The Pennsylvania State University

The Graduate School

College of the Liberal Arts

RESOLVING COMPETITION IN BILINGUAL STATISTICAL LEARNING PARADIGMS

A Thesis in

Psychology

By

Aaron D. Mitchel

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

May 2008

The thesis of Aaron D. Mitchel was reviewed and approved* by the following:

Daniel J. Weiss
Assistant Professor of Psychology
Thesis Adviser

Judith F. Kroll
Distinguished Professor of Psychology, Linguistics, and Women's Studies

Frank G. Hillary
Assistant Professor of Psychology

Henry (Chip) Gerfen
Associate Professor of Spanish and Linguistics

Mel Mark
Department Head
Professor of Psychology

*Signatures are on file in the Graduate School.

ABSTRACT

One of the fundamental issues in bilingual language research has been determining how multiple language representations develop over the course of acquisition. In an effort to explore this, previous research adapted a traditionally monolingual statistical learning word segmentation task to provide input from two artificial languages, thereby simulating the early stages of bilingual acquisition. The goal of that study was to determine whether learners could form multiple representations for each input language and encapsulate information in each of the input speech streams, an ability that would facilitate learning in cases where statistical interactions between the languages might interfere with proper learning. The results demonstrated that adult learners can track two sets of statistics at once (i.e., encapsulating the information within each language), suggesting that they can form multiple representations when confronted with multiple language input. This process was facilitated by an indexical cue of speaker voice. In the absence of such an indexical cue, learners combined input across both languages, resulting in reduced learning when presented with statistically incompatible language pairs.

The present study examines whether the process of forming multiple representations in bilingual segmentation tasks can be facilitated by visual cues such as videos of faces speaking the artificial languages. Previous research has demonstrated that synchronous visual displays (such as faces) can facilitate infants' performance on a segmentation task. Here, two artificial languages with incompatible statistics were paired with videos of two synchronous, dynamic faces. With the faces serving as an indexical cue to language, participants learned each language significantly above chance, replicating previous results. Further experimental conditions demonstrate that this effect hinges on indexical synchronous information being available to the learner. Neither a static visual cue of color background nor a dynamic display of faces that produce both languages (and therefore not indexical) facilitated segmentation. These results suggest that faces are particularly effective indexical cues for facilitating the encapsulation of statistical information in a bilingual word segmentation task, underscoring the potentially important role for audio-visual synchrony in the course of normal language acquisition.

TABLE OF CONTENTS

List of Figures

Acknowledgments

I would like to thank my masters committee for their guidance in the preparation of this thesis. In particular, I would like to thank my adviser, Daniel Weiss, for his invaluable contributions to this project. Finally, I would like to give special thanks to Marissa Weyer and Beth Buerger, for having the patience and generosity to let me videotape them producing the artificial languages.

“The face is the mirror of the mind—and eyes without speaking confess the secrets of the heart”

~St. Jerome

INTRODUCTION

During the process of language acquisition, one of the earliest challenges that an infant faces is the identification of words from a continuous speech stream. The goal of the present study is to explore the mechanisms underlying language acquisition, specifically word segmentation. There are a number of possible cues to word boundaries that infants can utilize, including isolated word presentation, acoustic cues, and phonotactic cues. However, with the exception of isolated word presentations, which are rare (see below), none of these cues can provide a foothold into the segmentation problem for the completely uninitiated learner. However, one candidate cue that may provide an early foothold into language acquisition is statistical learning, tracking the transitional probabilities between speech elements (explained below). Of particular interest here is how statistical learning mechanisms operate in a bilingual context. Specifically, what types of information enable and maximize statistical learning in the context of multiple language input.

This paper begins by exploring the segmentation problem and discuss a number of the cues and strategies available to infants to solve the problem. This is followed by an overview of statistical learning and work that explores the hierarchy of segmentation cues. Following a discussion of recent work on bilingual statistical learning, a series of experiments are presented that are designed to explore the information necessary to maximize statistical learning in a bilingual paradigm.

The Segmentation Problem

While there is considerable debate regarding the relative contributions of nature and nurture in the development of mechanisms underlying language acquisition, learning the words of a particular language cannot be innate since babies are born into the world with the ability to acquire any human language. Since it is impossible for there to be *a priori* knowledge about the language to be acquired, there cannot be innate knowledge of words or word boundaries of any specific language. Infants are not pre-wired to speak English or Japanese or any other language—there must be a word learning mechanism in place that is flexible enough to accommodate any language or combination of languages.

Given that words must be learned, an early task that faces an infant is to identify word boundaries. The question then becomes, how does the infant segment the continuous speech stream signal into discrete units, or words? An intuitive solution to this question, known as the segmentation problem, is to propose that infants use the pauses, or silences between words to mark word boundaries. Adults perceive pauses in between words (Cole, Jakimik, & Cooper, 1980), and so this would be a logical way for infants to segment the speech input into words. However, pauses in speech rarely co-occur with word boundaries, and so cannot be used as a reliable cue to word segmentation (Klatt & Stevens, 1973; Reddy, 1976; Kooijman, Hagoort, & Cutler, 2005). For instance, consider the example of "where are the silences between words." As Figure 1 illustrates, the silences do not consistently occur at word boundaries. Thus, if infants are not using pauses to segment continuous speech, then what strategy are they using?

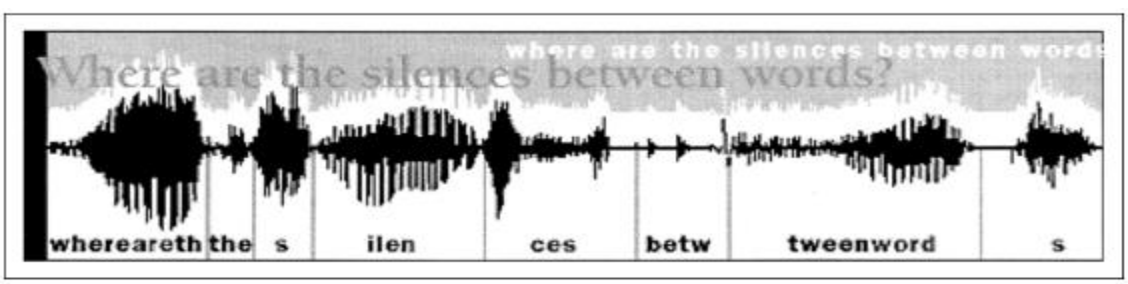


Figure 1. A waveform of the sentence "where are the silences between words," with markers illustrating breaks or silences. Source: Saffran (2003).

Possible Segmentation Strategies

Isolated word presentation. Several researchers have proposed that infants use isolated word presentation and distributional regularity to identify word boundaries (Brent & Cartwright, 1996; Dahan & Brent, 1999; Brent & Siskend, 2001). This theory is based largely on the INCDROP (incremental distributional regularity optimization) model of speech segmentation. The INCDROP model asserts that infants are able to segment speech through the recognition of familiar words or units (Dahan & Brent, 1999). There is evidence that distributional regularity, defined as the strategy of grouping sound patterns into words based on co-occurrence, can facilitate segmentation of speech streams (Brent & Cartwright, 1996). INCDROP advances the role of distributional regularity by postulating that once a word becomes familiar, it can be used to identify novel words in the sound stream. The example provided is that given the utterance "Lookhere!," if "Look" is familiar, then "Here" is inferred to be a new word (Dahan & Brent, 1999).

In a headturn preference procedure (in which longer looking times toward the test stimulus indicates a familiarity with the stimulus), six month-old infants looked significantly longer at words that followed a familiar word (Bortfeld, Morgan, Galinkoff, & Rathbun, 2005). During the familiarization task, the test item would follow either an easily recognizable item, such as "Mommy" or the participants own name, or a novel item such as another, unfamiliar name (e.g., "The girl rode Maggie's bike" and "A clown drank from Hannah's cup," where *bike* and *cup* are the test items; Bortfeld et al., 2005). The results of this study suggest that infants utilize familiar words as an "anchor" to identify novel words during speech segmentation. However, in this particular experiment, all test items immediately followed the familiar or unfamiliar name. The concern with this methodology is that a familiar item (particularly something as salient as one's name or "mommy") is going to draw attention to a greater extent than an unfamiliar item. This focusing of attention affects the items immediately following the attended item (c.f. Wood & Cowen, 1995), and so there would be greater attention for words following familiar items. This increased attention provides a plausible alternative explanation for the results reported in Bortfeld et al. (2005).

The chief criticism of isolated word presentation theories of segmentation, such as INCDROP (Brent & Cartwright, 1996), is that words are not presented in isolation often enough or as consistently as would be needed for such a strategy to be successful (Christiansen, Allen, & Seidenberg, 1998; Saffran, Newport, and Aslin, 1996; Aslin, Woodward, Lamendola, & Bever, 1996). Christiansen et al. (1998) argue that not only will many words never occur in isolation (such as articles and function words), but also the strategy is “underpowered” because of the continually increasing number of words the infant encounters. Aslin et al. (1996) report findings from several studies of infant-directed speech that while some mothers present words in isolation, others do not, even when provided explicit instructions to teach their children several target words. This bolsters the idea that isolated word presentation is inconsistent and insufficient as a segmentation strategy.

Acoustic cues. There is evidence that infants are able to extract information for language learning from the speech input (e.g. Werker & Tees, 1984). One type of information that infants can extract from the speech input are acoustic cues to word boundaries. There are a number of possible acoustic cues that could facilitate segmentation, including stress patterns (Houston, Jusczyk, Kuijpers, Coolen, & Cutler, 2000; Jusczyk, Houston, & Newsome, 1999; Houton, Santelman, & Jusczyk, 2004; Jusczyk, Cutler, & Redanz, 1993; Curtin, Mintz, & Christiansen, 2005; Nazzi, Iakimova, Bertoncini, Fredonie, & Alcantara, 2006), other prosodic cues (Mattys &

Juscyzk, 2001a; Christophe, Gout, Peperkamp, & Morgan, 2003), and phonotactic cues (Mattys & Jusczyk, 2001b; Mattys, Jusczyk, Luce, & Morgan, 1999).

Many languages follow a consistent stress pattern, such as French (Nazzi et al., 2006), Dutch, and English (Houston et al., 2000); thus, one of the prominent acoustic cues available to infants is the stress pattern of the language. In English, for example, most words follow a trochaic (strong/week) stress pattern (Houston et al., 2004). In an analysis of a large corpus of English (British) words, Cutler and Carter (1987) found that 90% of English words begin with a stressed syllable. Stress patterns provide a rich source of information for the infant to utilize in segmenting, and there is evidence that infants are sensitive to this information.

Jusczyk et al. (1993), using a preferential headturn procedure, found that by nine months, English learning infants prefer words that follow a strong/weak stress pattern (e.g. *donor*) to those that follow a weak/strong pattern (e.g. *define*); however, 6 month old infants did not exhibit such a preference, suggesting that this preference develops between six and nine months of age. Infants are also able to extract words from a sentence using stress patterns. In a preferential looking time task, 7 and a half month-old infants were able to extract strong/weak words from a sentence, but not weak/strong (Jusczyk et al., 1999). Infants were familiarized to a pair of words, and were then tested on sentences that contained either familiar or unfamiliar target words. Seven and a half month-old infants listened longer to familiar items only if the target word followed the strong/weak stress pattern (Jusczyk et al., 1999). This suggests that infants use stress patterns to segment the speech stream into words. Additionally, 10.5 month-old infants were able to extract weak/strong words from the sentence, indicating that using stress patterns to segment may be important initially, but as a larger corpus of words is acquired, other segmentation strategies are used (Jusczyk et al., 1999).

Using the same procedure, Houston et al. (2000) found that infants can segment words in an unfamiliar language using stress patterns. Nine month-old English-learning infants were able to successfully extract Dutch words, since despite the stark phonetic differences between the two languages, both Dutch and English follow a trochaic stress pattern. Likewise, Nazzi et al. (2006) found that French-learning 12 month-old infants utilized prosodic stress information for segmentation in a syllabic language (French). These studies suggest that infants are able to use the stress patterns of a language as a cue to word boundaries, facilitating word segmentation.

There are other potential acoustic cues to word boundaries that infants are sensitive to. Prosodic cues other than stress patterns, such as allophones (predictable variation within a particular phoneme; e.g. the aspiration for /p/ in word initial positions as in the aspirated *pit*, /p^hIt/, and the non-aspirated *spit*, /spIt /), are utilized by infants. Jusczyk, Hohne, and Baumann (1999) found that by 10.5 months, infants are sensitive to allophonic variations and that they use this information to segment words in sentential contexts. Mattys and Jusczyk (2001a) argue that prosodic cues allowed 8.5 month-old infants to correctly segment words in a sentential context, such as *ice* from "The city truck cleared ice and sand from the sidewalk," rather than *dice*, which was also phonemically present. Word initial vowels are often glottalized, and so the perception of glottalization informs the infant to segment the word *ice* instead of *dice* (Mattys & Jusczyk, 2001a).

In addition to prosodic cues, infants are able to use phonotactic cues to segment a speech stream (Mattys & Jusczyk, 2001b; Chambers, Onishi, Fisher, 2003). Phonotactic cues refer to the use of rule governed sound sequences within a language (e.g. the sound "ng" can occur at the end of words in English, as in *running*, but never occurs at the beginning of a word) Mattys and Jusczyk (2001b) familiarized 9 month-old infants to passages that either had good or poor phonotactic cues. The "goodness" of the phonotactic cues was determined by the consonant to consonant (C-C) clusters surrounding the target item (i.e. …C-**CVC**-C…). If the target item was bordered by C-C clusters that frequently occur between words in English, then the target item had good phonotactic cues. If the target item was bordered by C-C clusters that frequently occur within words in English, then the target item had poor phonotactic cues (Mattys & Jusczyk, 2001b). In a headturn preference procedure, infants listened longer to target items that had good phonotactic cues than to those that had poor phonotactic cues, suggesting that infants are able to use phonotactic information in speech to facilitate word segmentation.

Although infants are able to use several different acoustic cues in speech to identify word boundaries, acoustic information alone cannot account for word segmentation. The shortcoming of theories that stress the role of acoustic information in segmentation is that there are no invariant acoustic cues across all languages (Cole and Jakimik, 1980; Klatt, 1979). For example, while stress patterns provide a reliable cue to word boundaries in English, not all languages follow a stress pattern that reliably aligns with word boundary (e.g., Spanish). Since not all languages can rely on stress patterns (or other acoustic cues), there must be some other mechanism in place. This does not preclude the possibility that on some level infants use acoustic information, or even a combination of acoustic cues (see Christiansen, Allen, & Seidenberg, 1998), to identify word boundaries. Indeed, word learning likely relies a number of different cues (Gerken, 2002). However, the acoustic and lexical cues discussed above are insufficient for learning word boundaries due to their inconsistency across languages. There is one cue, though, that is invariant across all languages—statistics. Using the statistical structure and patterns of a language is a segmentation strategy that is viable for any language, and since statistical learning mechanisms are available early in development (Saffran, Aslin, & Newport, 1996), they may provide a foundation upon which the other segmentation cues build.

Statistical learning.

One possible cue to word boundaries in continuous speech comes from the statistical dependencies in the language. If an infant can keep track of correlations between the sounds of the language, it is possible to use this information to parse the input into words. The classic example of this is the phrase *pretty baby*. Because there are few words in English that begin with the syllable *pre*, the probability that *pre* will be followed by the syllable *ty* is relatively high. However, *ty*, occurring in the word final position, can be followed by a nearly limitless set of word-initial syllables (e.g. *pretty baby, pretty chair, pretty eyes*, etc.), and so the probability that *ty* will be followed by *ba* is relatively miniscule(Saffran, 2003).

These probabilities are known as *transitional probabilities*. Transitional probabilities are a conditional probability statistic between successive syllables (Aslin, Saffran, & Newport, 1998). The formula for computing transitional probabilities (here between two successive syllables, *X* and *Y*) is:

$$Y|X = [\text{frequency of XY}] / [\text{frequency of X}]$$

What this formula states is that the probability of Y occurring, given the occurrence of X, is equal to the frequency of co-occurrence of X and Y divided by the overall frequency of X. In other words, transitional probability can be seen as a relative frequency of co-occurrence; the chief difference is that transitional probabilities account for the overall frequency of the word, while a simple co-occurrence probability does not (Aslin, Saffran, & Newport, 1998).

To illustrate how transitional probabilities work, consider the example given earlier: *pretty baby*. To compute the transitional probability of *pretty*, one would take how often the sound *pre* is followed by the sound *ty* in English, and then divide that number by how often the sound *pre* occurs in English. Inserting these syllables, the formula above looks like this:

ty | *pre* = [frequency of *pre.ty*] / [frequency of *pre*]

Likewise, the transitional probability for *ty.bay* would be:

Bay | *ty* = [frequency of *bay.ty*] / [frequency of *ty*]

The difference in transitional probabilities between *pre-ty* and *ty-ba* (approximately 80% and 3%, respectively, in speech to infants) allows the infant to correctly segment the phrase *pretty baby* (Saffran, 2003). Word boundaries are marked by low transitional probabilities, and so by attending to statistical patterns and regularities, a non-acoustic cue in the input, infants are able to successfully segment a continuous speech stream into words. Indeed, there is evidence that both adults (Saffran, Newport, & Aslin, 1996) and infants (Saffran, Aslin, & Newport, 1996) are able to use these transitional probabilities to segment words from fluid speech streams. Additionally, this mechanism is both domain general (Saffran, Johnson, Aslin, & Newport, 1999; Creel, Newport, & Aslin, 2004; Kirkham, Slemmer, Johnson, 2002; Kuhn & Dienes, 2005; Fiser & Aslin, 2002; Trk-Browne, Junge, & Scholl, 2005) and species general (Hauser, Newport, & Aslin, 2001; Toro & Trobalon, 2005).

Adults. To this point, the discussion has focused on infants' abilities to use information in the speech stream to segment languages. While this population is important for studying the early mechanisms of language acquisition, there are several reasons why it is useful and important to study statistical learning in adults. First, it is necessary to establish the "endpoint" of development. Any trend found in infants is difficult to interpret without knowledge of the adult state of these abilities. If there are types of statistical learning that only infants can perform, the implications would drastically differ relative to a scenario where there are no changes across development. Related to this point is that, unlike other aspects of language acquisition (e.g. phonetic discrimination, see Werker & Tees, 1984), statistical learning abilities do, in fact, appear to be relatively stable across development (Newport, Weiss, Wannacott, & Aslin, in prep.; Newport & Aslin, 2004) and therefore testing adults affects the expectations for what we should expect to find in infancy. The final reason for initially testing adults is practical: it is easier to test adult subjects relative to infants and the results are often easier to interpret.

Saffran, Newport, and Aslin (1996) found that adults are able to use transitional probabilities to learn artificial languages. They tested adults in a statistical learning paradigm, wherein participants are exposed to an artificial language that has been stripped of any segmentation cue other than transitional probabilities and are then tested on their ability to segment the artificial language into words. The only way to successfully segment words from the continuous speech stream in a statistical learning paradigm is to keep track of the transitional probabilities between the sounds and use this information to identify word boundaries.

In this particular statistical learning task, participants were exposed to the artificial language of six trisyllabic words (*babupu, bupada, dutaba, patubi, pidabu,* and *tutibu*) via a continuous synthesized speech stream. All pauses and acoustic word boundary cues were removed. The transitional probabilities within the word ranged from 0.31 to 1.00, but were always higher than the transitional probabilities between words (ranging from 0.1 to 0.2). Participants listened to the synthesized artificial language for 21 minutes, and were then given a two-alternative, forced-choice task between either words and non-words or words and part-words. Part-words consist of the final syllable of one word and the first two syllables of another word (e.g. *pubupa* from *babupu* and *bupada*), thus differing from the word by only one syllable. If participants were able to segment the speech stream into words, then they should identify words over part-words at a better-than-chance rate. Indeed, participants correctly identified the words significantly above chance in both the non-word condition and the part-word condition (see Figure 2). Because they were able to segment the speech stream when the only cues to segmentation were transitional probabilities, this suggests that adults are able to keep track of the statistical regularities in the speech stream and are able to use this information to correctly segment artificial languages (Saffran, Newport, & Aslin, 1996).

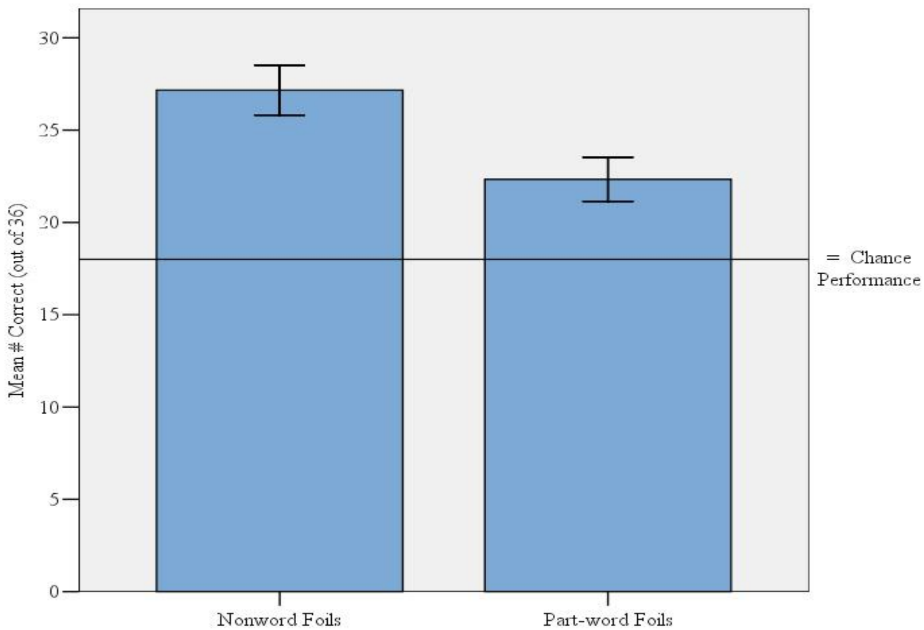


Figure 2. Mean test scores (out of a possible 36) for the non-word foil and the part-word foil conditions. Error bars represent +/- 1.00 Standard Error. Source: Saffran, Newport, & Aslin (1996).

Infants. Given that adults are able to use conditional statistics to parse a fluid speech stream, the question still remains; can infants in the process of acquiring their first language use statistical cues to identify word boundaries? Saffran, Aslin, & Newport (1996) replicated their earlier findings from their adult study with 8-month-old infants, using a headturn preference procedure (see Jusczyk & Aslin, 1995). In this task, infants listened to a two-minute familiarization passage of artificial language (again, where the only cues to word boundaries were the transitional probabilities). Because the participants were infants, the artificial language was simplified. This was done by reducing the number of words from six to four and by adjusting the transitional probabilities such that within words the transitional probability was always 1.0, and between words it was always .33 (no word was repeated, so each word could only be followed by one of the three other words).

Following the familiarization passage, infants were tested on four tri-syllabic strings (two words, two part-words). At test, a tri-syllabic string was repeatedly presented on either the left or right of the infant, accompanied by a red, flashing light. The critical value is looking time, which is measured from the onset of the infant orienting towards the flashing light until the infant looked away for longer than two seconds. In this way, looking time was regarded as a measure of preference, with the hypothesis that infants should be more interested in novel utterances. Thus, if the infants are learning the words of the artificial language, then they should display a preference for the novel part-words. The results of Saffran, Aslin, & Newport (1996) support this hypothesis, as infants showed a preference for the novel utterances, listening longer to both non-words and part-words than to statistically-defined words. This indicates that infants are not only able to segment words from speech, but also they are able to use statistical cues to do so. Since these were fairly young infants (8-months-old), this suggests that statistical learning is available early in development and that this may be an initial segmentation strategy.

Domain general. There is evidence that statistical learning is domain-general; that is, the ability to use statistical regularities for parsing is not limited to language. Statistical learning has been shown in a wide number of domains, including musical tones (Saffran et al, 1999; Creel et al., 2004) and visual patterns (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Turk-Browne, Junge, & Scholl, 2005). In these studies, participants were able to segment "words" from continuous auditory or visual streams, suggesting that statistical learning is not limited to the linguistic domain. Statistical learning has also been shown in other species such as rats (Toro & Trobalon, 2005) and non-human primates (Hauser et al., 2001). Since these species do not have language, it can be inferred that statistical learning is domain general.

Competition in Statistical Learning

The studies described above have examined statistical learning abilities in the absence of any other cues. These reductionistic studies have established that statistical learning is a tool that infants and adults can use to segment fluid speech into words. However, in natural language acquisition, there are multiple sources of information for segmentation (e.g. stress, coarticulation, phonotactics, and isolated word presentation). These multiple sources of information are not always consistent with each other, and so competition arises. How is this competition resolved? In the next section, two sources of competition are reviewed. The first arises when multiple cues to segmentation are present. The second is when multiple languages with incongruent statistical patterns are present.

Statistical learning vs. acoustic cues. When artificial languages are stripped of all cues to segmentation other than transitional probabilities, it is possible to use statistical learning mechanisms to learn word boundaries. However, natural languages contain cues to segmentation other than transitional probabilities. They are not stripped of all acoustic word boundaries and any possible phonotactic regularities, as they are in statistical learning experiments. A number of possible cues to segmentation were discussed earlier, and now the question arises: how do these cues interact with statistical information?

The first question is whether or not each cue is given equal weight. If multiple speech segmentation cues are available, how are these cues integrated into a single representation to identify word boundaries? Recent work that pitted these cues against each other suggests that these cues are integrated into a hierarchical framework (Mattys, White, & Melhorn, 2005). In order to pit prosodic cues (stress) against lexical level (word-context) and segmental level cues (phonotactics), Mattys et al. (2005) used tri-syllabic English words that either followed the dominant stress pattern of Englsih (strong-weak; e.g. *marathon*) or did not follow the stress pattern (weak-strong; e.g. *material*). Segmental (phonotactics) and lexical (word context) cues to word segmentation all were followed instead of the prosodic cue of stress, as congruent targets showed greater priming than incongruent targets, regardless of the stress pattern of the prime (Mattys et al., 2005). Additionally, lexical cues were pitted against segmental cues, and in this condition, lexical cues were used instead of segmental cues. This suggests that there is a hierarchical structure for segmentation cues, and that the order of dominance is lexical cues, followed by segmental cues, and finally prosodic and acoustic cues (Mattys et al., 2005). However, when background noise was added to the stimuli, participants followed the stress cue and not the other higher-order cues. Thus, it is possible that context and condition play a role, and that the hierarchy varies dynamically.

While statistical learning would be considered lexical level information, Mattys et al. (2005) did not directly examine the relative contributions of statistical learning and other segmentation cues. The first study to directly pit statistical learning against other cues was conducted by Johnson and Jusczyk (2001). Using the same language as Saffran, Aslin, and Newport (1996), they examined the relative contribution of multiple cues (stress and co-articulation) to the segmentation of an artificial language by 8-month old infants. In Experiment 2, they made the final syllable of the statistically defined words stressed, and in the stress pattern of English a stressed syllable suggests the beginning of a word. Thus, stress cues indicate different word boundaries, and actually line up with the part words (e.g. *tibudogolaTUdaropitibudodaroPIgolatu*, where upper case syllables are stressed; in this example, statistically-defined words are *golaTU* and *daroPI*, whereas stress-defined words are *TUdaro* and *Pigola*). Infants were tested using a headturn preference procedure.

If the infants used statistical learning, then they should listen longer to novel part-words (it is important to note that in this condition, the 3-1-2 part-words actually correspond to the stress-defined words, thereby providing a direct test between statistics and stress). However, if the infants followed the stress patterns, then they should listen longer to the words. Looking times were significantly longer for words than part-words, indicating that stress cues outweigh statistical cues (Johnson & Jusczyk, 2001). Additionally, in Experiment 3 they used a similar paradigm to test between statistics and co-articulation, finding that co-articulation also outweighed statistics.

The findings of Johnson and Jusczyk (2001) suggest that statistics are not necessarily a primary segmentation cue. However, as was mentioned earlier, it is possible that statistical learning is an initial strategy that is then built upon by subsequent strategies, such as stress patterns. Thiessen and Saffran (2003) also used competing cues in a statistical learning paradigm, only comparing the performance of 7- and 9-month-old infants. Like Johnson and Jusczyk (2001), Thiessen and Saffran (2003) found that 9-month-old infants followed stress patterns instead of statistical cues. However, unlike the previous study, they found that 7-month-old infants followed statistics over stress patterns. This suggests that statistical learning is an initial strategy to word segmentation, and that as a knowledge base begins to form, older infants switch strategies and begin to follow other acoustic and segmental cues. It also confirms that, as

Mattys et al. (2005) argue, segmentation cues are organized in a hierarchical structure, and that this structure is flexible, such that cues can move up or down in hierarchy position.

Recent research has explored the role of salience in these "colliding cues" experiments that pit several cues or strategies against each other. Weiss, Gerfen, Mitchel, and Rizzo (in prep) used a paradigm similar to that of Johnson and Jusczyk (2001) and Thiessen and Saffran (2003) to pit pauses against statistics. However, unlike previous studies, Weiss et al. (in prep) manipulated the salience of the acoustic cue by varying the length of pauses. In Experiment 1, pauses (25ms and 50 ms) were inserted at statistically-defined word boundaries, and so were consistent with the statistical information, such that both parsed the speech stream in the same place. As pause length increased, performance significantly improved (see Figure 3). This suggests that pauses are relevant cues to segmentation and thus it is possible to pit pauses against statistics. It also suggests that consistent cues are additive in nature (see also, Merkx & Monaghan, 2006).

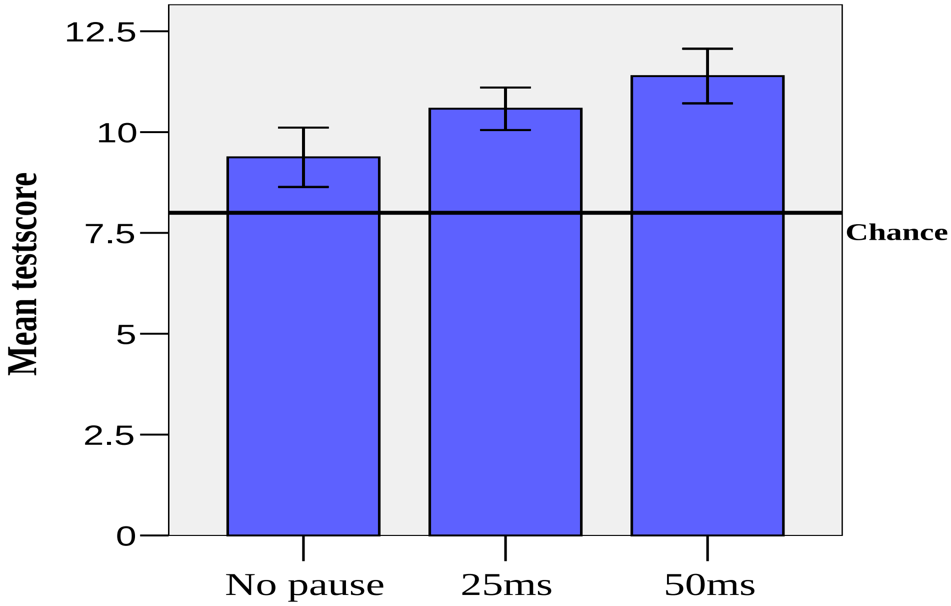


Figure 3. Mean number of words learned by pause length in compatible conditions.

In Experiment 2, pauses of varying length (25ms, 50ms, and 75ms) were inserted in between the second and third syllable of statistically-defined words (e.g. *bu.ti. # gu.,* where # represents a pause). Thus, pauses were inconsistent with statistics. It is important to note that in this condition, the pause-defined words were consistent with the part-word test items. Thus, if a participant were to follow pauses, then at test she/he would correctly select the part-words, and so in this experiment performance is determined as the distance from chance in either direction since a high score represents successful statistical learning and a low score represents successful segmentation using pauses. There was a significant main effect of pause salience, as participants in the 25ms pause condition followed statistics, participants in the 50 ms pause condition performed right at chance, and participants in the 75ms condition followed pauses (see Figure 4). This suggests that as the salience of a segmentation cue increases, participants are more likely to utilize this cue. By this logic, pauses and statistics were approximately equally salient in the 50ms condition, and this seems likely as participants learned both the pause-defined words and the statistically-defined words.

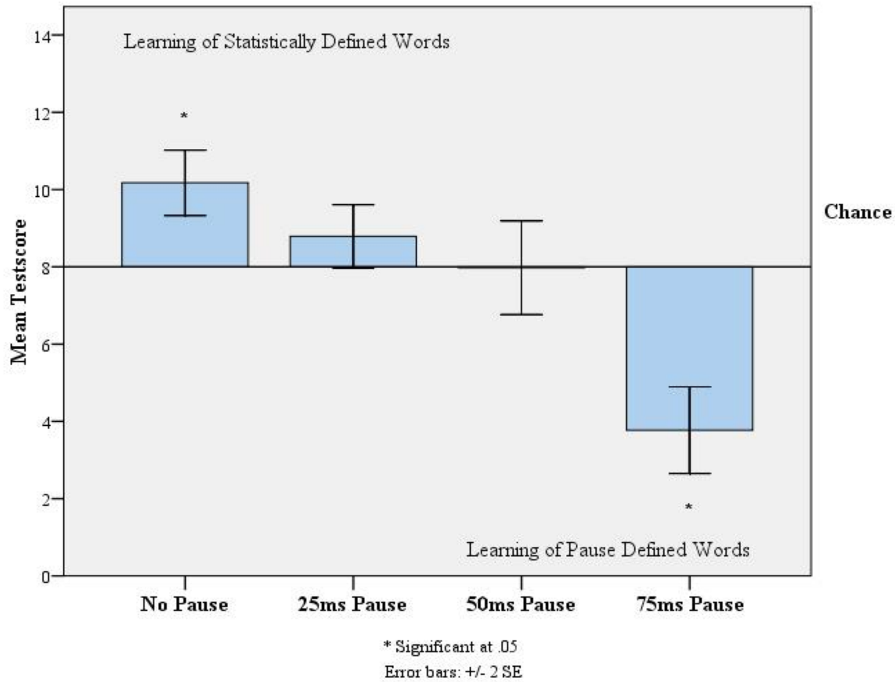


Figure 4. Mean number of statistically-defined or pause-defined words by pause length in incompatible conditions.

An additional finding of this study was that the Simon Effect, a standard measure of inhibitory control (the ability to suppress or ignore irrelevant information) negatively correlated (the Simon Effect is measured such that a low Simon score represents greater inhibitory control) with performance in the 50ms condition, but this correlation was not present in the other conditions. This seems to suggest that inhibitory control plays an important role when the cues are equally salient, requiring the participant to ignore or suppress one of the cues. Since inhibitory control is considered a general cognitive mechanism, it is interesting to find that it mediates segmentation as language acquisition and general cognitive functions are often thought to be distinct processes. The findings of this study indicate that when multiple cues to segmentation are present, cue salience and inhibitory control mechanisms play a role in resolving cue competition.

Bilingual statistical learning. The aforementioned studies strongly suggest a role for statistical learning in early word segmentation. However, these studies have focused on monolingual acquisition. The second type of competition in statistical learning comes from the presence of more than one language. Given that a considerable proportion of the world's population is bilingual or multilingual (Crystal, 1997), what role does statistical learning play in simultaneous bilingual language acquisition? Can language learners keep track of the statistical regularities in multiple languages?

One of the prominent questions of bilingual language acquisition is determining the point at which the infant realizes that multiple languages are being spoken. The Unitary Language System hypothesis (Volterra & Taeschner, 1978) states that, initially, bilingual children progress through two stages of development in which they are unable to separate the two languages. It is not until a third and final stage (between 2 and 3 years of age) that children are able to separate the languages into distinct representations. The Unitary Language System hypothesis arises from concern over bilingual children's language mixing, or code switching. Bilingual children often mix languages, using both languages seemingly interchangeably in the same sentence (Genesee, Nicoladis, & Paradis, 1995). This was seen as a sign of imperfect language learning, leading to the argument that bilingual children were unable to differentiate or separate the two languages (Taeschner, 1983).

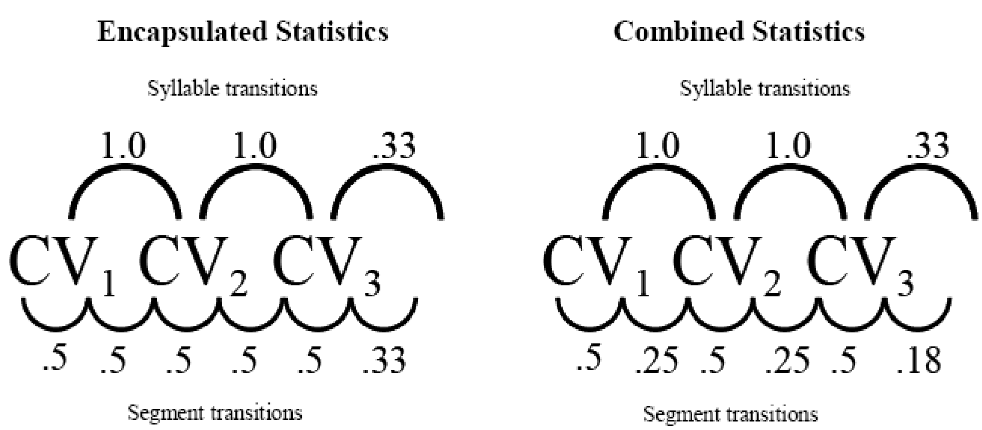
However, the preponderance of evidence supports a Differentiated Language System hypothesis where each language has a unique and separate representation from at least the stage of first words or earlier (Döpke, 1996; Genesee, 2000). For example, the Unitary Language Hypothesis (Taeschner & Volterra, 1978) predicts that children are biased to interpret novel utterances as new words (Junker & Stockman, 2002), and so learning cross-language synonyms, or translation equivalents, should be impossible. However, Junker and Stockman (2002) found no such performance deficits among bilingual-learning toddlers for translation equivalents. This was taken as evidence of an ability to separate two languages at an early stage in development. In addition, much of the evidence for a unitary language system comes from studies showing cross-linguistic influences; however, these influences may reflect structural similarities in the languages (Döpke, 1996) rather than due to an inability to separate the two languages (Muller & Hulk, 2001). Moreover, language mixing by children is not evidence for an inability to differentiate or separate the two language systems, as language mixing in children resembles adult language mixing (when, according to the Unitary Language Systems hypothesis, languages are differentiated) and is rule-governed (Genesee et al., 1995; Genesee, 1989; Garcia, 1980). Because code mixing takes on a structured, adult-like form, it is not likely the product of unsystematic mixing as a product of undifferentiated languages. Thus, there is little support for the Unified Language Hypothesis (Taeschner & Volterra, 1978), and there is considerable evidence that children maintain separate and distinct representations for each language (Genesee, 2000).

Maintaining distinct representations for each language during acquisition presupposes that, on some level, the infant is aware there are two languages in the input. At what point does this realization occur? Recent research indicates that children are able to discriminate not only between two languages of different rhythmic class (Nazzi, Jusczyk, & Johnson, 2000) at a very early age, but also within rhythmic class as early as four months of age (Bosch & Sebastian-

Galles, 2001). In a modified familiarization-preference procedure that measured looking time, 4-month-old bilingual Catalan/Spanish-learning infants looked longer at test items from the non-familiarization language (e.g. if the familiarization was in Spanish, then infants would look longer at Catalan test items than Spanish test items, regardless of maternal language). Furthermore, the patterns of results resembled performance of monolingual infants, where the non-familiarization language was a novel language. These results suggest that 4-month-old bilingual-learning infants are able to discriminate between languages (Bosch & Sebastian-Galles, 2001). However, the ability to discriminate between languages does not mean that separate representations are being formed and maintained as this could be the result of purely acoustic processes (as evidenced by animal studies, Toro & Trobalon, 2005; Hauser et al., 2001). Maintaining separate statistics for two languages, on the other hand, does require distinct representations. Thus, if it is possible to use statistical learning in a bilingual environment, then this would provide evidence that bilingual-learning infants are able to maintain separate, distinct representations.

To test whether it is possible to use statistical learning as a segmentation cue in simultaneous bilingualism, Weiss and Gerfen (2006) modified the statistical learning paradigm to incorporate two artificial languages (hereafter referred to as the bilingual statistical learning paradigm). The addition of a second language may present a problem for learners because the statistical regularities may differ across languages. For example, the sound *zeh*, as in the word *zealous*, in Hebrew often corresponds to the word meaning *this* (Weiss & Gerfen, 2006). Since *zeh* marks a word boundary in Hebrew, it has a low transitional probability. However, in English *zeh* occurs within words, and so the statistical regularities of these two languages are not compatible (at least in this particular instance). One potential solution to this problem is to maintain separate sets of statistical computations for each language.

The bilingual statistical learning paradigm follows the basic structure of the paradigms used in previous statistical learning studies (e.g., Saffran, Aslin, Newport, 1996). However, rather than one language, Weiss and Gerfen (2006) developed a second artificial language (L2) that was either congruent or incongruent with the first artificial language (L1). The languages were congruent if the transitional probabilities were invariant irrespective of whether the listener maintained separate statistics for each language or combined statistics across languages. This was done by having the final segment (and so in these languages, a vowel) be the same across languages and by not having any syllable repeated. Thus, the within word transitional probabilities (at the syllable level) remained at 1.0 and the between-word transitional probabilities remained at .33 (see Figures 5a and 5b). It should be noted that the segmental statistics, while variant depending on whether or not the participant combines statistics across languages, still upholds the higher transitional probabilities within words with the lowest transitional probabilities marking word boundaries.



Figures 5. This shows the transitional probabilities of the congruent languages used in Weiss and Gerfen (2006). Reproduced from Weiss and Gerfen (2006).

The languages were incongruent, however, if the transitional probabilities became less reliable (the transitions remain low at word boundaries, but the noise level is raised) if combined across languages (see Figure 6). The statistics in this condition would only provide a reliable, strong cue to word boundary if the participants were able to maintain separate statistics for each language. The segmental statistics also become noisier if participants combine statistics across language in the incongruent condition.

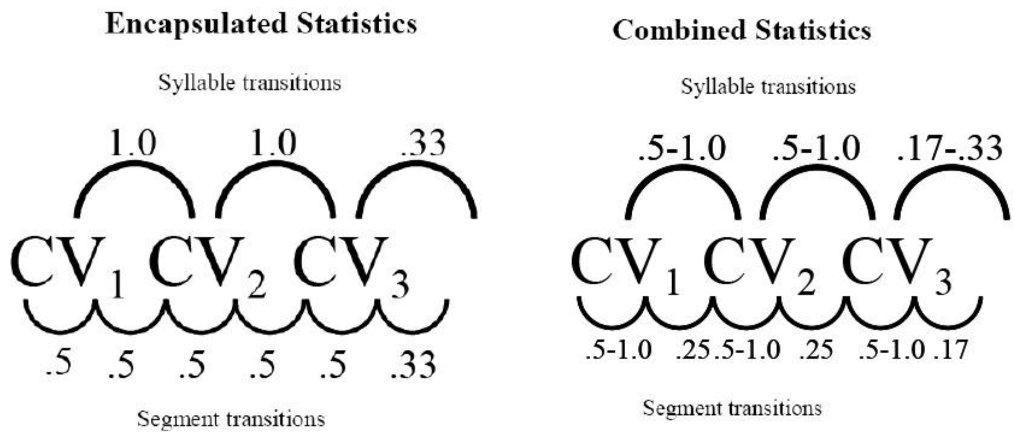


Figure 6. This shows the transitional probabilities for the incongruent languages from Weiss and Gerfen (2006). Note that the combined transitional probabilities change from word to word for both syllables and segments, creating a much noisier statistical environment. Source: Weiss and Gerfen (2006).

In the congruent language conditions, participants were able to successfully segment the statistically defined words in both languages (Weiss & Gerfen, 2006). In the incongruent condition, on the other hand, participants did not perform above chance, indicating that they were unable to segment the speech stream using statistics. This suggests that participants did not compute and maintain separate statistical representations for each language. However, when an indexical cue to language (voicing—L1 was in a male voice, and L2 was in a female voice) was incorporated into the audio stream, participants performed above chance. This provides evidence that participants are indeed able to compute separate statistics as long as they are able to distinguish that there are two languages in the input. Given the indexical cue of voice, performance is significantly better than when there are no such cues (Weiss & Gerfen, 2006). This raises the central question of the present study. That is, given that a voicing indexical cue facilitates segmentation in this bilingual statistical learning paradigm, what other indexical cues facilitate and maximize performance in this bilingual paradigm?

The Present Study

The goal of this thesis is to determine the relevant features underlying indexical cues. The speaker voice cue used in a previous experiment was found to facilitate segmentation in a bilingual statistical learning paradigm (Weiss & Gerfen, 2006). In thinking about this problem from the perspective of the language learner, there are a host of potential indexical cues. For example, there may be environmental cues such as where the language is heard (e.g., English at home, French at school), language-specific phonetic cues such as pitch or stress patterns, and/or phonotactic cues. However, not all of these cues may be reliable, and the process of cue selection is unknown. It is likely that the language learner may selectively attend to some of these cues (e.g. speaker voice) while ignoring other cues (e.g. environmental cues). Thus, it is important to establish which of these potential cues facilitate the formation of multiple representations and what dimensions allow them to be effective. In the subsequent experiments, indexical cues in the visual domain were considered. The importance of visual information in language acquisition, and more specifically, auditory speech perception, has been well documented (e.g. see Hollich, Newman, & Jusczyk, 2005). Infants are sensitive to the relations between visual and auditory information (Kuhl & Meltzoff, 1982), and so it is possible that visual information may serve as an indexical cue in bilingual language acquisition environments.

Indexical cues can be defined as any cue that provides systematic information about which language is being presented. In a bilingual statistical learning paradigm, there are two artificial languages, and an indexical cue would be anything that would indicate to the participant whether Language 1 or Language 2 is being presented. In Weiss and Gerfen (2006), the indexical cue was speaker voice, an auditory cue to language. The present studies tested two potential visual indexical cues: background screen color and the dynamic display of faces.

EXPERIMENT 1A: Background Color

What are the indexical cues that maximize segmentation of a continuous speech stream consisting of two artificial languages? To explore this question, Experiment 1 uesd the bilingual statistical learning paradigm used in Weiss and Gerfen (2006) and manipulated the indexical cue to language, removing the auditory cue of speaker voice and using instead a visual cue of background display color.

Methods

Participants

Thirty-six undergraduate introductory psychology students participated for class credit. I excluded from analysis any participant who reported themselves as being bilingual (2), had seven or more years of second language experience (9), if there was technical failure (2), or if the participant failed to follow instructions (1). This brought the number of participants included in analyses to 22. There were 10 men and 12 women, and all were monolingual English speakers.

Materials

The familiarization stimuli were two artificial languages. The artificial languages were those used in Weiss and Gerfen (2006). Each language consisted of four trisyllabic words, with a CV.CV.CV. structure (see Figure 6). The CV syllables were created by digitally recording a female speaker producing CVC syllables. The CVC was then hand edited in Praat©, removing the coda consonants and controlling for duration to create the CV syllable. The benefit of creating the CV syllables in this manner is that it preserves the vowel to consonant transitions when the CV syllables are combined to create the words. The syllables are also recorded without a coda consonant in order to create the test items. All words were normalized using SoundForge©, controlling the loudness and pitch contours of the words, and then resynthesized using Praat©. The resythesized words were concatenated in random order into continuous speech streams, with each word presented the same number of times. The only cues to word boundaries were the transitional probabilities.

Since the goal of the present study is to examine indexical cues that facilitate segmentation, the artificial languages were constructed to be incompatible, as in Weiss and Gerfen (2006). Languages are incompatible if the statistics become noisier and less reliable when combined across languages. Within individual languages, each word had a 1.0 internal transitional probability and a 0.33 transitional probability at word boundaries. However, when the languages are combined the transitional probabilities dip within the word, and so the transitional probabilities are less reliable cues to segmentation (see Figure 6). To make the languages incompatible, a word-final syllable from language 1 is inserted word-initially in language 2. Thus, the same sound has a low transitional probability in one language and a high transitional probability in the other, providing conflicting cues to word boundary. The key feature of these languages is that each individual language has reliable transitional probabilities, but if the statistics are combined then there is no reliable cue to word boundary. Therefore, in order to successfully segment these languages, participants must maintain separate statistical representations for each language.

In addition to the statistics, Experiment 1A included a visual indexical cue. The visual cue consists of the color of a blank background screen during the familiarization presentation, changing between purple and teal. This indexical cue marks language boundaries, such that each screen color was paired with one language. During L1 presentation, the background color was purple, and during L2 presentation the background color was teal. In this way, there is a visual cue to language that can enhance the participant's ability to maintain separate statistical representations.

The test stimuli included statistically-defined words and part words from the familiarization stream. The part words consisted of the second and third syllable of a statistically defined word and the first syllable of another statistically defined word; thus, participants have heard the sequence of syllables before. This removes the possibility that recognition of statistically-defined words is due to exposure rather than successful word segmentation. If the test items were to consist of novel non-words, then the participant would be more likely to identify the word simply because the participant had heard it before. However, the part-words occur in the familiarization passage, and so learning cannot be attributed to exposure, but instead must be the result of segmentation through statistical learning.

The test was comprised of each word, four from each language. The words were paired with two part words from the same language, and each pairing was presented twice, counterbalancing the order of presentation to account for order effects. The total number of test trials was 32, with 16 trials from each language (therefore chance is defined as 16 overall, and 8 for each individual language).

Procedure

The experiment was presented using E-Prime software (Psychology Software Tools, 2002) on Dell PC computers (Optiplex GX280) with Pentium IV processors. Participants listened to L1 for 1 minute and 56 seconds, followed by 1 minute, 56 seconds of L2. This is repeated to create a block of presentation lasting 7 minutes and 44 seconds. This block is repeated three times with a one-minute break in between each block, for a total of 23 minutes and 12 seconds of familiarization. Each word was presented an equal number of times, and no word was presented twice in succession. While the familiarization audio stream was played, the color of the screen would change when the language changed. During L1 presentation the background screen was purple. During L2 presentation, the screen was teal.

Instructions, given both on the screen and verbally by the experimenter, did not mention the screen color. Instead, participants were simply told that they would hear an audio stream and then would be tested on information garnered from the audio stream. Verbal instructions were read from a script in order to standardize them across experimenters.

During the test phase, participants would listen to the test items (a pair of trisyllabic items, a word and a part-word). For each pair of test items, the first item would be presented, followed by a one second pause, and then the second test item would be presented. Between each pair of test items there was a four second pause. Test items were presented in a counterbalanced order, with the word occurring equally often in the first position and the second position. Participants were then asked to identify which item was the word based the familiarization audio stream by pressing the corresponding key on the keyboard. Response, response time, and response accuracy were recorded. After the test, participants were given a paper questionnaire (see Appendix B) about language use, language background, and effort (a self report measure of how "hard" they tried to answer questions and complete the experiment).

Results

Overall, participants successfully segmented words from the continuous speech stream. The mean number of words learned overall was 17.77 out of a possible 32, with a standard deviation of 3.22 (see Figure 7). The mean number of words learned for L1 was 10.23 (SD = 2.37) and for L2 was 7.55 (SD = 2.26). Performance was significantly above chance (chance is defined here as a score of 16) overall, $t(21) = 2.58$, $p = .017$. However, this effect appears to be driven by performance in only one of the languages. A paired-sample t-test revealed that there was a significant difference in performance between languages, $t(21) = 3.78$, $p = .001$. When performance is viewed at the level of the individual languages, participants learn L1 ($t(21) = 4.41$, $p < .001$) but not L2 ($t(21) = -0.94$, $p = .357$). This suggests that performance is asymmetrical across languages, with participants showing greater performance in L1 than L2. The reasons for this are discussed below.

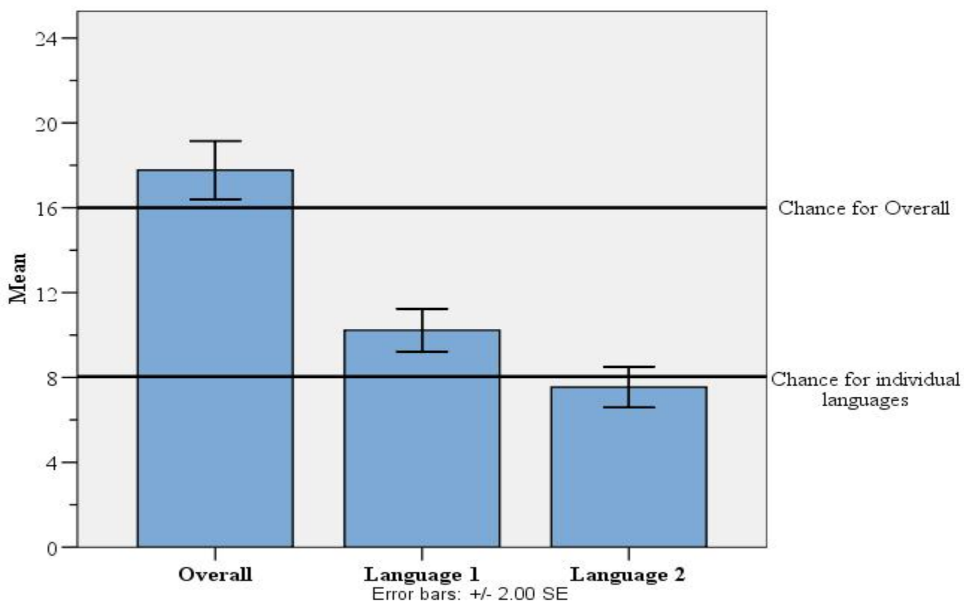


Figure 7. Mean number of words learned in Experiment 1A, overall and for each language.

Discussion

In Experiment 1A, the mean number of words learned was significantly above chance. However, this result may be the product of an order effect, as evidenced by the differential learning pattern for L1 and L2 in Experiment 1A. It is hypothesized that this asymmetry is, in essence, a primacy effect, in that L1 was always presented first and then followed by L2. L1 was presented for a block of two minutes, and then L2 for two minutes, and this blocking pattern repeated until the passage was completed. The finding of asymmetry in learning in Experiment 1A might suggest that participants were able to learn L1 in a short amount of time, and then tended to ignore the subsequent information, thus exhibiting learning for L1 (although to a lesser degree than with a longer block, see Weiss & Gerfen, 2006). Experiment 1B was designed to test this hypothesis. If the results of Experiment 1A were the result of a primacy, or order, effect, then reversing the order of presentation should yield opposite results.

EXPERIMENT 1B: Background color, L2 initial presentation

In experiment 1A, an asymmetrical learning pattern emerged, as L1 was learned while L2 was not learned. One possible explanation is that this asymmetry is due to an order effect. Thus, to test this hypothesis, Experiment 1B reversed the order of presentation, with L2 presented first followed by L1. If the primacy effect hypothesis holds, then participants should learn L2, but not L1.

Methods

Participants

Twenty-nine undergraduate introductory psychology students participated for class credit. I excluded from analysis any participant who reported themselves as being bilingual (4), had seven or more years of second language experience (3), if there was technical failure (1), or if the participant failed to follow instructions (1). This brought the number of participants included in analyses to 20. There were 8 men and 12 women, and all were monolingual English speakers.

Materials and procedure

The methods of Experiment 1B were identical to those used in Experiment 1A. The only variation is that order of language presentation was reversed. Thus, participants listened to L2 for 1 minute and 56 seconds, followed by 1 minute, 56 seconds of L1. This was repeated to create a block of presentation lasting 7 minutes and 44 seconds. This block was repeated three times with a one minute break in between each block, for a total of 23 minutes and 12 seconds of familiarization. Every other aspect of stimuli and procedure were held constant for Experiment 1B.

Results

The mean number of words learned overall was 17.35 out of a possible 32, with a standard deviation of 3.42 (see Figure 8). The mean number of words learned for L1 was 8.40 (SD = 2.85) and for L2 was 8.95 (SD = 1.73). Performance was marginally above chance (chance is defined here as a score of 16) overall, t (19) = 1.76, p = .094. At the level of the individual languages, participants learn L2 (t (19) = 2.45, p < .024) but not L1 (t (19) = 0.63, p = .538). Thus, the pattern of results in Experiment 1B was opposite of what was found in Experiment 1A, lending support to the primacy hypothesis.

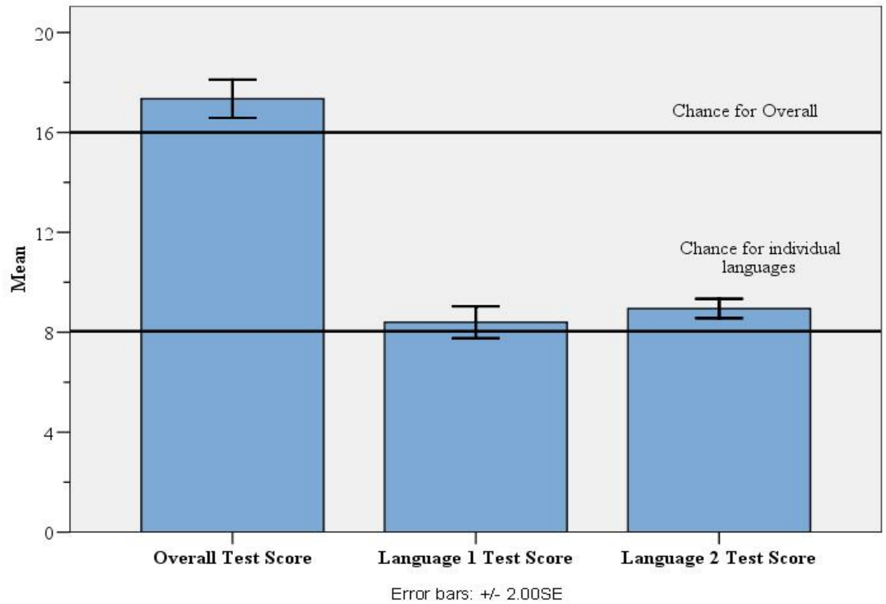


Figure 8. Mean number of words learned in Experiment 1B, overall and for each language.

A repeated measures ANOVA revealed that though L1 was learned better than L2 across both Experiments 1A and 1B, $F(1, 40) = 4.39$, $p = .043$, the interaction between language and order was significant, $F(1, 40) = 10.09$, $p = .003$ (see Figure 9). This supports the hypothesis that the results in Experiment 1A are at least partially due to an order or primacy effect. It does not rule out, however, the possibility that L1 is simply easier to learn than L2. When collapsed across both Experiment 1A and 1B (and thus counterbalanced for order), L1 was still learned significantly above chance, t (41) = 3.21, $p = .003$, while L2 was not learned above chance, t (41) = 0.65, $p = .517$. Thus, it is likely that L1 and L2 are not equivalent in learnability, and this will be addressed in the general discussion.

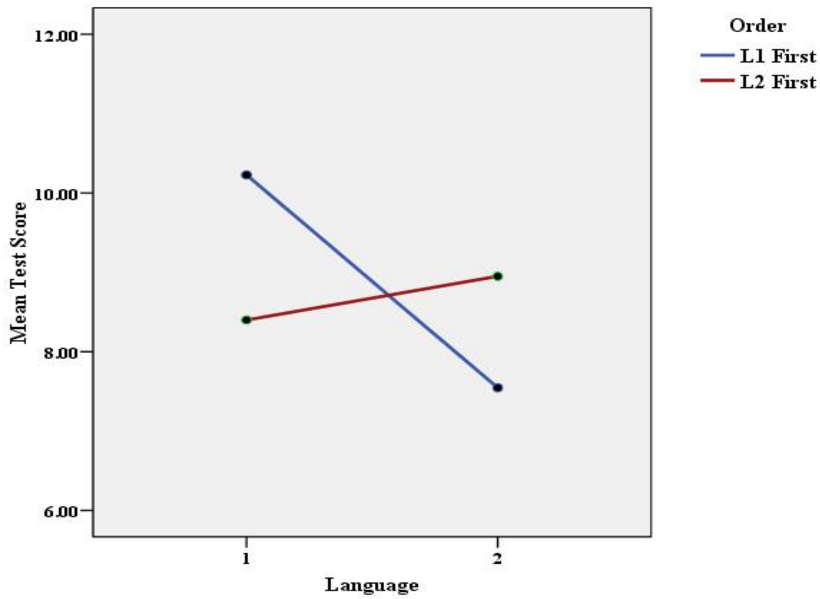


Figure 9. Mean number of words learned by language and order from Experiments 1A and 1B, showing a strong order effect. The blue line represents the order L1 first, while the red line represents L2 first.

Discussion

The purpose of Experiment 1B was to test the primacy effect hypothesis, or the claim that the results in Experiment 1A were due to the order of language presentation. While there was support for this hypothesis, other evidence seemed to support the alternative explanation that L1 is an easier language to learn than L2. Since L1 was learned even when the scores were collapsed across orders, the focus of subsequent experiments will be on learning in the L2. The definition of an indexical cue is any cue to language that would allow *bilingual* language learning. Thus, for a cue to be considered an effective indexical cue, it must facilitate learning in both languages. The asymmetrical learning pattern is consistent with previous experiments (see Weiss & Gerfen, 2006), but only with an effective indexical cue do participants learn both languages. Even though L1 was learned significantly above chance in Experiment 1A and L2 was learned above chance in Experiment 1B, the cue of background color is not an effective indexical cue. However, in both Experiments 1A and 1B, the cue was not present in the test, and so Experiment 1C was designed to explore the possibility that bilingual learning in the first two experiments did not occur because the cue was not present at test.

EXPERIMENT 1C: Background Color—Cue at Test

One confound of Experiment 1A is that during test, the background color remained white throughout. In Weiss and Gerfen (2006), the indexical cue (speaker voice) was present at test. Thus, in order to rule out this as a potential confound and alternative explanation for the results in Experiments 1A and 1B, Experiment 1C replicated the previous experiment, with the lone difference being the presence of the indexical cue at test.

Methods

Participants

Twenty-three undergraduate introductory psychology students participated for class credit. I excluded from analysis any participant who reported themselves as being bilingual (4), had seven or more years of second language experience (2), if there was technical failure (0), or if the participant failed to follow instructions (1). This brought the number of participants included in analyses to 16. There were 7 men and 9 women, and all were monolingual English speakers.

Materials and procedure

The methods of Experiment 1C were identical to those used in Experiment 1A. At test, the background screen was a consistent indexical cue to language. Thus, if the word in the pair of items was from language 1, then the background was purple, and likewise words from language 2 were paired with teal. The prediction is that this will have no effect on performance because the indexical cue allows for the separation of languages at the time of learning, forming distinct, abstract representations of the languages themselves that are not bound to the association with the cue. Other than the indexical cue at test, all other materials and procedures were identical to the previous two experiments.

Results

Overall, participants did not learn the languages above chance performance. The mean number of words learned overall was 15.69 out of a possible 32, with a standard deviation of 2.98 (see Figure 10). The mean number of words learned for L1 was 8.31 (SD = 2.09) and for L2 was 7.38 (SD = 2.06). Performance was not above chance overall, t (15) = -0.42, p = .681. At the level of the individual languages, participants did not learn either L1 (t (15) = 0.60, p = .558) or L2 (t (15) = -1.21, p = .244). Learning in this experiment was not significantly above chance, indicating that the poor performance in Experiments 1A and 1B was not due to the lack of a consistent indexical cue in the test phase of the experiment.

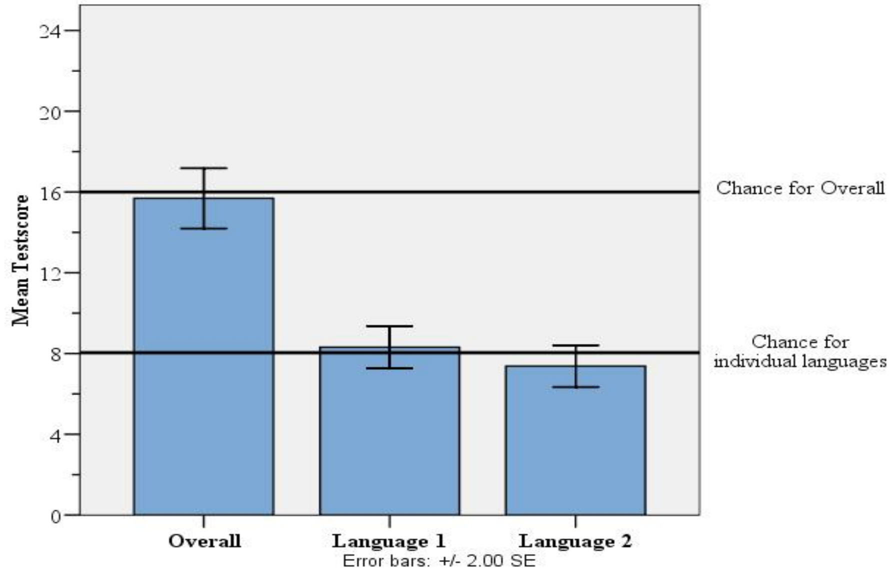


Figure 10. Mean number of words learned in Experiment 1C, overall and for each language.

Discussion

The purpose of this experiment was to rule out the possibility that participants weren't learning the languages because the cue was missing from the test section. Because Weiss and Gerfen (2006) included the indexical cue at test, it was necessary to rule out this alternative explanation. If this account were true, one would expect to find enhanced learning with the cue in test. However, no such facilitation was found, arguing against the hypothesis that a cue in test is necessary for an indexical cue to be effective.

EXPERIMENT 2A: Faces as indexical cues

Experiments 1A-1C demonstrated that a simple visual cue of background color is ineffective as an indexical cue in a bilingual statistical learning paradigm. This suggests that which stimuli can provide an effective indexical cue is limited. While speaker voice was an effective indexical cue in Weiss and Gerfen (2006), the simple visual cue of background color does not appear to facilitate segmentation in a bilingual statistical learning paradigm.

There are a number of possible reasons for why learning of both languages in Experiment 1 was not found. One possibility is that all visual stimuli are ineffective as indexical cues when learning two auditory languages. Another possibility is that background color is not particularly relevant to real-world language processing. That is, participants may either have been ignoring the visual cue altogether or at the very least not integrating it with the information in the auditory domain. To test the latter hypothesis, Experiment 2A explored the efficacy of an indexical cue, faces, that is both in the visual domain and relevant to language processing. Faces have long been known to be an integral feature in language processing and acquisition (see McGurk &

McDonald, 1976). Using faces as an indexical cue provided a test of the role of visual stimuli, for if faces facilitate segmentation in a bilingual statistical learning paradigm, then it suggests that participants are able to integrate information from two separate domains.

Methods

Participants

Forty-three undergraduate introductory psychology students participated for class credit. I excluded from analysis any participant who reported themselves as being bilingual (6), had seven or more years of second language experience (2), if there was technical failure (1), or if the participant failed to follow instructions (2). This brought the number of participants included in analyses to 32. There were 18 men and 14 women, and all were monolingual English speakers.

Stimuli

The same languages were used in this experiment that were used in Experiment 1 as well as in previous research (Weiss & Gerfen, 2006). The only difference will be the indexical cue to language. In Experiment 2A, the indexical cue will be a visual display of dynamic faces. Two female assistants (the faces were female to be consistent with the female voice of the audio stream) were videotaped lip-syncing each of the artificial languages (considered the "active" phase). During videotaping, the assistants sat in a chair approximately five feet from the camera (a Sony Handicam), which was mounted on a tripod four feet from the ground. The familiarization stream was played out on a nearby computer over speakers as assistants read along with the stream from a list of the artificial words in the thirty second stream (which would be looped to create the longer streams) that was held directly below the front of the camera. Each assistant was also videotaped during a rest or "silent" period, where they simply sat in the chair and were instructed to not move their mouth. By using the rest period videos instead of a simple static display, the slight movements in the face (e.g. blinking) simulated a conversation, maintaining the representation of the face as an individual even when not actively lip-synching. After the videos were created, they were then carefully hand edited and synched with the audio from the familiarization streams in Experiment 1 using Adobe Premiere© software. The videos were also cropped and adjusted to ensure an approximately equal size of the faces.

Once individual videos for each face-language pairing (4 total, with two faces and two languages) and the “silent” videos (2 total, one for each face) were created, the familiarization streams were created by taking two individual streams and playing them simultaneously. To do this, each individual video was given its own video track in Adobe Premiere© software, creating three total tracks (two video tracks and one audio track for the familiarization stream) that were then combined. Each individual video track was reduced in size by 51% and then the two tracks were positioned adjacently, such that both videos were equal in size and in the middle of the screen, with Face 1 always appearing immediately left of center and Face 2 always appearing immediately right of center (see Figure 11).

The videos were then concatenated so that one face would be active during one language. Whenever one face was “active” (i.e. lip-syncing to the audio stream), the other face was always passive, or “silent” (see Figure 11). Faces were active whenever the language they were paired with was presented. For example, in one familiarization block, Face 1 and L1 would be active (Face 2 silent) for 1:56, then Face 2 would be active (and Face 1 was silent) for 1:56. This would be repeated to create a block lasting 7:44 with the face ordering of F1 (face 1) active F2 (face 2) silent, F1 silent F2 active, F1 active F2 silent, F1 silent F2 active, and a language ordering of L1, L2, L1, L2. This block was repeated three times with a one minute break in between each block, for a total of 23 minutes and 12 seconds of familiarization.

The test stimuli were identical to that used in Experiment 1A, consisting of an auditory presentation of 16 word – part-word pairs in a two-alternative, forced-choice task with no indexical cue at test.

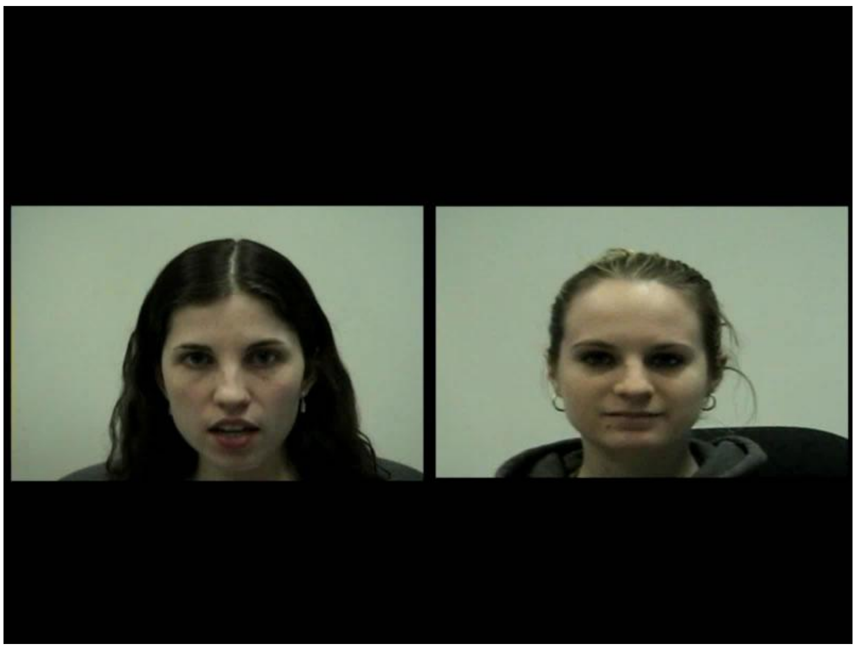


Figure 11. Still frame from the familiarization stimuli used in Experiment 2A. Here, Face 1 is active while Face 2 is silent.

A 2 (Face) X 2 (Language) counterbalancing design was used. Since there were two potential face orderings and two language orderings, four different familiarization streams were created to counterbalance any order effects. Order 1 was F1 and L1 first, Order 2 was F1 and L2 first, Order 3 was F2 and L1 first, and Order 4 was F2 and L2 first. There were 8 participants in each of the four conditions.

Procedure

The procedure was identical to that of Experiment 1A. No explicit verbal instructions were given with regard to the faces.

Results

Overall, participants successfully learned the languages above chance performance. The mean number of words learned overall was 20.00 out of a possible 32, with a standard deviation of 2.81 (see Figure 12). The mean number of words learned for L1 was 10.13 (SD = 2.03) and for L2 was 9.88 (SD = 2.01). Performance was above chance overall, $\underline{t}\,(31) = 8.07$, $p < .001$. At the level of the individual languages, participants learned both L1 ($\underline{t}\,(31) = 5.93$, $p < .001$) and L2 ($\underline{t}\,(31) = 5.27$, $p < .001$). Learning in this experiment was significantly above chance, indicating that faces are an effective indexical cue.

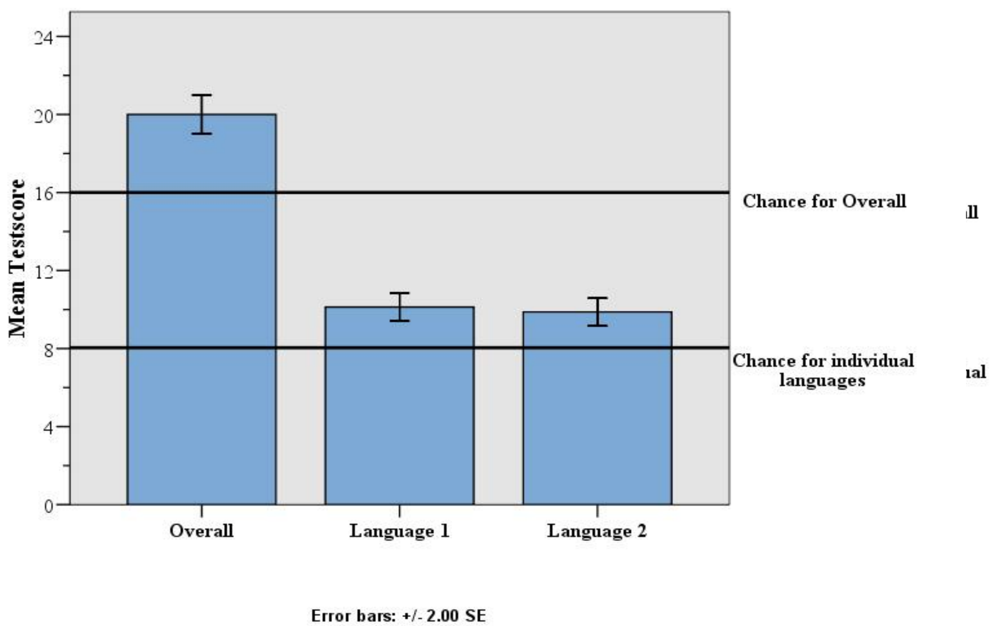


Figure 12. Mean number of words learned in Experiment 2A, overall and for each language.

Additionally, a between-subjects 2 (language order) X 2 (face order) ANOVA revealed no significant order effects, neither for the order of face presentation, $F(1, 28) < 1$, nor for the order of language presentation, $F(1, 28) < 1$. Furthermore, there was no significant interaction between face order and language order, $F(1, 28) < 1$. This suggests that unlike in Experiment 1A and 1B, there was no primacy effect. Participants successfully learned both languages, rather than relying on an initial learning period before the introduction of the second language, further arguing for the effectiveness of faces as an indexical cue.

Discussion

The results of Experiment 2A reveal that faces are an effective visual cue in a bilingual statistical learning paradigm. This discredits the hypothesis that the visual cue in Experiment 1A failed to indexically facilitate performance because it was outside of the auditory domain. Experiment 2A provides evidence that effective cues in the visual domain were integrated with the auditory speech stream, providing a means for participants to form multiple representations, encapsulating the statistics for each language .

EXPERIMENT 2B: Simultaneous presentation of faces

Though participants were able to successfully segment the two languages in Experiment 2A, it is possible that this facilitation was due to the mere presence of faces. To date, there are no studies that use faces in conjunction with a statistical learning paradigm. However, dynamic faces provide a rich source of information, which alone might account for the facilitation witnessed in Experiment 2A. Previous research has found that audio-visual synchrony facilitates segmentation (Hollich, Newman, & Jusczyk, 2005), and so it is possible that the facilitation may not be indexical, but rather the result of the enhanced audio-visual synchrony provided by the faces. To test this hypothesis, synchronous dynamic faces were presented during familiarization and test, but the faces' role as an indexical cue was removed. If faces were indexically facilitating segmentation in Experiment 2A, then one would not expect the simultaneous display of synchronous faces to enhance performance.

Methods

Participants

Twenty-four undergraduate introductory psychology students participated for class credit. I excluded from analysis any participant who reported themselves as being bilingual (3), had seven or more years of second language experience (1), if there was technical failure (0), or if the participant failed to follow instructions (0). This brought the number of participants included in analyses to 20. There were 12 men and 8 women, and all were monolingual English speakers.

Stimuli

The familiarization stream consisted of the same individual videos that were used in Experiment 2A. However, both faces were always active. When either L1 or L2 was presented, both F1 and F2 were active, in synchrony with the audio stream. By removing any relationship between an individual face and a particular language, the faces no longer conveyed information about the language being presented and hence were no longer indexical cues. Thus, any facilitation would be due to the mere presence of faces, rather than to an indexical role. All other aspects of the stimuli were identical to Experiment 2A, including test items.

Procedure

Procedure in Experiment 2B was identical to the procedure from Experiment 2A.

Results

Participants failed to segment the speech stream, as performance in Experiment 2B was not significantly above chance. The mean number of words learned overall was 18.30 out of a possible 32, with a standard deviation of 3.44 (see Figure 13). The mean number of words learned for L1 was 9.35 (SD = 2.94) and for L2 was 8.95 (SD = 2.67). Though performance was significantly above chance overall, t (19) = 2.99, p =.007, it was not above chance at the level of the individual languages, L1 (t (19) = 2.05, p =.054) and L2 (t (19) = 1.59, p = .127). The individual languages were not each learned above chance, indicating that the faces alone do not facilitate segmentation in a bilingual statistical learning paradigm. Additionally, there was no significant language order effect, $F(1, 18) < 1$.

An independent samples t-test revealed that performance overall in Experiment 2A was marginally significantly higher than performance in Experiment 2B, t (50) = 1.95, p = .057. Though performance for the individual languages was not significantly higher (t (50) = 1.13, p = .266 for L1; t (50) = 1.42, p = .161 for L2), performance for both languages trended that way, as learning in Experiment 2A was consistently greater than in Experiment 2B.

Discussion

The goal of Experiment 2B was to determine whether faces alone facilitate segmentation in a bilingual statistical learning paradigm. When the faces no longer acted as an indexical cue, participants were unable to form multiple representations, failing to segment the speech stream. It is possible that audio-visual synchrony provided by the dynamic face display may alone facilitate segmentation. However, participants were unable to segment multiple conflicting speech streams without an effective indexical cue. Though the trend was only marginally significant, it nonetheless suggests that the indexical nature of the face cue augments and facilitates segmentation. However, future work will need to replicate this finding. Overall, the results from Experiment 2B suggest that the learning in Experiment 2A was not due merely to the presence of faces.

GENERAL DISCUSSION

The findings from the present study suggest that visual stimuli are differentially effective as indexical cues. Previous research found that in statistical learning tasks with two languages, participants only successfully segment the speech stream when given a consistent cue to language, or an indexical cue (Weiss & Gerfen, 2006). In the present study, participants were asked to segment a speech stream consisting of two artificial languages. Additionally, two different consistent indexical cues to language were tested: background color and dynamic faces. Participants were able to segment the bilingual speech stream when given an indexical cue of faces, but were failed to segment the speech stream when the indexical cue was background color. Specifically, participants were able to use the face stimuli in order to distinguish between two artificial languages whose statistical structure was incongruent. Thus, when faces appeared indexically learners were capable of forming multiple representations, thereby facilitating the acquisition of both languages. This learning contrasted results from conditions in which the indexical cue was background color, as well as previous results with no indexical cue (Weiss & Gerfen, 2006).

Experiment 1A demonstrated that a consistent and salient visual cue of background color was not an effective indexical cue and did not facilitate the acquisition of multiple languages. Though participants learned Language 1 (L1) above chance, this was likely due to an order effect, as evidenced in Experiment 1B in which the order of language presentation was reversed and the resulting pattern of results was also reversed with Language 2 (L2) being the only language learned above chance. This supports the conclusion that the order of presentation influenced the learning of L1 in Experiment 1A. This order effect is similar to other unpublished experiments that report order effects for the learning of multiple statistical streams (e.g., Aslin, personal communication). Experiment 1C replicated the findings from 1A, finding no indexical facilitation of background color when the cue was incorporated into the test (i.e. test items from L1 were paired with one color, while test items from L2 were paired with another). These results suggest that the overall poor performance evidenced in Experiment 1A was not related to the absence of a visual cue presented during the test. Taken together, the results from Experiment 1 suggest that the simple visual cue of background color is not an effective indexical cue in this experimental paradigm.

In Experiment 2A, participants were given a visual indexical cue that is highly relevant to language: faces. Faces are known to be integrated with auditory processing of language (e.g., McGurk & McDonald, 1976; Hollich, Newman, Jusczyk, 2005; Bertelson, Vroomen, & Gelder, 2003). Due to the prominence of faces in language processing, it was hypothesized that an indexical cue of faces (i.e. each language is paired with a face) would facilitate the segmentation of two artificial languages. The results from Experiment 2A support this hypothesis, as participants successfully segmented both languages above chance, suggesting that faces are an effective indexical cue.

It is possible that the facilitation in Experiment 2A was produced by the introduction of faces, irrespective of their role as indexical cues. To date, no experiment has integrated a dynamic face display in a statistical learning paradigm, and it is unknown what the effect of this would be. To rule out this possibility, the indexical nature of the cue was removed in Experiment 2B. In this experiment, both faces were active simultaneously (i.e. both F1 and F2 were active during L1 presentation) and thus the audio-visual synchrony of the faces and familiarization stream was still intact, but because both faces were active simultaneously, they no longer provided an indexical cue to language. When the indexical cue was no longer present, participants failed to learn either language above chance performance. This suggests that the above chance performance in Experiment 2A was due to the ability of participants to use faces to separate the two languages and maintain distinct representations, rather than the presence of faces.

An important finding from these experiments is that learners are selective with regard to indexical cues, such that not all stimuli are effective indexical cues. That is, there is something specific about indexical cues to which the learner is attuned. If this were not the case, then one would expect that any stimulus that is correlated with language presentation should facilitate segmentation. The visual cue of background color in Experiments 1A-1C was a reliable, consistent cue to language, and yet despite this, failed to elicit segmentation. This suggests that not all potential cues are effective, and therefore learners selectively identify which cues to attend. The following section will address two questions that are raised this finding. First, given that learners are selective with regard to indexical cues, which cues are effective, and what are the relevant features of these cues? Second, how are cues selected; what are the underlying mechanisms of cue selection?

As mentioned earlier, there are an unlimited number of potential indexical cues that learners could attend to. In order to narrow this set, learners must identify the relevant properties of effective indexical cues, and selectively attend to cues based on those properties. It is therefore important to establish what the necessary features of an indexical cue are. Four features of indexical cues that learners may attend to are addressed: the cue’s relevance to language, the domain congruency of the cue and stimulus, whether the cue is informative to speaker identity, and audio-visual synchrony.

The first potential feature of effective indexical cues is the cue's relevance to language processing. The results of the present study, as well as results from previous studies (Weiss & Gerfen, 2006), suggest that an important aspect of indexical cues is their ability to be integrated with the auditory speech stream. In Experiment 1, a cue (background color) that is relatively unrelated with real-world language processing failed to facilitate segmentation while in Experiment 2 a cue that is highly integrated with real-world language processing (faces) produced indexical facilitation. The findings of Experiment 1 suggest that the benefits from indexical cues require an integration of the statistical information in the audio stream and the indexical cue. Thus, cues that are relevant to language processing are more likely to be effective.

A second possible feature of effective indexical cues is domain congruency. While the present study found that indexical cuing works across domains, what is unclear is the effect of domain congruency, when both the cue and the stimulus are in the same domain. Indexical cues in both the auditory (speaker voice, Weiss & Gerfen, 2005) and visual (faces) have been shown to be effective. This suggests that domain congruency is not required for indexical cues, but does it have an effect? It is possible that domain congruency, while not the only relevant feature, plays a role in an indexical cue's effectiveness. For example, if an auditory cue that was not a relevant feature of language processing facilitated segmentation, then this would suggest that a cue can be an effective indexical cue as long as the cue is in the same domain as the stimulus. Future experiments will test the efficacy of an auditory, language-irrelevant cue: background tone. Similar to Experiment 1, the cue would be two simple pure tone pitches overlaid on top of the audio stream. This would provide a direct test of the effect of domain congruency.

The third feature is whether the cue is informative to speaker identity. In the original Weiss and Gerfen (2006) study, a possibility was raised that participants were forming speaker-specific representations, rather than forming representations based on language. Work in perceptual learning and categorization suggests that learners incorporate speaker identity into perceptual representations (Nygaard & Pisoni, 1998; Eisner & McQueen, 2005; Krajlic & Samuel, 2005, 2006, 2007).

Krajlic & Samuel (2007) presented participants with a standard perceptual learning task with multiple speakers. In Experiment 1, participants heard a sound midway between /d/ and /t/ (labeled *?dt* for clarity) in the context of an auditory lexical decision where the target words were produced by two different speakers (one male and one female). In this experiment, *?dt* replaced the /d/ consonant in items produced by one speaker, (e.g. *croco [?dt] ile*), while *?dt* replaced /t/ for items produced by the other speaker (e.g. *café [?dt] eria*). Participants were then given a category identification task for sounds on the /d/-/t/ continuum, with some tokens in the same voice. Experiment 2 was identical except that the sounds occurred in /s/-/Σ/ continuum (e.g. *bro [?sΣ] ure* and *ob [?sΣ] ene* for *brochure* and *obscene* respectively).

In perceptual learning tasks like this, perceptual learning occurs if the perception of sounds in the continuum is influenced by the prior lexical context. For example, perceptual learning would occur if participants more often classified sounds on the /d/-/t/ continuum as /d/ when the ambiguous sound occurred in the context of *crocodile* than if the sound had occurred in the context of *cafeteria*. In Krajlic and Samuel (2007), if perceptual learning is speaker-specific, then opposing information from the two speakers during familiarization should be irrelevant, and perceptual learning should occur. However, if perceptual learning is the product of general phonemic adjustments, then perceptual learning should not occur. Krajlic and Samuel (2007) found that participants exhibit perceptual learning only when the differences between target items in the continuum are informative to speaker identity (e.g. spectral shifts the /s/-/Σ/

continuum). When the differences are uninformative with respect to speaker identity (e.g. temporal shifts, or Voice Onset Time shifts, in the /d/-/t/ continuum) learners "reset" their representations to baseline when they encounter stimuli from different speakers during familiarization, resulting in a lack of perceptual learning. This effect was also found in dynamic, synchronous displays of faces (Bertelson, Vroomen, & Gelder, 2003).

The results of Krajlic and Samuel (2007) suggest that when there were speaker cues (/s/ and /Σ/), perceptual learning was speaker specific and information was encapsulated, but when there were no speaker cues (/d/ and /t/), perceptual learning was speaker-general and information was combined. The findings from the present study mirror this. The way the languages in the current study were created, if participants encapsulated the languages, then participants should show learning. However, if they combined representations across languages, then the statistics should conflict and participants should fail to learn. Thus, the findings of the current study are consistent with the Krajlic and Samuel (2007) study, as when the cue was informative to speaker identity (faces), the statistical representations were encapsulated and the languages learned, but when the cue was uninformative with respect to speaker identity (background color), the representations were combined and the languages were not learned. The current study suggests that in bilingual statistical learning experiments, speaker-specific representations are formed. Therefore, being informative to speaker identity is an important feature of effective indexical cues.

However, it is important to note that the results of the present study do not rule out the possibility that participants form language specific representations. Only a total of three cues have been tested, and as mentioned, there are numerous possible explanations for why Experiment 1 yielded no learning. In order to more directly test for language-specific representations, one could create two different familiarization streams, one with a speaker-specific indexical cue (e.g. spectral shifts) and one with a speaker-independent indexical cue (e.g. voice onset time). If learners form speaker-specific representations, then only the speaker-specific indexical cue should facilitate bilingual learning. However, if learners form language-specific representations, then both cues should be effective, as both are auditory (domain congruent), language relevant cues, which are hypothesized here to be the features of effective indexical cues.

A final potential feature of effective indexical cues is synchrony. It is possible that learners are able to use the correlation of a synchronous visual display to augment the transitional probabilities in the audio stream. Hollich, Newman, and Jusczyk (2005) found that in addition to dynamic faces, a synchronous oscilloscope display facilitated the segmentation of a monolingual speech stream. Additionally, in the present study, the static display of background color did not facilitate bilingual segmentation, suggesting that synchrony may be an important aspect of indexical cues. However, it is important to note that synchrony alone cannot account for the findings in Experiment 2A. The dynamic face display in Experiment 2B presented participants with a synchronous audio-visual display, and this did not facilitate segmentation. A further test of this would be to present participants with a static face display as an indexical cue. Previous research has found that unlike dynamic displays, static face displays do not facilitate segmentation (Hollich, Newman, & Jusczyk, 2005). However, the static faces may still be an effective indexical cue. Such an experiment would test the importance of synchrony as a feature of indexical cues.

The previous section addressed *what* the relevant features of effective indexical cues are, highlighting four potential attributes. However, the question that remains is *how* cues are selected. The findings from Experiment 1A point to the necessity of a cue selection process. If not all cues facilitate the formation of multiple representations (and thus acquisition of both statistical streams), then how and when does a learner decide which cues, given myriad possibilities, to track? Is the cue selection process guided by experience, or is it guided instead by innate preferences for certain cues or specific features of cues?

One possible way of addressing these questions would be to inquire about the developmental trajectory of this selective process. The results of the current study suggest that adult learners are sensitive to certain properties of the cues, and this may guide cue selection. However, this study tested adults, who likely have already formed cue selection strategies. The participants here may have failed to integrate the background color cue because experience dictated that background color is not a reliable cue to language. Alternatively, background color may simply not enjoy the benefit of an inborn preference to attend to background color. There is evidence that newborns have an innate preference for attending to faces (Goren, Sarty, & Wu, 1975; Morton & Johnson, 1991; Valenza, Simion, Macchi Cassia, & Umiltà , 1996), whether through an innate sub-cortical mechanism (Morton & Johnson, 1991; Tsao, Freiwald, Tootell, & Livingstone, 2006; Kanwisher, 2006) or due to biases from general properties of the infant visual system (Kleiner, 1987; Banks & Ginsburg, 1985; Simion, Macchi Cassia, Turati, & Valenzia, 2001; Macchi Cassia, Turati, & Simion, 2004; Turati, 2004). Thus, it is possible that newborns have innate biases to attending certain cues, and this drives the cue selection process.

By testing cue preferences in newborn learners, who have yet to acquire significant language experience, it may be possible to contrast these opposing hypotheses. If cue selection is driven by experience, then one might expect that newborns should treat the visual cue from Experiment 1 similarly to the faces from Experiment 2, as initially cues would be weighted equally (these weights would be shifted based on experience, resulting in the preferences observed in the present study). However, if cue selection is innately guided, then one would expect that newborns should show a degree of cue selectivity and that cue preferences should be relatively stable across the developmental trajectory. Future studies will adapt the bilingual statistical learning paradigm for use with an infant population, first replicating the Weiss and Gerfen (2006) study, and then exploring possible indexical cues.

Moving away from the issue of cue selection, the present study also has implications for real-world bilingual language acquisition. A central issue in bilingual language acquisition is whether infants in a bilingual setting have the capacity to discriminate between the two languages in the speech input. As mentioned earlier, there is evidence that infants as early as 4 months of age can discriminate two languages based on rhythmic properties of the input (Bosch & Sebastian-Galles, 2001). However, this does not indicate whether infants are actually able to use this discrimination to parse and learn the two languages. Indeed, the One-Parent One-Language (OPOL) hypothesis suggests that in order for infants to successfully learn multiple languages simultaneously, the languages must be systematically separated. Briefly, the OPOL hypothesis states that in order to reduce confusion between languages, the optimal way to raise a child bilingually is to separate the languages by parent, with one parent paired with one language (Ronjat, 1913; Barron-Hauwaert, 2004; Döpke, 1998). For example, if the goal is to raise a Spanish-English bilingual, one parent would speak *only* Spanish to the child and the other parent would speak *only* English.

Though the OPOL hypothesis is widespread, and seems to receive popular acceptance as the best possible strategy for raising a child bilingually (Barron-Hauwaert, 2004), “there is no proof of its psycholinguistic reality” (Hamers & Blanc, 1989; 38). The majority of the evidence for the OPOL hypothesis comes from cases studies, primarily done with the researcher's own children (e.g., Leopold, 1939-1949; Taeschner, 1983; Hoffman, 1985; Harding & Riley, 1986; Dopke, 1992). These case studies have no control or comparison groups, and are hard to generalize to a more general population (Goodz, 1989; Arnberg, 1987). Indeed, there is evidence that a lack of any strategy is as effective as OPOL (Doyle, Champagne, and Segalowitz, 1978).

Despite the lack of empirical evidence to support its claims, OPOL still receives considerable popular support; therefore, it is important to test this hypothesis. The present study is an important first step towards that end. OPOL essentially proposes that infants need an indexical cue (parents), in order to maintain separate representations of the languages. The findings from the present study, as well as the results from Weiss and Gerfen (2005), provide evidence, at least superficially, for this claim. The present findings suggest that adult learners require an indexical cue to successfully segment two incongruent languages. Furthermore, the visual cue in Experiment 1 was not an effective indexical cue, suggesting that cues specific to speaker identity (voice and face) are effective, whereas other cues are not. Thus, it is possible to interpret this as evidence consistent with the OPOL hypothesis. However, this support must be tempered for a number of reasons. First, it is possible that the faces might facilitate bilingual acquisition even if deployed inconsistently. Future studies will test this by manipulating the face to language mapping such that each face is paired with each language (e.g. F1/L1, F1/L2, F2/L1, F2/L2 in two-minute blocks). This would loosely simulate a bilingual, code-mixing environment in which both parents speak both languages. This would provide empirical evidence as to whether OPOL is a necessary strategy. Second, the present study is reductionistic, removing any possible indexical cues other than one. However, there are numerous potential cues to language (e.g. phonology) and real languages may fundamentally differ, so it is it is possible many other cues would work just fine may facilitate bilingual segmentation, without requiring OPOL.

In conclusion, the present study found that a dynamic, synchronous display of faces as an indexical cue facilitated the acquisition of two artificial languages in a bilingual statistical learning paradigm. Further experimental conditions demonstrated that successful learning hinged on indexical synchronous information being available to the learner. Neither a static visual cue of color background nor a dynamic display of faces that actively produce both languages (and therefore was not indexical) produced successful learning of both languages. In addition to the planned experiments mentioned, future work is needed to replicate the finding with new artificial languages in order to eliminate any possible asymmetries in the languages learability. Finally, it will be important to explore the developmental trajectory of cue selection through infant studies, as well as examining how multiple indexical cues interact.

REFERENCES

Arnberg, L. (1987). *Raising children bilingually: The preschool years*. Clevedon: Multilingual Matters.

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

Aslin, R., Woodward, J., Lamendola, N. and Bever, T. (1996). Models of word segmentation in fluent maternal speech to infants. In Morgan, J. and Demuth, K. (eds), *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Language Acquisition*. Mahwah (NJ): Lawrence Erlbaum Associates, 117–134.

Banks, M., & Ginsburg, A.P. (1985). Infant visual preferences: A review and new theoretical treatment. In H.W. Reese (Ed.), *Advances in child development and behavior* (Vol. 19, pp. 207–246). New York: Academic Press.

Barron-Hauwaert, S. (2004). *Language strategies for bilingual families: The One-Parent-One-Language approach*. Clevedon, England: Multilingual Matters.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: a McGurk aftereffect. *Psychological Science, 14*, 592–597.

Bortfeld, H., Morgan, J., Golinkoff, R., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech stream segmentation. *Psychological Science, 16*, 298-304.

Bosch, L., & Sebastian-Gallés, N. (2001). Evidence of early language discrimination abilities in infants from bilingual environments. *Infancy, 2*(1), 29-49.

Brent, M. and Cartwright, T. (1996), Distributional regularity and phonotactics are useful for segmentation. *Cognition, 61*, 93–125.

Brent, M. and Siskind, J.(2001), The role of exposure to isolated words in early vocabulary development. *Cognition, 81*, 31–44.

Christiansen, M., Allen, J. and Seidenberg, M.(1998), Learning to segment speech using multiple cues. *Language and Cognitive Processes. 13*, 221–268.

Christophe, A., Gout, A., Peperkamp, S. & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics, 31*, 585-598.

Cole, R., & Jakimik, J. (1980). "A model of speech perception." In Cole, R. A. (ed), *Perception and Production of Fluent Speech*. Hillsdale: Lawrence Erlbaum Associates.

Cole, R., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America, 64*, 44-56.

Cook, G. & Sprague, R. (1995). Feeling the parts: A developmental study of separable perception with tactile dimensions. *Journal of General Psychology, 122*, 171-186.

Creel, S. C., Newport, E. L., and Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1119-1130.

Curtin, S., Mintz, T.H., & Christiansen, M.H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition, 96,* 233-262.

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

Dahan, D., & Brent, M. (1999). On the Discovery of Novel Word-like Units from Utterances: An Arti cial-language Study with Implications for Native-language Acquisition. *Journal of Experimental Psychology: General, 128*, 165-185.

Döpke, S. (1992) *One parent–one language: an interactional approach*. Amsterdam: Benjamins

Döpke, S. (1997). Is the simultaneous acquisition of two languages in early childhood equal to acquiring each of the two languages individually? In E. Clark (ed.), *Proceedings of the 28th Annual Child Language Research Forum*, 95-113.

Dopke, S. (1998). Can the principle of 'One person – one language' be disregarded as unrealistically elitist? *Australian Review of Applied Linguistics, 21*(1), 41-56.

Doyle, A., Champagne, M., & Segalowitz, N. (1978). Some issues on the assessment of linguistic consequences of early bilingualism. In M. Paradis (Ed.), *Aspects of bilingualism*. Colombia, SC: Hornbeam Press.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*, 224–238.

Fiser, J., & Aslin, R.N. (2002). Statistical learning of higher-order temporal structure from visual shape-sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 458-467.

Garcia, E.E. (1980). The function of language switching during bilingual mother-child interaction. *Journal of Multilingual and Multicultural Development, 1*(3), 243- 252.

Garner, W.R. & Felfoldy, G.L. (1970). Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology, 1*, 225-241.

Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language, 16*, 161-179.

Genesee, F. (2000). Introduction. *Bilingualism: Language and Cognition, 3*(3), 167-172.

Genesee, F., Nicoladis, E., & Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language, 22*, 611-632.

Gerken, L. A. (2002). Early sensitivity to linguistic form. *Annual Review of Language Acquisition, 2,* 1–36.

Goodz, N.S. (1989). Parental language mixing in bilingual families. *Infant Mental Health Journal, 10*, 25–44.

Goren, C., Sarty, M., & Wu, P. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*, 544-549.

Hamers, J., & Blanc, M. (1989). *Bilinguality and bilingualism*. Cambridge: Cambridge University Press.

Harding, E. & P. Riley (1986). *The bilingual family: a handbook for parents*. Cambridge: Cambridge University Press.

Hauser, M. D., Newport, E. L., & Aslin, R. N. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, B53-B64.

Hoffmann, C. (1985) Language acquisition in two trilingual children. *Journal of Multilingual and Multicultural Development, 6,* 479–495.

Hollich, G., Newman, R. S., & Jusczyk, P.W. (2005). Infants' Use of Synchronized Visual Information to Separate Streams of Speech. *Child Development, 76*, 598-613.

Houston, D. M., Jusczyk, P. W., Kuijpers, C., Coolen, R., & Cutler, A., (2000). Cross language word segmentation by 9-month-olds. *Psychonomic Bulletin & Review, 7*(3), 504-509.

Houston, D.M., Santelmann, L., & Jusczyk, P.W. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes, 19*, 97-136.

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*, 548-567.

Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology, 11*, 381- 394.

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., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of english words. *Child Development, 64*, 675—687.

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

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

Kanwisher, N. (2006). What's in a face? *Science, 311*, 617-618.

Kirkham, N.Z., Slemmer, J.A., & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence of a domain general learning mechanism. *Cognition, 83*, B35- B42.

Klatt, D., & Stevens, K. (1978). On the automatic recognition of continuous speech: Implications from a spectrogram-reading experiment. *IEEE Transactions on Audio and*

Electroacoustics, 21(3), 210- 217.

Klatt, D.H. (1979). Speech perception: A model of acoustic phonetic analysis and lexical access. *Journal of Phonetics, 7*, 279-312.

Kleiner, K. A. (1987). Amplitude and phase spectra as indices of infants' pattern preferences. *Infant Behavior and Development,10,* 49-59.

Kooijman, V., Hagoort, P., & Cutler, A. (2005). Electrophysiological evidence for prelinguistic infants' word recognition in continuous speech. *Brain research. Cognitive brain research, 24*(1), 109-116.

Kraljic, T., & Samuel, A.G. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language, 56*, 1-15.

Kraljic, T. & Samuel, A.G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*, 262–268.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*, 141–178.

Kuhn, G. & Dienes, Z. (2005). Implicit learning of non-local musical rules. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *31*(6), 1417-1432.

Leopold, W.F. (1939–49). *Speech development of a bilingual child. Vols. I–IV*. Evanston, Ill.: Northwestern University Press. Reviewed in Barron-Hauwaert, S. (2004). *Language strategies for bilingual families: The One-Parent-One-Language approach*. Clevedon, England: Multilingual Matters.

Macchi Cassia, V., Turati, C., & Simion, F. (2004). Can a non-specific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15*, 379-383.

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,* 465- 494.

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

Mattys, S.L., & Jusczyk, P.W. (2001b). Phonotactic cues for segmentation of fluent speech by infants. *Cognition 78*, 91-121.

Mattys, S.L., White, L., & Melhorn, J.F (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477-500.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Merkx, M. & Monaghan, P. (2006). A hierarchy of prosodic cues in speech processing. *Proceedings of the 28th Annual Conference of the Cognitive Science Society.*

Morton, J., & Johnson, M. H. (1991). Conspec and Conlern: A two-process theory of infant face recognition. *Psychological Review, 98,* 164-181.

Müller, N. & Hulk, A. (2001). Crosslingusitic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition, 4*, 1-21.

Nazzi, T., Iakimova, G., Bertoncini, J., Fredonie, S., & Alcantara, C. (2006). Early segmentation of fluent speech by infants acquiring French: Emerging evidence for crosslinguistic differences. *Journal of Memory and Language, 54*, 283-299.

Nazzi, T., Jusczyk, P.W., Johnson, E.K., 2000. Language discrimination by English learning 5-month-olds: effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1-19.

Newport, E. L., Weiss, D. J., Wonnacott, E., & Aslin, R. N. (2004). Statistical learning in speech: Syllables or Segments? In B. U. C. o. L. Development (Ed.). Boston.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-spcific leanring in speech perception. *Perception & Psychophysics, 60*, 355-376.

Onishi, K.H., Chambers, K.E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13-B23.

Reddy, D.R. (1976). Speech recognition by machine: A review. *Proceedings of the IEEE , 64* (4), 501- 531.

Ronjat, J. (1913). *Le développement du langage observé chez un enfant bilingue*. Paris: Champion. Reviewed in Barron-Hauwaert, S. (2004). *Language strategies for bilingual families: The One-Parent-One-Language approach*. Clevedon, England: Multilingual Matters.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Directions in Psychological Science, 12*, 110-114.

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

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*, 27-52.

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

Simion, F., Valenza, E., Macchi Cassia, V., Turati, C., & Umilta` , C. (2002). Newborns' preference for up-down asymmetrical configurations. *Developmental Science, 5*, 427–434.

Taeschner, T. (1983). *The sun is feminine: A study on language acquisition in bilingual children.* Berlin: Springer.

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

Thiessen, E.D., Hill, E.A., & Saffran, J.R. (2005). Infant directed speech facilitates word segmentation. *Infancy, 7*, 49-67.

Toro, J. M., & Trobalon, J. B. (2005). Statistical computations over a speech stream in a rodent. *Perception & Psychophysics, 67*(5), 867-875.

Tsao, D. Y., Freiwald, W. A., Tootell, R. B. H., &. Livingstone, M. S (2006). A cortical region consisting entirely of face-selective cells. *Science*, *311*, 670-674.

Turati, C. (2004). Why Faces Are Not Special to Newborns: An Alternative Account of the Face Preference. *Current Directions in Psychological Science*, *13*, 5-8.

Turk-Browne, N.B., Jungé, J.A., & Scholl, B.J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General, 134*, 552-564.

Volterra, V., & Taeschner, T. (1978). The acquisition and development of language by a bilingual child. *Journal of Child Language*, *5*, 311-326.

Valenza, E., Simion, F., Macchi Cassia, V., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 892–903.

Weiss, D.J. & Gerfen, C. (2006) Language segmentation in a bilingual environment. *Proceedings of the 30th Annual Boston University Conference on Language Development*.

Weiss, D.J., Gerfen, C., Mitchel, A., & Rizzo, W. (In review). The role of pauses and statistics in word segmentation: A new perspective on colliding cues. Submitted to *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

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.

Werker, J.F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2006). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition*.

Wingfield, A., & Titone, D., (1998). Sentence Processing (chapter). In Gleason, J. & Ratner, N. (Eds), *Psycholinguistics*, 2nd ed. Fort Worth: Harcourt Brace College Publishers.

Wood, N., & Cowan, N. (1995). The Cocktail Party Phenomenon Revisited: Attention and Memory in the Classic Selective Listening Procedure of Cherry (1953). *Journal of Experimental Psychology: General, 124* (3), 243-262.