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

Marc Ettlinger, Amy S. Finn, and Carla L. Hudson Kam
University of California, Berkeley

Abstract

It has been well documented that language specific cues—such as transitional probability (TP), stress and phonotactics—can be used for word segmentation. In our current work, we investigate what role a phonological universal, the sonority sequencing principle (SSP), may also play. Participants were presented with an unsegmented stream of speech from an artificial language with non-English onset phonotactics that pitted adherence to the SSP against TP. The participants favored using the SSP over TP in assessing wordhood, suggesting that the SSP may be used by language learners for word segmentation. Importantly, the SSP can be used absent any language experience and represents a new, potentially powerful cue for word segmentation that is grounded in phonological theory.

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

A critical question in language acquisition concerns the interaction between the language-specific information the learner is exposed to by hearing a language and the invariant acoustic, articulatory and psychological biases that are intrinsic to the learner and shape linguistic universals. A number of psycholinguistic studies point to a set of very general perceptual biases that constitute some of the innate abilities of the language learner. These include a preference for human speech over acoustically matched non-speech (Vouloumanos & Werker, 2004) or speech played backwards (Tincoff, Hauser, Tsao, Spaepen, Ramus, & Mehler, 2005), as well as an ability to factor out sources of variability (e.g. speaker variation, Jusczyk, Pisoni, & Mullennix, 1992; Kuhl 1983), to perceive the phonetic contrasts used in speech (Eimas, 1974; Levitt, Jusczyk, Murray, & Carden, 1988), and even to distinguish lexical from grammatical words (Shi & Werker, 2003). On the other hand, myriad studies have shown how native language input can affect language learning in very specific ways: by impacting phone perception and categorization (e.g. Aslin, 1981; Best, McRoberts, & Goodell, 2000; Pitt, 1998; May, Werker, & Gerken 2002; Werker & Tees, 1984), word recognition (Church, 1987), and of particular interest for this study, word segmentation (e.g. Jusczyk, Houston, & Newsome, 1999; Gómez & Gerken, 2000). In the present study we investigate the possible role a linguistic universal - the Sonority Sequencing Principle (SSP) - may play in word segmentation. The results of the study suggest that the SSP indeed represents a specific phonetic bias that guides word segmentation adding to the set of very general universals mentioned above.

Background

The segmentation of the speech stream into words is a non-trivial task as there are no obvious acoustic signals, such as pauses, that consistently mark word boundaries (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Woodward & Aslin, 1990). At the same time, it is a necessary task in language acquisition (Newman, Jusczyk, & Howe, 2003), and there is evidence that infants at least as young as 7.5 months of age can extract at least some words from running speech (Jusczyk & Aslin, 1995).

Numerous studies have investigated the potential cues and strategies that language learners might use in this task. Cutler and Carter (1987) suggest that the stress pattern of a language guides word segmentation, and indeed, Jusczyk, Houston, and Newsome (1999) found that by 7.5-months, English-learning infants are sensitive to the dominant trochaic metrical pattern found in English bi-syllabic words. The phonotactics, allophonic variation, and articulatory cues of a language also influence word segmentation. For example, Mattys, Jusczyk, Luce, and Morgan (1999) found that 9-month-old infants are sensitive to the frequency of certain consonant-consonant (CC) sequences within a word versus across word boundaries in English. Specifically, infants treated nonce words containing CC sequences with a high within-word frequency in English as one word, while nonce words containing CC sequences with a low within-word frequency were treated as two words. In another study, Johnson and Jusczyk (2001) found that infants are sensitive to the sub-phonemic phonotactic patterns of English and can segment words according to subtle differences in articulation that depend on adjacent segments within a word.

Finally, transitional probabilities (TP) also play a significant role in segmenting continuous speech into words.[1] In a landmark study, Saffran, Aslin, and Newport (1996) exposed 8-month-old infants to a continuous stream of 12 different CV-syllables devoid of pauses or stress. The 12 syllables were grouped into four three-syllable 'words' such that within a word, the TP between syllables was 1.0 – the three same syllables appeared together in a word 100% of the time. The words appeared in random order in the speech stream with no repeated words. Therefore, the TP was .33 across words, since one of three words (and its initial syllable) can occur after a given word (and its final syllable). As in natural speech, then, the contingencies between syllables within a word were greater than those for syllables across word boundaries. The results showed that infants extracted the 1.0 TP words from the speech stream; this ability has also been demonstrated in adults (Saffran, Newport, Aslin, Tunick, & Barrueco, 1997; see also Saffran, Newport, & Aslin, 1996). Since the only indicator of word-hood in these studies was transitional probability (TP), the results provide strong evidence that learners are sensitive to the input-specific cue of TP when segmenting speech.[2]

In all of the above cases, the cues examined are language-specific. The stress and phonotactic cues were the stress and phonotactic patterns of English, presumably extracted over time from exposure to English, and the TP experiments assessed participants' ability to extract information from the stimulus during the time course of an experimental session. None of the experiments looked for or pointed to any potential cross-linguistic universals that may be involved in word segmentation. Typological data suggests, however, 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 their 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 (Clements & Keyser, 1983; Jakobson, 1962), for words to have at most one primary stress (Chomsky & Halle, 1968; Liberman & Prince, 1977; Gambell & Yang 2005), for syllables to have a greater variety of consonants in onset position as compared to coda position (Trubetskoy, 1939; Beckman, 1998), and--relevant for the purposes of this study--for syllables to adhere to the Sonority Sequencing Principle (SSP; Sievers, 1881; Whitney, 1885).[3]

The SSP expresses the generalization that syllables generally rise in sonority through the onset to a peak at the nucleus, then fall in sonority through the coda. Put another way, the more sonorous elements are closer to the nucleus than the less sonorous elements. It is not completely clear how sonority is manifested acoustically or articulatorily and a number of phonetic correlates have been suggested including perceptual salience, loudness (Ladefoged, 1993), airflow (Goldsmith, 1995) and resonance (Clements, 2006) but there is, at this time, no consensus (Ohala, 1990). Whatever the actual correlates of sonority are, typological evidence suggests that 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 the least. (The hierarchy is shown in Figure 1.) In Figure 1, we show four 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 onset cross-linguistically, as opposed to *lb,* and why *lb* is generally a valid syllable coda, while *bl* is not.

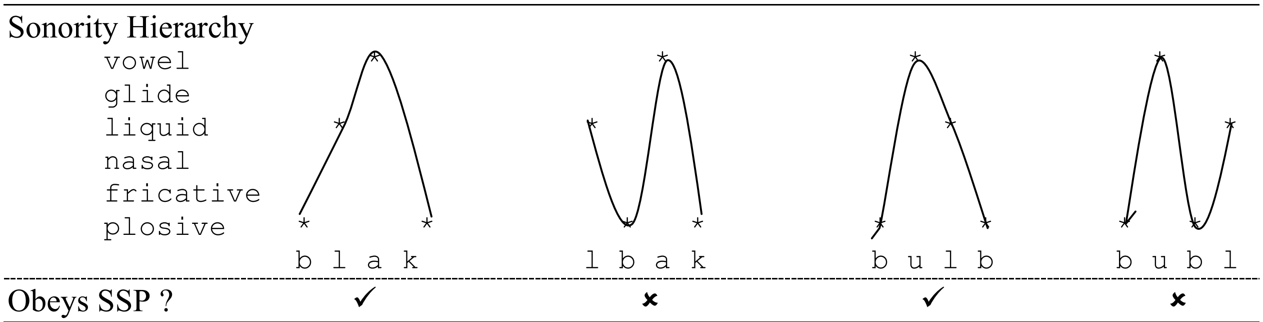

Figure 1: Words that adhere to and violate the SSP

This observed typological pattern reflects a universal that can aid the language learner with word segmentation by biasing her to identify word breaks between speech sounds that violate the SSP. For example, if the learner were to hear the unsegmented string *feelbad*, the SSP would direct the learner 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. Similarly, the unsegmented string *grablots* would be segmented as *grab* and *lots* and not *grabl* and *ots* because the coda of *gra**bl*** violates the SSP.

For any language that allows clusters and codas, the SSP has the potential to be a very useful cue for word segmentation. The aim of our experiment is to test whether language learners use this cue and exhibit a bias towards segmenting words according to the SSP when it is pitted against the conflicting cue of transitional probability. The broader aim is to test for the existence of this specific initial phonological bias in learners to add to the set of general biases outlined above.

A recent study (Berent, Steriade, Lennertz, & Vaknin, in press) found evidence for the SSP being a phonological universal that shapes the perception of onset clusters. In a series of experiments, they demonstrated that the sonority hierarchy was used by adults in syllable counting, discrimination, and lexical decision tasks. The participants in their study were more likely to perceive an illusory epenthetic schwa in the clusters which violate the SSP (*lb* perceived as [ləb] and *bd* as [bəd]) than in clusters that did not (*bn* perceived as [bən]). Because English gives no indication as to whether *bn* or *bd* are valid clusters, Berent et al. concluded that this knowledge must be innate. Pepperkamp (in press) challenges the conclusion, and suggests that the lack of epenthesis on *bn* clusters (crucial for Berent et al.'s argument) may not reflect acceptance of the cluster, but rather the (in)appropriateness of schwa-epenthesis (*bn* perceived as [bən]) as a repair for this particular cluster. Like Berent et al., we are interested in probing the impact of SSP knowledge. Instead of looking at cluster perception, however, we are interested to see how this knowledge may contribute to language learning and are therefore starting with an early step: word segmentation.

The Present Study

We investigated the interaction of the SSP and TP by constructing a set of languages made up of words that have complex onsets that adhere with varying degrees to the SSP. To ensure that the results reflect the abilities of all learners, independent of their actual native language, onsets that are not found in English were used and the phonetic evidence that cue English speakers as to the position of a segment within a syllable was eliminated. If adults used transitional probabilities alone, there would be no difference in the participants' ability to segment words that adhered to the SSP versus those that violated it. However, if participants segment words that adhere to the SSP according to transitional probabilities but segment words that violate the SSP in conflict with transitional probabilities, it would suggest that the SSP could play an important role in language acquisition, in particular, word segmentation.

Method

Participants

Participants were 30 undergraduates (21 females, 9 males) at University of California, Berkeley with a mean age of 21.2 years. All were native English speakers and none reported a history of speech or hearing problems.[4] Individuals participated in partial fulfillment of course requirements.

Materials

Stimuli. Each subject was exposed to 18 minutes of an artificial language consisting of four 2-syllable CCV.CV words. Each word was repeated 240 times in pseudo-random order such that a word never followed itself. Over the course of the entire stimulus, each word was equally likely to occur after another word, but at any given point, the chance of a word occurring was inversely proportional to the number of times the word had already occurred.

There were two main constraints on word construction. First, it was necessary to eliminate any potential influence from English phonotactics. Second, we wanted to include all permutations of the different levels of the sonority hierarchy in the complex onsets (initial CC) of the words.

To avoid interference that could have arisen from the participants' knowledge of valid English onsets, we used only voiced coronal consonants. The only valid English complex onsets consisting of two coronal consonants are *st*, *dr* and *tr*, so using only voiced consonants and excluding *r* avoided these clusters.[5] Glides, or semi-vowels, were also excluded because of the potential confound with English vowel diphthongs and because of the variation of the phoneme /u/ found in English in words like *tune* [tun] ~ [tjun] in certain Southern American English and British English dialects. This leaves the plosive, fricative, nasal and liquid tiers in the sonority hierarchy. The voiced coronal consonants corresponding to those tiers are [d], [z], [n], and [l], respectively.

English phonotactics can also influence word segmentation through syllable-based allophonic variation (Mattys et al., 1999). For example, /t/ is aspirated ([t^h]) in foot initial position but unaspirated ([t]) after [s], so an *st* sequence would be segmented as *s#t* if the [t^h] allophone is heard, but as *#st* if [t] is heard. This particular allophony was avoided by using only voiced segments. Furthermore, the phonemes /n/ and /z/ do not exhibit syllable-based allophonic variation in the speech synthesis program we used, and we used an allophone of /l/ that was ambiguous between dark (coda) and light (onset) *l*. Finally, while the word initial and word final allophones of /d/ are less distinct than those of /t/, differences do nevertheless exist; word-initial /d/ has a greater center of gravity and the pre-voicing has less harmonic structure and stronger fundemental frequency than word-final /d/ (Umeda & Coker, 1974; van Son & van Santen, 2005). The allophone used was word-final /d/. Although this is in itself a potential cue, it did not bias the participants in favor of our hypothesis as addressed in the discussion section.

These can be combined into a two-consonant complex onset in 12 different ways (all of which were used), shown in Figure (2).

dl	zl	nl	--
dn	zn	--	ln
dz	--	nz	lz
--	zd	nd	ld

Figure 2. Combination of the four voiced coronal consonants in complex onsets.

Each of the above combinations has an initial consonant (C_1) followed by three different second consonants (C_2). For example, Figure 2 shows three different *d*-initial clusters, *dl, dn* and *dz* which results in a segmental TP between C_1 and C_2 of .33. To maximize the within-word segmental TP, the 12 clusters were separated into three groups of four so that each C_1 was followed by a unique C_2. Each group of four was a different input language. The rest of each word was made up of a monophthong English vowel ([i], [E], [æ], [u], [o], [a]), a non-coronal English consonant ([p], [k], [b], [g], [m], [v], [f]), and another non-lax monophthong vowel that can appear word-finally in English ([i], [u], [o], [a]). These remaining slots were generated randomly with the condition that each vowel occurred either once or twice and that each consonant appeared only once in any given language.

The three four-word languages are shown in Table 1. Each participant was exposed to only one language and one corresponding set of forced-choice questions. A sample stimulus from Language 2 is shown in (2). Italics indicate word boundaries and are not indicative of any acoustic difference; they are included only to facilitate readability.

2) *zlapo*dnævu*ndEma*dnævu*lzubi*ndEma*lzubi*zlapo*ndEma*zlapo*dnævu*lzubi*ndEma*lzubi*dnævu*ndEma

Table 1

Words in the Three Stimulus Languages

Language		
1	2	3
dlifo	**dn**ævu	**dz**ipa
zdæka	**zl**apo	**zn**æko
nzapu	**nd**Ema	**nl**Emi
lnEmi	**lz**ubi	**ld**ogu

In addition to varying the SSP-adherence of the onsets, we also controlled the transitional probabilities (TP) within words and across word boundaries, such that the TPs were higher within than across word boundaries. We describe these both at the level of the segment and syllable. For all of the words in each language, segmental TPs within each word ranged from .5-1.0. Between-word segmental TPs were always lower, ranging from .33 to .17. In *lzubi*, for example, the lowest within-word segmental TP is .5 since *l* is followed by one of two segments, the *z* in ***l**z̲ubi* and the *a* in *zl̲apo*. The *z*→*u* TP is also .5 (*lzu̲bi*, *z̲lapo*), *u*→*b* is .5 (*lz**u**b̲i*, *dnæv**u***), and *b→i* is 1 (*lzu**b**i̲*). (Formally, the computation is frequency of pair *LZ*/ frequency of *L*, which is ½, or .5.) Across word boundaries, the segmental TP is either .33 or .17 because each word (and its final segment) is followed by one of the three other words (and its initial segment). For example, between the two words *dnævu* and *ndEma*, the TP that *n* follows *u* in stimulus is the number of occurrences of *un* (.33, since for every 3 instances of *dnævu,* it is followed by *ndEma* once) divided by the total number of occurrences of *u*, (.33/2 or .17). The syllabic TP for each word reflects the probability that the second syllable will follow the first. In the stimulus, each syllable in each word is unique so the syllabic TP for each word is 1.0 and across word boundaries, the syllabic TP is always .33.

The stimuli were generated with the text to speech program SoftVoice (Katz, 2005) which uses terminal analog formant synthesis as opposed to pre-recorded di-phones. Natural speech was not used because it potentially provides word-segmentation cues through varying degrees of co-articulation, vowel length, amplitude and frequency, as do di-phones. With SoftVoice, the only co-articulation factored in is string-based rule-governed segment-to-segment acoustic transitions and does not include any information on where in the syllable each segment is. Thus, segment length and loudness were consistent regardless of syllable position and segments were pronounced the same irrespective of whether they were in onset or coda position. The synthesizer produced syllables with a monotonic F0 (fundamental frequency) of ~84 Hz which eliminates F0 pitch cues for stress and intonation. All vowels were the same length regardless of placement next to particular consonants or location within the word. The consonants were not equal in length and are shown in Table 2.

Table 2

Segment Duration (ms)

Segment	Duration
d	80
z	140
n	80
l	60
p	120
b	120
k	100
g	100
m	110
f	130
vowels	170

Tests. For all three languages, the test consisted of 24 forced-choice items comprised of three different types, with each type examining a different aspect of the participants' possible knowledge of the language. The items each consisted of two tokens, one that appeared as a word in the exposure stimulus and one that did not.

The first test type (4 items) was designed to assess whether participants were attending to the task, and to ensure that participants could track words with complex onsets in this type of an artificial language learning task. We asked participants to compare a word that was in the stimulus (syllable TP=1.00) to a non-word consisting of the first syllable of one word followed by the second syllable from another (syllable TP=0). A test item of this type (hereafter referred to as syllabic NW) asked, for example, whether *lzubi* or *lzuma* was a better example of a word. The syllable TP for *lzuma* is 0 because *ma* never follows *lzu* in the stimulus. This test replicated previous studies, such as Saffran *et al.* (1996; 1997), where it is referred to simply as a 'non-word test', except that we used complex onsets in the initial syllable instead of simple CV syllables. The SSP was not a factor since both choices had complex onsets that were equally good (or equally bad) according to the SSP.

The second test type (4 items) assessed participants' sensitivity to segmental transitional probabilities. The participants were asked to compare a word from the stimulus to a non-word made up of the last four segments of an actual stimulus word plus its initial consonant in coda position. Thus, the non-words in this test had simple, as opposed to complex, onsets and closed final syllables. A test item of this type (hereafter referred to as Segmental NW), for example, asked whether *lzubi* or *zubil* was a word. The minimum within-word segmental TP for the non-word is 0 since words can never repeat themselves and each language only has one word with any given initial C. The minimum segmental TP for each correct word was .5 (see above). If participants are sensitive to segmental TPs, they should consistently select the word, as they have never heard the non-word in the input. (The test item choice *zubil* is never heard in the stimulus because only *ndEma*, *dnovu*, or *zlapo* can follow *lzubi*.) In addition to testing participants' sensitivity to the segmental TPs, these items also allow us to rule out the possibility that participants simply prefer words with simple onsets. (Such a preference would result in consistent selection of the non-words.) While potentially interesting, such a preference would interfere with our ability to test the hypothesis that the SSP guides word segmentation.

The final type of test (16 items) assessed whether word segmentation is (at least initially), preferentially directed by TPs alone or is also sensitive to the SSP. In this test participants were asked to compare a word from the stimulus to a part-word with a simple onset and a coda that corresponded to the initial consonant of one of three other possible words. The words had a minimum within-word segmental TP of .5 and the part words had a minimum segmental TP of .17 or .33 (see above). For example, a test item of this type (hereafter referred to as Segmental PW) asked whether *lzubi* or *zubin* 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.[6] If segmentation is completely dictated by TPs, then participants should always prefer words to part-words. If, instead, segmentation is sensitive to the SSP in addition to TPs, we would expect participants to prefer words whenever they do not conflict with the SSP. For words that do conflict with the SSP, participants should prefer the part-words, despite the fact that they have lower TPs than the words.

Syllabic Non-Word (NW):
Stimulus: *zlapo*lzubi*dnovu*ndEma

Test Items: lzubi vs. lzuma
Syllabic TP 1.0 0

Segmental Non-Word (NW):
Stimulus: *zlapo*lzubi*dnovu*ndEma

Test Items: lzubi vs. zubil
Min. segmental TP .5 0

Segmental Part-Word (PW):
Stimulus: *zlapo*lzubi*dnovu*ndEma

Test Items: lzubi vs. zubid
Min. segmental TP .5 .17

Figure 3: Example test items for the 3 test types in the experiment.

A summary of the three different question types is shown in Figure 3. As before, italics are only for clarity.

Procedure

Participants were run individually in a quiet room. The auditory stimulus was presented over headphones for both exposure and test to eliminate any outside noise. Participants were told they were listening to a new language and asked to listen to the stimulus, not to tune it out and not to analyze or think too much about it. To encourage this, participants were instructed to complete a coloring task during exposure. Each participant was exposed to the stimulus for 18 minutes, after which they completed the test. Immediately prior to the test, participants were told they would hear two things, and that their task was to indicate which of two was more likely to be a word in the new language they had just been exposed to. Pairs were presented in a random order using E-Prime (Schneider, Eschman, & Zuccolotto, 2002). The two forms in a pair were separated by a 1 second pause and participants had as long as they wanted to indicate their selection by pressing the “1” key for option 1 or the “2” key for option 2. After the test, the individuals completed a questionnaire regarding their demographic and linguistic backgrounds.

Results

Recall that there were 3 different types of test items, each designed to ask a different question: Syllabic NW items assessed learning at a very general level; it was included primarily to ensure 1) that participants were extracting information from the input, and 2) that the inclusion of complex onsets did not impede learning. Segmental NW items examined whether participants can track TPs over segments within this kind of paradigm, and Segmental PW items investigated the involvement of sonority sequencing in word segmentation. Mean percent correct for all three types are shown in Figure 4, where correct indicates selection of the word that is defined by greater transitional probabilities (i.e., the word that occurred during exposure).

Participants performed well on Syllabic NW items. A one-sample t-test comparing performance to chance (50%) indicates that participants choose the correct word significantly more often than chance ($t(29) = 3.565$, $p < .001$). These results are consistent with previous studies that have shown that adults are sensitive to TPs (e.g. Saffran et al., 1996, 1997). Here, however, individuals tracked TPs when initial syllables had complex, instead of simple, onsets replicating the results in Finn and Hudson Kam (2006). Participants also performed very well on Segmental NW items. Again, participants choose the correct word significantly more often than chance ($t(29) = 2.644$, $p < .05$), suggesting that they can track transitional probabilities at the level of individual segments. Importantly, this choice of TPs also seems to outweigh any dispreference for complex onsets.

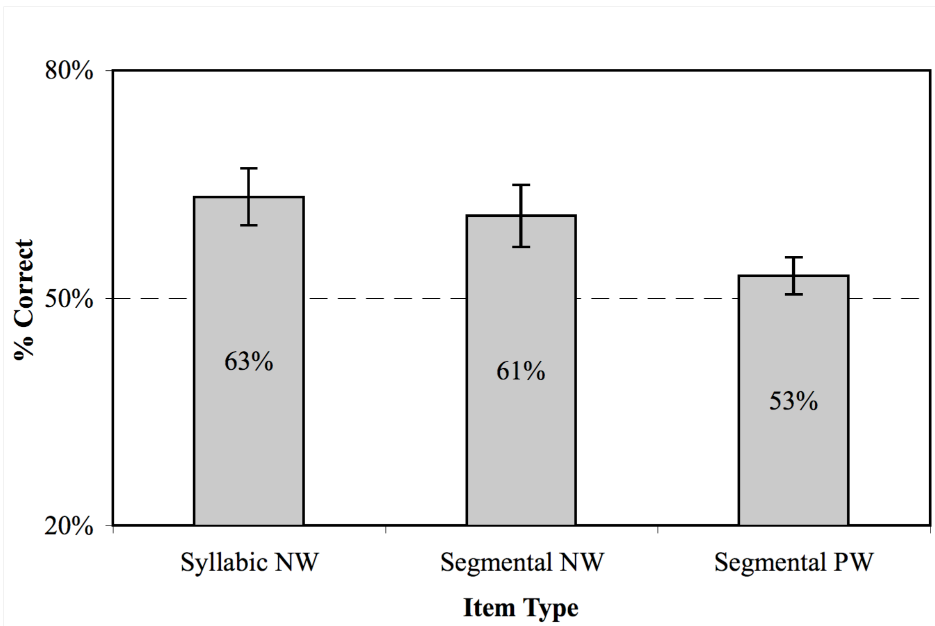

Figure 4: Percent correct by Item type[7]

The test type of greatest interest for us was the third – Segmental PW. If participants are simply performing word segmentation on the basis of TPs, performance here should similarly be above chance. In contrast, if they are influenced by other factors, in particular adherence to the SSP of the various onsets, we would not expect performance to differ from chance, as some of the words would be correctly segmented but others not. (Recall that half of the words in the stimuli adhere to the SPP and half do not). As is clear from the figure, performance on these items is much worse than on the other two types of items and a t-test shows that performance did not differ significantly from chance ($t(29) = 1.183$, $p > .25$) despite the disparity in segmental TP between the words and part-words.

Chance performance in and of itself of course does not mean that the SSP affects word segmentation. To asses this directly, we must look to see whether performance differs on words that accord with or violate the SSP. We therefore coded words according to their adherence to the SSP (onsets are shown in Table 3), and examined the data for each group of words.

Table 3

SSP Status of the 12 Experimental Clusters

SSP status	
Adheres	Violates
dl dn dz	ld lz ln
zl zn	nd nz
nl	zd

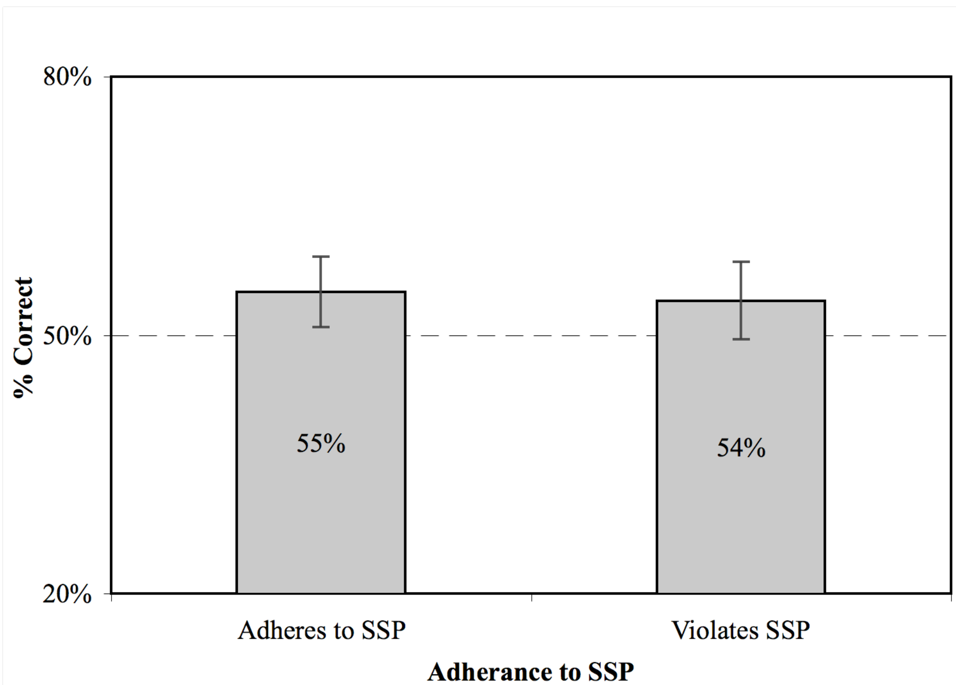

Figure 5: Percent of words segmented correctly by adherence to SSP

Figure 5 shows the percentage of words correctly segmented broken down by whether the word adhered to or violated the SSP. A t-test shows that participant responses did not significantly differ based on whether the onset clusters adhered to or violated the SSP ($t(29) = 0.935$, $p = 0.358$).

Since the SSP is based on a cline (see Figure 1), it is possible that simple adherence or violation is too gross a distinction; maybe participants' learning is sensitive to degrees of goodness or violation. To test this we calculated the degree of adherence to the SSP for each cluster and used this in additional analyses. Each cluster was given a sonority rating according to the number of tiers between the two consonants, in essence, a difference score.[8] For example, for *dl*, plosives are three tiers above liquids, so that cluster receives a rating of three (3), while for *nz*, nasals are one tier below fricatives, so that cluster receives a rating of negative one (-1). The sonority rating for all 12 clusters is shown in Table 4. If participants' segmentation is sensitive to these tiers, we would expect to find a correlation between percent correct and degree of adherence to the SSP. In other words, as the sonority rating increases from -3 to 3, we would expect the percent correct to increase, indicating that participants are more accurate at identifying words the more they adhere to the SSP.

Table 4

Degree of Adherence or Violation of Clusters to the SSP

Ranking	Adherence	Value	Clusters		
Best	Yes	3	dl		
		2	dn	zl	
		1	dz	zn	nl
	No	-1	zd	nz	ln
		-2	nd	lz	
Worst		-3	ld		

Figure 6 is a plot with SSP ranking on the X axis and percent correct on the Y axis. To test whether percent correct differed according to SSP rating, we performed a linear regression clustering by participant. Although a look at the figure suggests that the sonority hierarchy does influence people's word-segmentation, the regression coefficient is .0212 and is not significant ($t(59)=1.34$, $p = .189$).

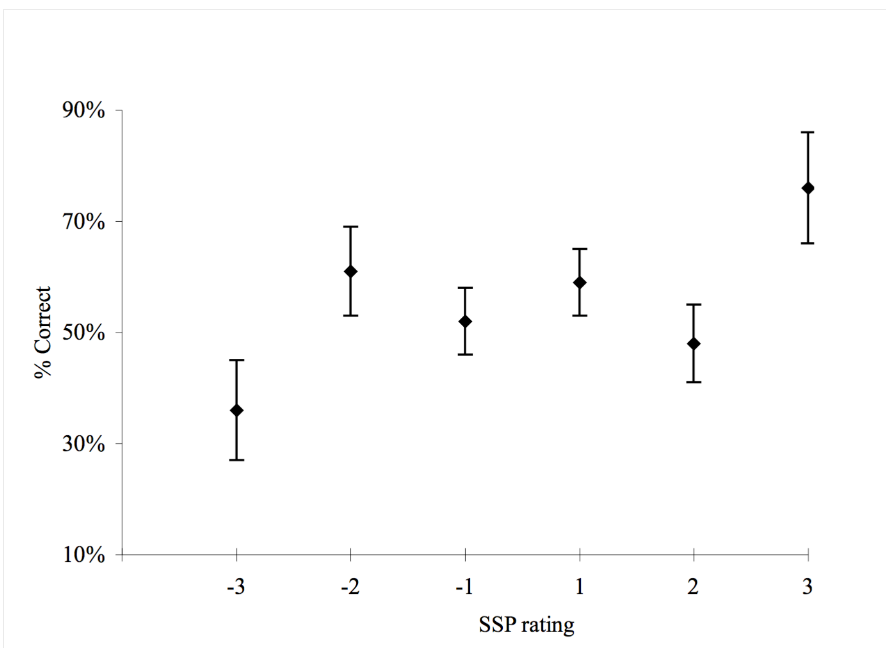

Figure 6: Percent correct by SSP score

The initial hypothesis was that complex onsets interact with TPs by making words with complex onsets that adhere to the SSP easier to segment and onsets that violate the SSP harder to segment, something not supported by the results presented thus far. However, further examination of the data from individual clusters suggests that something more interesting is going on. The lack of significant overall results might be due to four clusters that behave in ways not predicted by the SSP.

Three of the exceptional clusters are ones that, while not the same, are similar to valid onset clusters found in English. Crucial to the hypothesis set out in the introduction was that none of the clusters appear in English lest their acceptability be based on English phonotactics rather than on linguistic universals in the form of the SSP. The onset selection criteria outlined above ensured that no cluster was *identical* to those found in English. However, only voicing distinguishes *zl, zn* and *zd* from the valid English clusters *sl*, *sn*, and *st* respectively.[9] *zd* clusters were segmented correctly significantly above chance at 64% ($t(9) = 2.8$, $p < .05$) despite violating the SSP, while *zl* and *zn* were segmented correctly at 40% and 44% of the time respectively which are not below chance ($t(9) = -1.06$, p = .32; $t(9) = -0.54$, p = .60). As discussed below, the percent correct for these clusters seems to correlate with the frequency of their English analogues.

The fourth exception is the *nd* cluster. Although there is no direct evidence for this, it is quite possible that participants were not actually perceiving *nd* as a complex onset cluster, and thus the cluster did not violate the SSP. Homorganic nasal + plosive onsets, like *nd*, are found relatively frequently cross-linguistically and are treated, not as complex onsets, but rather as pre-nasalized plosives. Maddieson's (1984) survey showed that 5.6% (18) of the 300+ languages surveyed had a pre-nasalized stop series, so the nasal + plosive onset, and specifically the *nd* onset, appears to be a patterned exception to the SSP. Therefore, *ndEma* is potentially perceived as [^{n}dEma]. If this were indeed the case, it would not be in violation of the SSP. Another possibility is that the *n* is actually being perceived as a syllabic nasal, which are found in natural speech in English in words like *indifferent* [n̄.dɪ.frənt]. In this scenario, *ndEma* would be parsed as [n̄.dE.ma], which again, is not in violation of the SSP because the [n] is its own syllable. English lacks word-initial syllabic *d, l* or *z* so this would not be a possibility for clusters beginning with those sounds. Words beginning with the *nd* cluster were segmented correctly 88% of the time despite violating the SSP suggesting that this exception affected the participants' performance.

Removing these four clusters – the cross-linguistically acceptable *nd* and the English-like *zd, zl* and *zn* – from the overall results shows that the SSP does, in fact, appear to play a role in the segmentation of novel speech. Figure 7 shows the percentage of words segmented according to TP by whether the cluster adheres to or violates the SSP (minus *nd*, *zd*, *zl* and *zn*). Those that adhere to the SSP were segmented correctly significantly more often than chance ($t(29) = 2.19$, $p < .05$) while those that violate it were segmented correctly significantly less often than chance ($t(29) = -2.12$, $p < .05$). This is exactly the pattern of results we would expect; when the structure of the word accords with the SSP participants accept it as a word, and when it does not, they reject it.

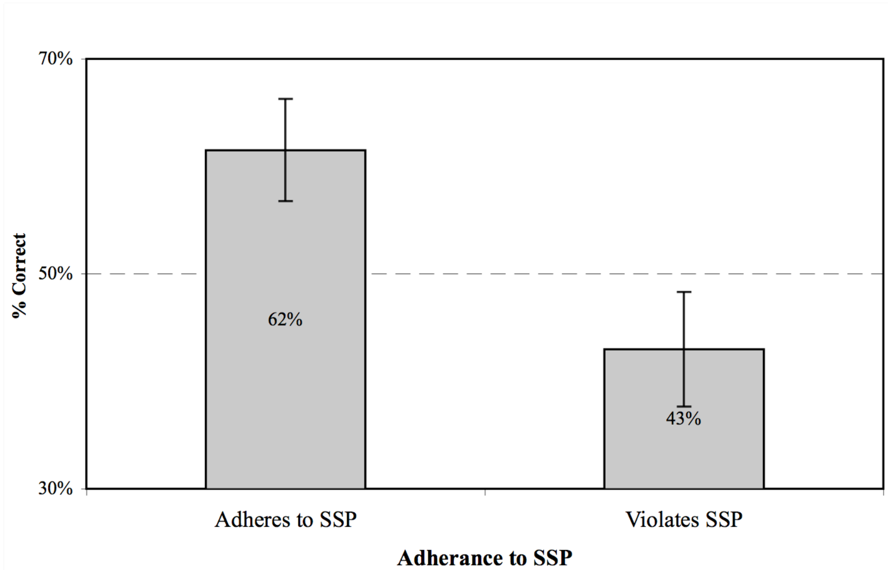

Figure 7: Percent Correct by Adherence to SSP (minus *nd, zl, zn* and *zd* clusters)

The difference is not simply a categorical one, however. It also appears that participants' segmentation was sensitive to the SSP cline, as shown in Figure 8. An additional linear regression clustering by participant yielded a regression coefficient of .05 which is significant ($t(58) = 2.91$, $p<.01$).

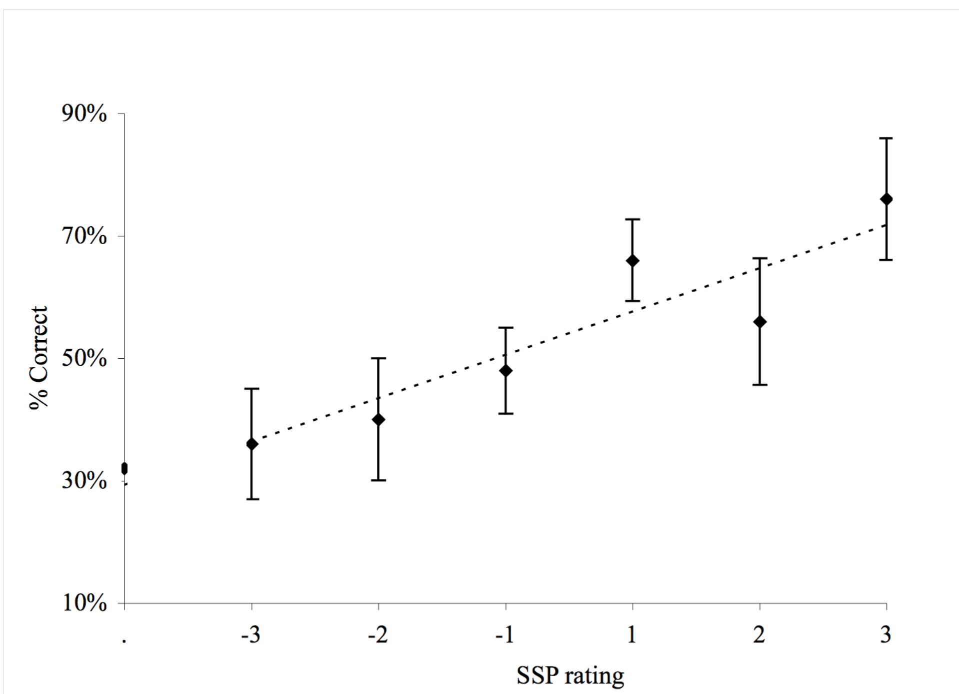

Figure 8: Percent correct by SSP score (minus *nd, zl, zn* and *zd* clusters)

Discussion

The results of the present study support the hypothesis that learners' word segmentation is affected by the SSP. Once we removed the four clusters that seemed to be processed differently, we found that words that adhere to the SSP were segmented correctly more often than those that violate it. Participants relied on TP when the TP of the incorrect choice was zero, but when it was not, their segmentation was guided more by the SSP than TP. This is similar to the results in Johnson & Juscyk (2001), for example, where when faced with conflicting phonotactic and TP cues, infants opted for phonotactic word segmentation. Moreover, we found that the effect of the SSP on segmentation was gradient; clusters were not simply treated as being either licit or illicit, rather, participant's performance suggests that there are degrees of goodness (or badness). Thus, the SSP, a language-independent linguistic universal related to the sequencing of segments within syllables, appears to be available to language learners as a cue to word boundaries.

One potential counter-argument to this conclusion is that the effect found here is not indicative of any linguistic universal, but rather, is based on interpolation or analogy from participant's knowledge of English. However, although this argument could hold for some of the clusters we tested (*zd, zl* and *zn* as described above; *dl, ld,* and *dz* described below), it cannot explain the results for the rest. Table 5 shows the eight onsets included in Figures 7 and 8 along with their potential English onset analogues. The validity of *dl* and invalidity of *ld* could have been extrapolated from participants' knowledge that English allows similar stop-liquid onsets like *dr, bl* and *gl*, but not their opposites. Similarly, the *dz* results could be based on the validity of English affricates *ch* and *j*. However, for the remaining clusters, there is no way for an English speaker to ascertain the clusters' relative rankings based on a knowledge of English words either directly, from the existence of similar sequences in English, or indirectly, by interpolating the relative ranking based on other clusters. For example, nothing in English says whether *nl* is a valid or invalid cluster because there are no nasal-liquid or liquid-nasal onset clusters nor are there any clusters that allow a ranking of nasals and liquids by reference to a third tier.

Also, as mentioned previously, the synthesized *d* segment had the acoustic properties of English word-final *d*. This could have biased participants to segment *d*s in a particular way, in particular, to put a boundary after the *d*. When the *d* is in cluster-initial position, this would cue a word-break between the two consonants, and thus should lead participants to select the split-cluster word at test. However, as all *d*-initial clusters adhere to the SSP, this cue creates a bias against our hypothesis. Performance on words with *d* in initial position was good (mean = 76%), demonstrating that this allophonic cue does not appear to be driving participants performance on these words. All onsets with *d* in C_2 position violate the SSP, and so the cue would lead to a different segmentation for these words, a word boundary between the cluster and the first vowel in the word. This option was not available at test, and so if the allophonic cue was driving performance on these items we might expect chance performance. Again, this is different from what our hypothesis predicts, which is below chance performance (due to a predicted preference for the split-cluster items). On these items as well it seems that it is the SSP, and not the allophonic cue, that is driving performance (mean = 36%).

Table 5

Complex onsets and potential English analogues

Ranking	Adherence	Value	Clusters	English Analogue
Best	Yes	3	dl	bl, gl, dr
		2	dn	*n/a*
		1	dz	tΣ, dZ
			nl	*n/a*
	No	-1	nz	*n/a*
			ln	*n/a*
		-2	lz	*n/a*
Worst		-3	ld	*lb, *lg, *rd

Another possibility that may come to mind is that participants deemed acceptable clusters in which the two individual segments are each very likely to occur in their respective positions in clusters in English, irrespective of the other member of the cluster. Or put another way, that participants learned words when the first phone occurs very frequently in first position in other English clusters, and the second phone likewise occurs very frequently in the second position in English clusters. For example, based on token frequencies in the Brown corpus (Kucera & Francis, 1967) *d* occurs cluster initially 1336 times (in words like *draw*) while *l* occurs in C_2 position 7473 times (*clap*). While this can explain the acceptability of *dl* as a valid onset, it fails to explain most of our results: None of the other segments appear cluster-initially and only *n* appears as a C2 in English. This is problematic. It would suggest, for example, that *dz* would be unacceptable, despite adhering to the SSP, and yet it is segmented correctly 96% of the time.

English-specific knowledge of word-internal syllabification also does not provide a basis for the results. While the syllabification of a word like *ill.ness* correctly predicts that *ln* would not be treated as a valid word onset, the syllabification of *man.ly* incorrectly predicts that the *nl* onset would be similarly rejected contrary to the results above. Thus, we contend that performance reflects the phonological or phonetic universal embodied by the SSP, and not language-dependent experience.

Participants' performance on the two sets of exceptional onsets (*zd, zl, zn,* and *nd*) suggests some additional conclusions. The first concerns the nature of phonotactic knowledge. In the present experiment, participants correctly segmented onset clusters that were *similar* to valid English onsets at the rate of 48%; individually, the percent correct was 64% for *zd*, 44% for *zl* and 40% for *zn*. The Brown corpus token frequency of the English analogues of these three clusters is 8045 (*st)*, 443 (*sl)*, 149 (*sn*) respectively suggesting frequency may play a role when extrapolating the acceptability of English clusters to novel ones. While not within the scope of investigation of this study, these results suggest further areas of investigation.[10]

The results also suggest that *nd* is a special onset with a perceptual bias that patterns independently of the SSP, something that likely generalizes to all nasal + plosive clusters, such as *mb* and *nk*. Our data gives little evidence as to the ultimate phonological analysis of this onset

– either as a pre-nasalized plosive or syllabic nasal. As mentioned above, nasal + plosive sequences are common cross-linguistically, and are often treated as pre-nasalized plosives. In contrast, there are few, if any, pre-nasalized fricatives in the world's languages. This suggests that if the nasals were perceived as pre-nasalization, *nz* should have been unaffected and the segmentation of words containing *nz* should have been affected only by the SSP, in contrast to *nd* and *nl*. That is, *nz* should be illicit whereas *nd* and *nl* should not be. However if the results instead reflect a syllabification of the nasal segment (as in [n̄.dI.frənt]), something which is equally possible for the other nasal initial sequences, *nd, nz* and *nl*, it should be treated the same. Therefore, we would expect similarly high performance on items containing all three of these clusters. The data do not clearly support one or the other possibility. On the one hand, *nz* onset clusters, which violate the SSP, are segmented correctly 64% of the time (as compared to 84% for nd), suggesting that nasal syllabification is a possible explanation for the acceptability of both. On the other, *nl* is segmented correctly only 46% of the time despite both adhering to the SSP and being nasal-initial which suggests that it is not the syllabification of the nasal alone that is causing the observed results in the *nd* and *nz* clusters.[11]

The most important implication of our results, however, is that word segmentation is based on the *interaction* of the SSP and segmental TP, and not on the SSP (or TP) alone. Participants were perfectly willing to accept words which violate the SSP when there was very clear evidence that the alternative was not a word (i.e., it was a string of phones which had never occurred in their input). However, when the word violated the SSP and there was another possible way to segment the input, participants selected it, despite the fact that it was less likely than the TP-defined word. This suggests that the SSP's roll in word segmentation partially depends on TP. This makes sense: if segmentation of complex onsets were based on only the SSP and operated independently of any other factors, it would be difficult to account for the many languages that have words that violate the SSP. For example, in Ladakhi (Koshal, 1979) initial liquid-obstruent clusters /lg-/, /lz-/, /rb-/, /rg-/, etc. are allowed, and Russian exhibits a number of SSP violations in words like /mgnovenije/ 'moment' /lstets/ 'flatterer' /mzda/ 'bribe' and /rvat'/ 'vomit'. When other cues to word boundaries, such as allophonic variation or TP, clearly indicate a word that violates the SSP, as in the real world examples above and the word vs. segmental non-word tests in our study, violations can be learned.

With this conclusion, we join Jusckyk (1997), Saffran (2002), Thiessen & Saffran (2003), and many others who have suggested that cues to word-segmentation 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. So, if the t^h~t~t|~P allophonic alternation for /t/ is to be used to segment words in English, the learner first needs to learn which of those allophones occur word internally and word-initially, which of course requires knowing at least some word boundaries.

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 the others (Saffran, Aslin, & Newport, 1996). The SSP, being language independent, requires no knowledge of word-hood to be effective so, it, too, has the potential to be an early cue facilitating word-segmentation, enabling learners to subsequently acquire other cues that require more language-specific knowledge. This study demonstrated how language specific information, transitional probabilities, interacts with a linguistic universal, the Sonority Sequencing Principle, in the task of word segmentation. Ultimately, this experiment represents an attempt towards unifying a theory of language acquisition with theories of language itself.

word count: 7106

References

Aslin, R. N. (1981). Experiential influences and sensitive periods in perceptual development: A unified model. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception: Psychobiological perspectives, Vol II* (pp. 45-93). New York: Academic Press,.

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

Beckman, J. (1998). Positional Faithfulness. (Doctoral dissertation, University of Massachusetts, 1998). *Dissertation Abstract International 59*(02), 468A.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (in press). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition.* (2007) doi:10.1016/j.cognition.2006.05.015.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). American listeners' perception of nonnative consonant contrasts varying in perceptual assimilation to English phonology. *Journal of the Acoustical Society of America, 1097*, 775-794.

Blevins, J. (2003). The independent nature of phonotactic constraints: an alternative to syllable based approaches. In C.Féry & R. van de Weijer (Eds.), *The Syllable in Optimality Theory*. Cambridge: Cambridge University Press.

Blevins, J. (2004). *Evolutionary phonology: The emergence of sound patterns*. Cambridge: Cambridge University Press.

Chomsky, N., & Halle, M. (1968). *The sound pattern of English.* New York: Harper and Row.

Church, K.W. (1987). *Phonological parsing in speech recognition.* Dordrecht: 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 I* (pp. 283-333). Cambridge: Cambridge University Press.

Clements, G. N., & Keyser, S. J. (1983). Chapter 2: A three-tiered theory of the syllable. *CV Phonology: A Generative Theory of the Syllable* (pp. 25-56). Cambridge: MIT Press.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Lawrence Erlbaum.

Cutler, A. & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2*, 133-142.

Eimas, P.D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception & Psychophysics, 16*, 513–521.

Finn, A., & Hudson Kam, C. L. (2006). Use of Word Segmentation Cues in Adults: L1 Phonotactics versus L2 Transitional Probabilities. In R. Sun (Ed.) *Proceedings of the 28th Annual Meeting of the Cognitive Science Society* (pp. 1229-1304). Mahwah, NJ: Erlbaum.

Gambell, T., & Yang, C (2005). *Mechanisms and Constraints in Word Segmentation*. Unpublished manuscript, Yale University.

Gaygen, D. E. (1997). The effects of probabilistic phonotactics on the segmentation of continuous speech. (Doctoral dissertation, State University of New York at Buffalo, 1997). *Dissertation Abstracts International 60*(02) 848A.

Goldsmith, John A. (1995). Phonological Theory. In J. Goldsmith (Ed.) *The Handbook of Phonological Theory* (pp. 1-23). Oxford: Basil Blackwell.

Gouskova, M. (2004). Relational hierarchies in OT: the case of Syllable Contact. *Phonology, 21*, 201-250.

Gómez, R. & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences, 4,* 178-86.

Hockema, S. (2006). Finding words in speech: An investigation of American English. *Language Learning and Development, 2*, 119-146.

Jakobson, R. (1962). *Selected writings. Volume I: Phonological studies*. Gravenhage: Mouton.

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

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

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*, 323-328.

Jusczyk, P., & Aslin, R. (1995). Infant's detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 46*, 65-97.

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

Jusczyk, P., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630-645.

Jusczyk, P.W., Pisoni, D.B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month old infants. *Cognition, 43* , 253–291.

Koshal, S. (1979). *Ladakhi Grammar.* Delhi: Motilal Banarsidass.

Kucera, H. & Francis, W.N. (1967). *Computational Analysis of Present-Day American English.* Providence: Brown University Press.

Kuhl, P.K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6*, 263–285.

Ladefoged, P. (1993). *A Course in Phonetics* (3rd ed.). Orlando: Harcourt Brace & Company.

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

Liberman, M., & Prince, P. (1977). On Stress and Linguistic Rhythm. *Linguistic Inquiry, 8*, 249-336.

Levitt, A., Jusczyk, P.W., Murray, J., & Carden, G. (1988). The perception of place of articulation contrasts in voiced and voiceless fricatives by two month-old infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 361–368.

MacNeilage, P., & Davis, B. (2000). On the Origin of Internal Structure of Word Forms. *Science*, 288, 527 – 531.

Maddieson, I. (1984). *Patterns of sounds*. Cambridge: Cambridge University Press.

Mattys, S., & Jusczyk, P. (2001). Do infants segment words or recurring contiguous patterns? *Journal of Experimental Psychology: Human Perception and Performance, 27*, 644-655.

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

Maye, J., Werker, J.F., & Gerken, L.A. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101-B111.

Newman, R. S., Jusczyk, P. W., & Howe, M. (2003). The effect of previously learned words on children's acquisition of similar word forms. In D. Houston, A Seidl, G. Hollich, E. Johnson, & A. Jusczyk (Eds.) *Jusczyk Lab Final Report*. Retrieved January 12, 2007, from http://hincapie.psych.purdue.edu/Jusczyk.

Ohala, J. (1990). Alternatives to the Sonority Hierarchy. In M. Ziolkowski, M.Noske & K. Deaton (Eds.) *Papers from the 26th Regional Meeting of the Chicago Linguistics Society* (pp. 319-338). Chicago: Chicago Linguistic Society.

Peperkamp, S. (in press) Do we have innate knowledge about phonological markedness? Comments on Berent, Steriade, Lennertz, and Vaknin. *Cognition* (2007), doi:10.1016/j.cognition.2006.12.009

Pitt, M. A. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics 60*, 941-951.

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

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month-olds. *Science, 274,* 1926-928.

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

Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science, 8*, 101–105.

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

Sievers, E. (1881). *Grundzüge der Phonetik.* Leipzig: Breitkopf und Hartel.

Shi, R., & Werker, J. F. (2003). Basis of preference for lexical words in six-month-old infants. *Developmental Science, 6,* 484–488.

Steriade, D. 1999. Alternatives to the syllabic interpretation 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.

Thiessen, E., & 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*, 706–716.

Tincoff, R., Hauser, M., Tsao, F., Spaepen, G., Ramus, F., & Mehler, J. (2005). The role of speech rhythm in language discrimination: Further tests with a nonhuman primate. *Developmental Science*, *8*, 26-35.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language, 27*, 87-104

Trubetskoy, N. S. (1969). *Principles of Phonology*. (C.A.M. Baltaxe, Trans.). Berkeley/Los Angeles: University of California Press. (Original work published 1939).

Umeda, N. & Coker, C. H. (1974). Allophonic variation in American English. *Journal of Phonetics*, *2*, 1-5.

van Son, R.J.J.H. & van Santen, J.P.H. Duration and spectral balance of intervocalic consonants: A case for efficient communication. *Speech Communication*, *47*, 100-123.

Vennemann, T. (1988). *Preference Laws for Syllable Structure and the Explanation of Sound Change*. Berlin: Mouton de Gruyter.

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

Werker, J.F., & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49-63.

Whitney, W. D. (1865) The relation of vowel and consonant. *Journal of the American Oriental Society*, *8*. 357-73.

Woodward, J. Z., & Aslin, R. N. (1990). *Segmentation cues in maternal speech to infants.* Poster presented at the International Conference on Infant Studies, Montreal, Canada.

Endnotes

[1] While Gambell and Yang (2005) show that syllable TPs only have limited application for segmenting English, this does not preclude syllable TPs from being part of a more intricate boot-strapping mechanism for acquiring word boundaries. Additionally, Gambell & Yang do not include segmental transitional probabilities in their assessment which can provide a significantly higher rate of word identification. (Gaygen, 1997; Hockema, 2006).

[2] In most statistical learning studies, string frequency is confounded with TP, so it is possible that learners are extracting frequently occurring strings as chunks, as opposed to using TPs to extract words. However, the available evidence suggests that this is not the case (e.g. Aslin, Saffran, & Newport, 1998; Saffran, Newport, & Aslin, 1996).

[3] There is a great deal of discussion and disagreement in the literature as to whether the SSP is an independent linguistic universal (Clements, 2006; Vaux 2006) or is instead a label for or description of other more basic phonetic, phonological, or historical facts (i.e. it is epiphenomenal, Blevins 2004; Harris, 2006; MacNeilage & Davis, 2000; Ohala, 1990; Steriade 1999). An answer to this question is not necessary for the present work, and although we see much merit in the latter view (see e.g. Berent, Steriade, Lennertz, & Vaknin, to appear; Blevins, 2003; Pitt, 1998; Steriade, 1999), we take no strong position here.

[4] Five of the participants were bilingual. None of the other languages the participants were fluent in (Spanish, Korean, Mandarin, Punjabi, or Tagalog) have any of the clusters used in the experiment. Clusters that were valid in the other language that were invalid in English were voiceless (e.g. Mandarin t⌋) and included phonemes not found in English (e.g. Spanish tP; compare English t♦).

[5] This also served to avoid using cluster-initial *s* which has special status in English. The *s* in a word like *stop* has been considered by some to not be in the same syllable as the following *t* (Pierrehumbert & Nair, 1995; Treiman, Gross, & Cwikiel-Glavin, 1992; Treiman & Zukowski, 1990; Vaux, 2006) and so the word *stop* does not violate the SSP. It should be noted that Trieman et. al. (1992) also argue that *s* is not in the same syllable as following sonorants either, as in clusters like <*s*>*.not*. Also, see fn. 12 below.

[6] This is also true for items in the second test type. However, we anticipated that participants would be willing to override the SSP when there was no evidence for the other form in the input, and so expected those items to be relatively unaffected by the SSP.

[7] In this graph and all others, error bars represent standard error.

[8] This is a simplified version of the 15-point rating rubric outlined in Gouskova (2004) where place of articulation and voicing also play a role.

[9] This raises the question of why the *st* cluster is valid in English in the first place since it violates the SSP. Most accounts that attempt to maintain the applicability of the SSP argue that the *s* is extra-syllabic, or not part of the syllable, and therefore not relevant to the SSP. Despite the fact that this approach has been argued to be an ad hoc way of saving the generalization (Ohala 1990), there is ample linguistic and psycholinguistic evidence to suggest that s-initial clusters are, indeed, special (Stemberger & Treiman, 1986; Treiman & Fowler, 1991; Treiman & Zukowski ,1990). The word-segmentation results here suggest that *z* is being treated like an s and is similarly extra-syllabic. Further experimentation on nonce words with *z* + consonant onsets or on the segmentation preference on words like mesmerize [mEzmraiz] would further illuminate this issue.

[10] Open questions include how similarity is assessed and what effect this similarity has on the acquisition of an artificial language. One possibility is that similarity may be based on feature

sharing (Clements & Hume, 1984) with voicing being the only feature distinguishing the novel clusters (*zl, zn, zd*) and valid English clusters (*sl, sn, st*). Alternatively, similarity could be based on the fact that voiced segments have similar cues and cue-bearing properties as their voiceless counterparts. This is based on the idea that the perception of segments is not only the result of the perception of acoustic cues during the articulation of each individual segment, but is also based on cues found on adjacent segments. While our results suggest that similarity to high frequency items aids artificial language acquisition, while similarity to low frequency items hinders it, at this point we only conclude that there is some sort of interference.

[11] A third possibility may present itself to some readers, that *nd* and *nz* commonly occur as codas within the same syllable in English, whereas *nl* does not. However, if this were what was guiding word segmentation it would actually predict that participants would consistently reject words starting with *nd* and *nz*, since they occur at the *ends* of words, not the beginning.

Figure Captions

Figure 1. Words that adhere to and violate the SSP
Figure 2. Combination of the four voiced coronal consonants in complex onsets
Figure 3. Example test items for the 3 test types in the experiment.
Figure 4. Percent correct by Item type
Figure 5. Percent of words segmented correctly by adherence to SSP
Figure 6. Percent correct by SSP score
Figure 7. Percent Correct by Adherence to SSP (minus *nd, zl, zn* and *zd* clusters)
Figure 8. Percent correct by SSP score (minus *nd, zl, zn* and *zd* clusters)