistic structure are argued to be innate. Instead, with the SSP, we have evidence for a phonetic constraint serving to shape language by facilitating its processing or acquisition (Evans & Levinson, 2009; Hawkins, 1994).

Our findings also have implications for understanding the nature of linguistic universals more generally. The presence of the SSP effect in Experiment 1 juxtaposed with the absence of an effect in Experiments 2 and 3 corroborate the observation that few, if any, ''universals,'' are actually immutably universal (Evans & Levinson, 2009; Hyman, 2008). Thus, this study represents evidence for a view of linguistic universals as being universal biases instead of immutable constraints on language. Instead of learners just using the co-occurrence probabilities and statistical information present in their input, as a proverbial blank

slate, and instead of being constrained by hard restrictions on what languages may look like, we find evidence that universals guide the segmentation process in concert with properties of the input.

This study also allows for further elaboration of how bootstrapping might operate in language learning. Our findings suggest that universal phonological or perceptual biases can guide the earliest stages of word-segmentation in concert with TPs and words used in isolation (Brent & Siskind, 2001). Language-specific phonotactic generalizations may then be extracted from these initial words, which can then be used to ascertain further word boundaries (Saffran, 2002). Even if the SSP is only providing information on syllable breaks, this would be a crucial component of statistically driven processes of word segmentation that use syllables and syllable counting (Swingley, 2005).

Finally, these findings can contribute to the growing literature on computational models of word segmentation (Blanchard, Heinz, & Golinkoff, 2010; Davis, 2004; Monaghan & Christiansen, 2010; Rytting et al., 2010; Swingley, 2005). Generally, most of these models seek to minimize external components (Monaghan & Christiansen, 2010), or constraints and biases applied to the model that cannot be inferred from language exposure alone. This may needlessly limit precision, however, as models have successfully incorporated experimentally justifiable biases (Frank, Goldwater, Mansinghka, Griffiths, & Tenenbaum, 2007). The current study represents evidence of just such a bias, and models incorporating the SSP may show improved precision (Bartlett, Kondrak, & Cherry, 2009).

5.3. Summary

To conclude, the present results suggest that the SSP is a language-independent bias that can interact with language-specific information to guide word segmentation. With this conclusion, we join Jusczyk (1997), Johnson and Jusczyk (2001), Saffran (2002), Mattys and Melhorn (2007) and others who have suggested that cues to word segmentation, including both universal and language-specific cues (Tyler & Cutler, 2009), operate in concert and not independently. An open question is precisely how these cues interact both synchronically, at any one stage in development, and ontogenetically. Cues that depend on word-based phonotactic restrictions and stress patterning crucially depend on knowing something about the language being learned. TP has been forwarded as a solution to this chicken-and-egg problem as an early, if not the first, cue that is used to boot-strap others (Saffran, Aslin, et al., 1996). The SSP, being language independent and reflected in children's earliest word productions (Ohala, 1996), requires no knowledge of word-hood to be effective. So it also has the potential to be an early cue facilitating word segmentation, enabling learners to subsequently acquire other more language-specific cues requiring exposure to a particular language. This study demonstrated how language-specific information interacts with a linguistic universal, the SSP, in the task of word segmentation. Ultimately, this experiment represents an attempt toward unifying a theory of language learning with theories of language itself.

Notes

1. Although it is possible that learners are extracting frequently occurring strings as words, rather than using TPs between elements to extract words, the available evidence suggests that this is not the case (e.g., Aslin, Saffran, & Newport, 1998).
2. This is an interesting counter-argument to the idea that English word-likeness may have affected participants' segmentation in Experiment 1: since English allows SSP violations, participants should have been comfortable with SSP-violating words—they are used to them. However, as in Experiment 2, native-language knowledge at this level of abstraction (broad descriptions of what is allowed, such as clusters, SSP violations) does not seem to impact performance on these tasks. While this may seem to contradict Finn and Hudson Kam (2008), they were investigating the impact of much more specific L1 knowledge.
3. Silence in and of itself is not a particularly reliable cue for word segmentation as silences more typically coincide with consonantal closures than with word boundaries. However, our silences were generally longer than voiceless consonants and so were likely more salient, in addition to being correlated with boundaries defined by TP. Furthermore, previous studies (Finn & Hudson Kam, 2008; Gomez, 2002) have shown that pauses can positively impact word segmentation.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324.

Bartlett, S., Kondrak, G., & Cherry, C. (2009). *On the syllabification of phonemes*. Paper presented at the Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL, Boulder, CO.

Beckman, J. N. (1997). Positional faithfulness, positional neutralisation and Shona vowel harmony. *Phonology*, *14*, 1–46.

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(14), 5321–5325.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*(3), 591–630.

Blanchard, D., Heinz, J., & Golinkoff, R. (2010). Modeling the contribution of phonotactic cues to the problem of word segmentation. *Journal of Child Language*, *37*(3), 487–511.

Brent, M. R., & Siskind, J. M. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, *81*(2), B33–B44.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris.

Church, K. W. (1987). *Phonological parsing in speech recognition*. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Clements, G. N. (1990). The role of the sonority cycle in core syllabification. In J. Kingston & M. Beckman (Eds.), *Papers in laboratory phonology 1: Between the grammar and physics of speech* (pp. 283–333). New York: Cambridge University Press.

Clements, G. N. (2006). Does sonority have a phonetic basis? In E. Raimy & C. E. Cairns (Eds.), *Contemporary views on architecture and representations in phonological theory* (pp. 165–176). Cambridge, MA: MIT Press.

Daland, R., Hayes, B., Garellek, M., White, J., Davis, A., & Norrmann, I. (2011). Explaining sonority projection effects. *Phonology*, *28*(2), 197–234.

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*, 104–137.

Davis, M. (2004). Connectionist modelling of lexical segmentation and vocabulary acquisition. In P. Quinlan (Ed.), *Connectionist models of development: Developmental processes in real and artificial neural networks* (pp. 151–188). Hove, UK: Psychology Press.

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Science*, *32*(5), 429–448. discussion 448–494.

Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499.

Frank, M. C., Goldwater, S., Mansinghka, V., Griffiths, T., & Tenenbaum, J. (2007). Modeling human performance on statistical word segmentation tasks. In D. S. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 281–286). Mahwah, NJ: Lawrence Erlbaum.

Gierut, J. A. (1999). Syllable onsets: Clusters and adjuncts in acquisition. *Journal of Speech, Language, and Hearing Research*, *42*(3), 708–726.

Gomez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*(5), 431–436.

Hawkins, J. A. (1994). *A performance theory of order and constituency*. Cambridge, England: Cambridge University Press.

Hyman, L. M. (2008). Universals in phonology. *Linguistic Review*, *25*(1–2), 83–137.

Jakobson, R. (1962). *Selected writings I: Phonological studies*. The Hague, Netherlands: Mouton.

Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, *44*(4), 548–567.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics*, *61*(8), 1465–1476.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in english-learning infants. *Cognitive Psychology*, *39*(3–4), 159–207.

Katz, J. (2005). *SoftVoice*. Los Angeles, CA: SoftVoice, Inc.

Levitt, A., Jusczyk, P. W., Murray, J., & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 361–368.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6), 431–461.

MacNeilage, P. F., & Davis, B. L. (2000). On the origin of internal structure of word forms. *Science*, *288*(5465), 527–531.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology*, *38*(4), 465–494.

Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *The Journal of the Acoustical Society of America*, *122*(1), 554–567.

Monaghan, P., & Christiansen, M. H. (2010). Words in puddles of sound: Modelling psycholinguistic effects in speech segmentation. *Journal of Child Language*, *37*(3), 545–564.

Ohala, J. (1990). Alternatives to the sonority hierarchy for explaining segmental sequential constraints. In M. Ziolkowski, M. Noske, & K. Deaton (Eds.), *CLS 26: Volume 2: The Parasession on the syllable in phonetics and phonology* (Vol. 26, pp. 319–338). Chicago, IL: Chicago Linguistic Society.

Ohala, D. (1996). *Cluster reduction and constraints in acquisition*. Unpublished Ph.D. dissertation, University of Arizona.

Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics*, *1*, 55–90.

Peperkamp, S. (2007). Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition*, *104*(3), 631–637. discussion 638–643.

Redford, M. A. (2008). Production constraints on learning novel onset phonotactics. *Cognition*, *107*(3), 785–816.

Rytting, C. A., Brew, C., & Fosler-Lussier, E. (2010). Segmenting words from natural speech: Subsegmental variation in segmental cues. *Journal of Child Language*, *37*(3), 513–543.

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172–196.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, *35*(4), 606–621.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime*. Pittsburgh, PA: Psychology Software Tools.

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: Karolinum Press.

Swingley, D. (2005). Statistical clustering and the contents of the infant vocabulary. *Cognitive Psychology*, *50*(1), 86–132.

Thiessen, E. D., & Saffran, J. R. (2003). When cues collide: Use of stress and statistical cues to word boundaries by 7- to 9-month-old infants. *Developmental Psychology*, *39*(4), 706–716.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, *126*(1), 367–376.

Vaux, B., & Wolfe, A. (2009). The appendix. In E. Raimy & C. E. Cairns (Eds.), *Contemporary views on architecture and representations in phonology* (pp. 101–143). Cambridge, MA: MIT Press.

Vouloumanos, A., & Werker, J. F. (2004). Tuned to the signal: The privileged status of speech for young infants. *Developmental Science*, *7*(3), 270–276.

Werker, J. F., & Tees, R. C. (1992). The organization and reorganization of human speech perception. *Annual Review of Neuroscience*, *15*, 377–402.