All words are not created equal: Expectations about word length guide infant statistical learning

Casey Lew-Williams *, Jenny R. Saffran

Department of Psychology and Waisman Center, University of Wisconsin-Madison, United States

1. Introduction

Smith (2000) describes learning as a historical process. Learners do not simply make connections between stimulus events. Instead, ‘the formation of initially simple associations changes what is attended to and, in doing so, changes what will be learned in the future’ (p. 172). There have been many demonstrations of how learning occurs against the backdrop of what has been learned previously, across diverse domains including categorization (Lin & Murphy, 1997), concept learning (Wisniewski & Medin, 1994), shape learning (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002), scene perception (Biederman, Mezzanotte, & Rabinowitz, 1982), visual pattern detection (Thiessen, 2011), musical pattern detection (Tillman & Poulin-Charronnat, 2010), syntax learning (Lany & Gómez, 2008), and classroom education (Ennis, 1992). This approach to learning has also dominated research on second language learning for decades, yielding a highly robust and convergent finding: an individual’s first language knowledge shapes and constrains acquisition of a second language (e.g., Bates & MacWhinney, 1981; Lado, 1957; Lew-Williams & Fernald, 2010; Nemser, 1971). Here we applied this dynamic, temporal perspective on learning to a basic issue in early language acquisition, in order to refine what is currently known about the operation of mechanisms that support learning in infancy.

In the last 15 years, developmental researchers have become interested in ‘statistical learning,’ a mechanism argued to drive learning from the earliest months of life. Statistical learning refers to the domain-general ability to extract structure from patterned input (Fiser & Aslin, 2002; Kirkham, Slemmer, & Johnson, 2002; Romberg & Saffran, 2010). Much of the research on statistical language learning has investigated how infants learn to segment word-like units from continuous speech. Using artificial-language (Aslin, Saffran, & Newport, 1998; Saffran, Aslin, & Newport, 1996) and natural-language paradigms...
Lew-Williams, Pelucchi, & Saffran, 2011; Pelucchi, Hay, & Saffran, 2009), studies have demonstrated that infants are sensitive to the regularity with which one syllable will transition to another syllable, known as transitional probability (TP). It is argued that infants’ general ability to detect statistical dependencies helps them identify where words begin and end in the ambient language, in tandem with many other cues specific to that language (Sahni, Seidenberg, & Saffran, 2010).

Researchers are only just beginning to understand how experience shapes infant statistical learning. For example, highly frequent lexical items such as ‘Mommy’ are easy for infants to detect in continuous speech (Bortfeld, Morgan, Golinkoff, & Rathbun, 2005; Singh, Nestor, & Bortfeld, 2008). Previous learning of stress patterns also influences segmentation skills: infants more readily segment unfamiliar words if they match the predominant stress pattern in the native language (Houston, Santelmann, & Jusczyk, 2004; Johnson & Jusczyk, 2001; Thiessen & Saffran, 2003), and shift their segmentation strategy based on exposure to novel stress patterns (Thiessen & Saffran, 2007). Moreover, adults’ phonological knowledge shapes the detection of new statistical structure. For example, Catalan (Toro, Pons, Bion, & Sebastián-Gallés, 2011) and English speakers (Finn & Hudson Kam, 2008) have difficulty segmenting words that violate subtle, language-specific rules for word formation. Saffran and Thiessen (2003) illuminated how this perceptual specialization may take shape in infancy. Given brief exposure to specific phonological patterns (artificial words with syllable-initial voiceless consonants, e.g., *todkad*), infants successfully segmented words that fit the pattern, but failed to segment words that violated expectations (words with syllable-initial voiced consonants, e.g., *bupgok*). Together, this research suggests that learners are biased by prior experiences with word frequency, stress, and phonology when determining word boundaries in new linguistic input.

A recent study with English-speaking adults took this idea in an important direction, showing that such biases can actually override attention to sequential statistics. Frank, Tily, Arnon, and Goldwater (2010) designed an artificial language that could be segmented into disyllabic, trisyllabic, or hexasyllabic units. Crucially, only the hexasyllabic lexicon was supported by segmenting at locations with low TPs, but few adults actually segmented the language into these larger chunks. Most participants preferred disyllabic or trisyllabic segmentations, imposing an apparent bias for shorter units. Indeed, English words are short relative to some languages, including Spanish: the 680 words included on the MacArthur-Bates Inventario del Desarrollo de Habilidades Comunicativas are a mean 2.7 syllables in length (vs. 1.7 syllables in English). In contrast, words in Mandarin Chinese are predominantly monosyllabic. Across time, experience with shorter vs. longer words may lead infants to form perceptual expectations for future language input, and thereby impact statistical learning from fluent speech.

The current study was designed to determine whether we could change infants’ expectations about where words begin and end in a statistical learning task. Does prior experience mediate the deployment of language learning mechanisms in infancy? Rather than manipulating the infants’ native language, we used a laboratory simulation of learning across time, in order to avoid the introduction of biases related to phonology and stress inherent in cross-linguistic comparisons. Infants were first pre-exposed to a list of nonsense words that were deterministically either disyllabic or trisyllabic. This between-subjects manipulation was intended to induce a specific expectation about word length. Infants were then exposed to fluent speech generated using new nonsense words that were either disyllabic or trisyllabic, with only sequential statistical cues to word boundaries. Thus the length of words heard during the pre-exposure phase was either consistent or inconsistent with the length of words embedded in fluent speech during the segmentation phase (Fig. 1). Some infants heard words of uniform length throughout the experiment (disyllabic or trisyllabic in both phases), while other infants were led down a “garden path” of sorts, where initial expectations about word length did not match the statistical structure of the fluent speech stream (Gebhart, Aslin, & Newport, 2009). Infants were then tested by assessing their preference for repetitions of target words (TP = 1.0) vs. frequency-matched part-words that spanned word boundaries (TP = .5).

If pre-exposure to words of a particular length has no effect upon subsequent statistical learning, the pattern of listening times should not differ as a function of whether the pre-exposure was consistent or inconsistent with the structure of the fluent speech. That is, hearing a list of trisyllabic words should not impact infants’ ability to detect disyllabic words in fluent speech, or vice versa. However, if word-length expectations do influence statistical learning, listening times to words and part-words should vary across the consistent and inconsistent conditions. Pre-exposure to trisyllabic words should facilitate infants’ detection of trisyllabic words relative to disyllabic words in fluent speech. Similarly, pre-exposure to disyllabic words should facilitate detection of disyllabic words relative to trisyllabic words. This outcome would indicate that prior experience shapes statistical learning.

2. Method
2.1. Participants

Participants were 96 healthy, full-term infants (43 female) from monolingual English-speaking households. Infants ranged in age from 9.0 to 10.9 months ($M = 10.0$ months), matching the approximate age of
participants in previous studies on infant word segmentation using a similar three-stage task (Saffran & Thiessen, 2003; Thiessen & Saffran, 2007). Parents reported no pervasive developmental delays and no history of hearing or visual problems. Twenty-four infants were randomly assigned to one of four between-subjects conditions. Forty additional infants were tested but not included in analyses due to fussiness (29), inattentiveness (6), experimenter error (3), or equipment malfunction (2). This rate of participant exclusion was expected due to the length of the experimental session relative to past artificial-language studies.

2.2. Stimuli

In the pre-exposure phase, infants heard a list of 30 nonsense words, each separated by a 300-ms pause. Words were either disyllabic (CVCV; e.g., boga...giku...) or trisyllabic (CVCVCV; e.g., bogapi...gikutu...), depending on condition. Pre-exposure words were disyllabic in the Consistent\textsubscript{2–2} and Inconsistent\textsubscript{2–3} conditions, and trisyllabic in the Consistent\textsubscript{3–3} and Inconsistent\textsubscript{2–3} conditions; subscripts indicate the number of syllables in pre-exposure and target words (Fig. 1). All speech stimuli were recorded in citation form by a female native English speaker and presented at approximately 67 dB SPL.

During the segmentation phase, infants heard pause-free speech consisting of four new nonsense words, repeated in random order. Two of the words occurred 35 times each (low-frequency words) and two occurred 70 times each (high-frequency words). This phase was modeled after Aslin et al. (1998), except that frequency of exposure to each target word was reduced by 22.2% to increase the potential impact of the pre-exposure manipulation. Target words were disyllabic in the Consistent\textsubscript{2–2} and Inconsistent\textsubscript{2–3} conditions, and trisyllabic in the Consistent\textsubscript{3–3} and Inconsistent\textsubscript{2–3} conditions. Successive syllables within all target words had TPs of 1.0. There were two counterbalanced disyllabic languages and two counterbalanced trisyllabic languages, such that each test item was a word for half of the infants and a part-word for the other half (see Appendix). To approximate the operations of a speech synthesizer and maintain natural coarticulation, we followed the speech recording procedures used in Graf Estes, Evans, Alibali, and Saffran (2007, Experiment 2). We recorded each target syllable in the middle of a 3-syllable sequence, in every possible coarticulation context. Middle syllables were spliced together into a fluent speech stream with consistent rate (3.1 syllables/s) and pitch ($F_0 = 196$ Hz); there were no acoustic cues to word boundaries.

In the test phase, infants heard repetitions of a single item on each trial, separated by 300-ms pauses. Test items for each infant were either all disyllabic (Consistent\textsubscript{2–2} and Inconsistent\textsubscript{2–3} conditions) or all trisyllabic (Consistent\textsubscript{3–3} and Inconsistent\textsubscript{2–3} conditions), matching the segmentation materials. Test items consisted of two low-frequency target words (TP = 1.0), and two frequency-matched part-words, consisting of the last syllable of one high-frequency word and the beginning of the other high-frequency word (TP = .5 for disyllabic items; TP = 0.5 and 1.0 between successive syllable pairs for trisyllabic items).

2.3. Procedure

The experiment consisted of three phases: pre-exposure (disyllabic, 33 s; trisyllabic, 40 s), segmentation (disyllabic, 2:12; trisyllabic, 3:18), and Test (~3 min), all using the head-turn preference procedure. Infants were seated on parents’ laps; parents listened to unrelated speech over closed-ear headphones. During the pre-exposure phase, infants heard a list of nonsense words. During the segmentation phase, infants heard a stream of continuous speech; blinking side-lights were flashed contingent on looking behavior. The experimenter coded head-turns using custom-designed MATLAB software (R2010b, Mathworks, Inc.). During the test, a trial was initiated when infants looked to a flashing center light, at which point the experimenter extinguished the center light and one of the side-lights began flashing. When infants looked at the side-light, a token of a word or part-word was repeated.

![Fig. 2. Mean looking times to words and part-words in the four conditions. Error bars represent standard errors of the mean.](image-url)
of either disyllabic or trisyllabic words changed the course of learning. Infants only showed successful discrimination between words and part-words when pre-exposure and target words were consistent in length. When word lengths were inconsistent across phases of the experiment, infants did not successfully use TPs to discriminate between test items. Our results thus provide evidence that expectations about word length shaped infants’ processing of statistical cues to word boundaries. These findings are revealing about the manner in which prior experience and new learning interact. Obviously, real language environments do not provide such overt and deterministic cues to word length, and infants do not have fixed expectations about word length. Instead, infants become gradually attuned to the structure inherent in the global language input. Prior experience with the distribution of word lengths in the native language should facilitate subsequent analyses of input, in tandem with knowledge of phoneme distributions (Christiansen, Onnis, & Hockema, 2009), lexical stress (Johnson & Jusczyk, 2001), and highly frequent words (Bortfeld et al., 2005).

Why did expectations about word length change infants’ sensitivity to statistical relations? One possibility is that top-down knowledge led infants to seek words of a particular length in fluent speech. Infants may have attempted, but ultimately failed, to divide a trisyllabic language into disyllabic words, or vice versa. However, it is not necessary to posit that infants derive word structure so actively. Swingley (2005) questioned the idea that infants insert dividing points into unsegmented speech, suggesting instead that infants cluster syllables that co-occur reliably. By studying how prior word knowledge influences the detection of syllable clusters, our results support an integration of these two descriptions of segmentation: during the pre-exposure phase, infants may have internalized a two- or three-part rhythm, which interacted with their perception of transition statistics. When TP-defined word boundaries aligned with this rhythmic expectation, infants used convergent evidence from syllabic co-occurrence to accurately segment words. But a mismatch derailed successful detection of syllable clusters. Computational, adult, and developmental data indicate that both top-down expectations and bottom-up processes are inevitably involved in learning (Bortfeld et al., 2005; Lew-Williams et al., 2011; Reber, 1989; Sun & Zhang, 2004). Our experiment design illuminates how an interaction between past and present learning may operate as infants build knowledge of structure in their linguistic environments.

Our results converge with and advance recent findings in research on statistical learning. Johnson and Tyler (2010) showed that infants have difficulty using TPs to segment continuous speech when the component words vary in length. Similarly, infants in our study who heard words of inconsistent lengths across the two learning phases of the experiment were not successful in segmenting continuous speech. These findings could indicate that statistical learning mechanisms cannot account for the variability inherent in real language, leading to catastrophic interference when the structure of the input diverges across time. Alternatively, our findings suggest an efficient processing strategy: infants used prior language experience to quickly
narrow in on the most accurate analysis. This perspective is compatible with a growing body of research suggesting that learners do not blindly attend to transition statistics (Frank & Tenenbaum, 2011; Giroux & Rey, 2009). Bayesian approaches to word segmentation posit that learning is constrained by prior assumptions about the nature of words (Goldwater, Griffiths, & Johnson, 2009). Our experiment provides a snapshot of this learning process in infant learners. Pre-exposure to a list of disyllabic or trisyllabic words induced a prior, or learning bias, that infants adopted when processing subsequent language input. This process mirrors the fine-tuning of infants’ phonological perception to sounds in the ambient language over the first year (Werker & Tees, 1984), and supports an experimentally provocative description of learning by Smith (2000): “we learn about what we attend to, and we learn what to attend to” (p. 170). A complete theory of statistical learning will need to account for how infants accumulate language experience and update word knowledge.

Future research should aim to uncover how cross-linguistic differences in word length guide statistical learning in infancy. Multisyllabic words are more common in Spanish than in English (Cutler & Carter, 1987; Roark & Demuth, 2000), and the consequences of this difference emerge early in language production: infants exposed to Spanish produce more multisyllabic babbles than infants with no Spanish exposure (Ward, Sundara, Conboy, & Kuhl, 2009). Our findings suggest that similar effects will emerge early in language comprehension. Understanding how the distribution of word length and other types of variability in home-language experience shape subsequent learning will be important for analyzing the biases that infants bring to lab-based experiments, and for uncovering how learning unfolds over time.

Acknowledgments

This research was funded by grants from the Eunice Kennedy Shriver National Institute of Child Health and Human Development to C.L.W. (F32HD069094), J.R.S. (R01HD037466), and the Waismann Center (P30HD03352). Additional support was provided by a grant from the James F. McDonnell Foundation to J.R.S. Thanks to the participating families and the members of the Infant Learning Lab. We are also grateful to Jon Willits and Erik Thiessen for feedback on an earlier version of this manuscript.

Appendix A

Segmentation phase

<table>
<thead>
<tr>
<th>Disyllabic</th>
<th>Trisyllabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language 1</td>
<td>Language 2</td>
</tr>
<tr>
<td>baku&quot;</td>
<td>pido&quot;</td>
</tr>
<tr>
<td>tiro&quot;</td>
<td>lagu&quot;</td>
</tr>
<tr>
<td>dola&quot;</td>
<td>roba&quot;</td>
</tr>
</tbody>
</table>
| gupi" | kuti" | daropi" | budopa"

* = Low-frequency.
** = High-frequency.

Test phase

<table>
<thead>
<tr>
<th>Disyllabic</th>
<th>Trisyllabic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language 1</td>
<td>Language 2</td>
</tr>
<tr>
<td>baku&quot;</td>
<td>baku&quot;</td>
</tr>
<tr>
<td>tiro&quot;</td>
<td>tiro&quot;</td>
</tr>
<tr>
<td>pido&quot;</td>
<td>pido&quot;</td>
</tr>
</tbody>
</table>
| lagu" | lagu" | pigola" | pigola"

* = Word.
** = Part-word.

Pronunciation: |a| = ‘ah’; |i| = ‘ee’; |o| = ‘oh’; |u| = ‘oo’.

References


