INVESTIGATING LEXICAL SEGMENTATION IN 7.5-MONTH OLD INFANTS

By

COURTNEY ANN MEOLA

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree
With Honors in

Psychology

THE UNIVERSITY OF ARIZONA

M A Y 2 0 1 7

Approved by:

Dr. Rebecca Gómez
Department of Psychology

Abstract

The present study aims to better pinpoint the amount of exposure a 7.5-month-old infant requires to become familiar with an unfamiliar word such that it helps infants correctly identify, or segment, a subsequent novel word in running speech. Previous literature has shown that statistical learning and segmentation are instrumental in language acquisition in humans. Infants detect the predominant strong-weak stress pattern of their native language (Jusczyk, Cutler, and Redanze, 1993). At certain points in development, infants weight segmentation strategies and stress patterns differently leading them to incorrectly segment TARis after hearing "The guitar is…". Infants have shown that they are able to override a metrical bias (that otherwise leads them to missegment words) when presented with a highly familiar phrase containing "Mommy" or "Mama" but not when presented with an unfamiliar word like "Lola" (Sandoval, LeClerc, & Gómez, 2016). We tested the hypothesis that infants could learn "Lola" with sufficient exposure and whether or not this previously unfamiliar word would similarly allow them to override the metrical bias to accurately segment a weak-strong word. Infants became familiar enough to "Lola" after two weeks of exposure that it assisted them in segmenting weak-strong patterned words following it in running speech.

Introduction

From birth, infants face a difficult task: acquiring their native language. One major challenge in their development is gaining the ability to detect words in spoken language. As an example, if an individual did not have any familiarity with the French language, they may hear a stream of spoken French sentences as a long run on of sounds rather than words broken up by pauses. In a written format, distinguishing where one sentence begins and ends would be easier. But obviously infants cannot read they can only listen and observe those who are speaking to

them. Infants have trouble recognizing where one word begins and ends in a speech stream. Nevertheless, infants do grow up and become functioning language using toddlers, adolescents and adults. A strategy discussed by Jusczyk et al. (1993) is metrical segmentation where infants pay more attention to predominant stress patterns in their native language. In English, the predominant stress pattern is a strong-weak pattern (zebra). A less common pattern would be a weak-strong pattern (giraffe). The important ability to detect stress in syllables acts as a foundation for segmentation and ultimately, word learning.

Previous literature has examined statistical learning and the way it impacts language acquisition in adults and children (Saffran, Aslin & Newport, 1996). However, what is less understood is exactly how young babies acquire a large lexicon and how retention is involved in this process. A puzzling finding about statistical learning in infants is that they have excellent capabilities for phonetic strategies and demonstrate very rapid learning in a laboratory setting. However, real world language acquisition and retention is quite slow in comparison (Gómez, 2017). A leading explanation for this mismatch is that infants' exposure to certain situations and lexicons change frequently since the real world is quite different than a controlled laboratory setting. Therefore, there is a gap in the literature with an almost lack of infants being tested for their retention over time versus testing their abilities to perform in tasks immediately after learning. There is also a gap in the literature with infants' learning occuring almost solely in a lab versus at their own home (which is where real world learning would normally occur).

One study that did investigate retention by Jusczyk and Hohne (1997) tested infants that were 8-9 months of age on their ability to recognize words from a storybook. In this particular study, infants listened to 3 short prerecorded stories for 30 minutes each day at their homes for ten days. 14 days later the infants came into a lab and heard words from the stories. Using the

head-turn preference procedure, infants were tested on their ability to differentiate frequent words from the stories from words they did not hear in the stories. The infants' listening times were significantly longer to the familiar words from the recordings suggesting that they were able to recognize the sound patterns of words from the stories. This is important because the ability to recognize sound patterns will help infants draw inferences about statistical patterns that occur in their native language as they hear more and more language. The more inferences infants are able to make the better they will be able to detect frequent patterns in their language.

Frequency is another important component in SL as it has been well documented to establish and strengthen perception of words over time. Shi et al. (2006) set out to discover whether or not 8 through 11-month-old participants would be better at using a high-frequency function morpheme word such as "the" versus a less common morpheme like "she" to segment words from speech. Their results indicated that both older infants (11 months of age) and 8-month-olds were more successful in the usage of "the" to segment adjacent words. They suggest that because "the" has such an extremely high frequency in speech, infants are more familiar with the word and distinguish the beginning and end of "the" to help them detect and segment the following word.

Language contains lexical, metrical, and segmental-acoustic cues that become apparent at various points of development (Sandoval & Gómez, 2016). Lexical cues refer to words themselves or a vocabulary. Metrical cues are otherwise known as "stress cues" and refer to the syllable structure of a word. Segmental-acoustic cues refer to the spoken sound of a word and how the brain transduces it. Depending on the age of the infant, these components are thought to combine and assist with language learning. Using one of these strategies alone would most

certainly lead to inaccurate segmentation (Sanders & Neville, 2000). The weighting of these cues is what determines an infant's success in word segmentation.

The predominant strong-weak syllable pattern in English is also called a trochaic pattern. For example, the word apple (APP-le) is trochaic. According to Jusczyk et al. (1993) infants show a preference for trochaic words because they hear this pattern statistically more often. Subsequent research conducted by Jusczyk et al. (1999) tested 7.5-month old infants' ability to pick out bisyllabic words from speech. They discovered that the infants were unable to segment iambic words, but they were able to segment trochaic words. This preference leads English-learning 7.5-month-old infants to have a bias known as a metrical bias, which negatively impacts their ability to correctly segment iambic (weak/strong syllable patterned) words in a stream of speech. Metrical cues are very important for word learning but when this bias occurs, an infant would missegment a word such as "guitar" as: TAR+ the first syllable of the next word (mistaking a weak strong syllable pattern for a strong weak). See Figure 1 for an illustration of this. The authors also found that by 10.5 months infants segment iambic words. They believe this points to the older infants' advanced brain development and better ability to integrate multiple sources of information about the boundaries of words. Is there a way 7.5-month infants can overcome this bias? Or must they rely on development and simply wait to begin segmenting iambic words until they are months older?

In research conducted by Sandoval and Gómez (2016), 7.5-month old infants were familiarized with passages containing Mommy/Mama followed by two out of four chosen iambic words (guitar, device, beret, or surprise) and Lola/Lolly followed by the other two words. Whether the child heard "Mommy" or "Mama" depended on how the mother referred to herself, with "Lola" or "Lolly" matching. For instance if the mother referred to herself as "Mommy", the

child heard "Mommy" and "Lolly" passages. Infants successfully segmented the two iambic test words following "Mommy" or "Mama" because they showed a preference in longer listening times when compared to the words following "Lola" or "Lolly". In a second study, when exposed to Lola or Lolly (unfamiliar to the infants), they did not show significant listening times for iambic words following Lola or Lolly compared to control words the infants had not heard. These results suggest that segmentation of iambic words for this age group requires much more than a repetition of a word, but rather a *familiarity* with a word.

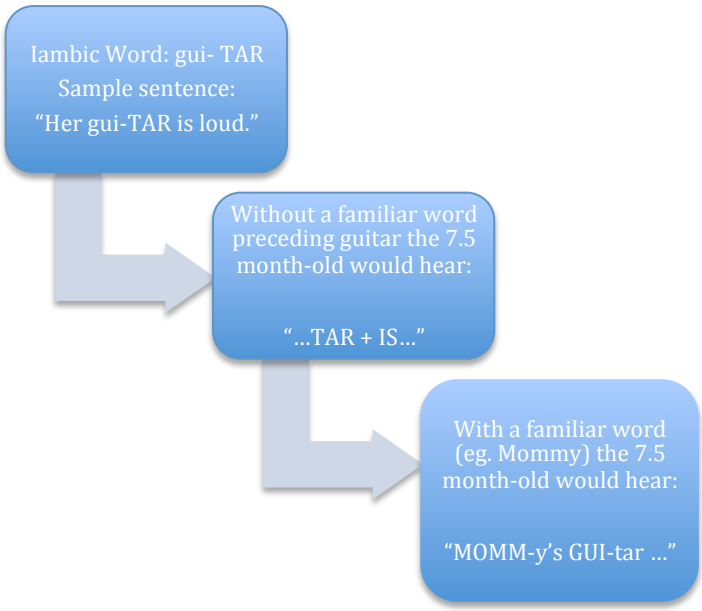

Figure 1: Diagram illustrating missegmentation and segmentation in 7.5 month-olds

The present study aims to investigate the amount of exposure infants this age require to become familiar with an otherwise unfamiliar word (Lola) in order for them to correctly segment iambic words from speech. We predict that if exposure to Lola helps the infants extract iambic words from speech, then listening times for words after Lola will be significantly longer than words that did not follow Lola in a speech stream. If Lola does not assist the infants in segmentation, then we would expect the listening times between the control words (words not

after Lola and not occurring in the materials) and the target words (words occurring after Lola) to be indistinguishable from each other statistically.

Methods

Participants

The participants of this study were 12 (4 male, 8 female) healthy, monolingual English-learning 7.5-month-old infants from the city of Tucson. They were recruited through the use of the Child Cognition Lab's database of signups. This database has been compiled from parents volunteering their names and phone numbers at various recruiting events around Tucson, from birth announcements, and from signing up at the University Medical Center after giving birth. Guardians were contacted, told about the study, and then were invited to participate as long as their infant was not familiar with the name Lola. The average age at the first exposure to the story was 7.40 months and at test was 7.85 months. Two infants kept in the data set listened to the storybook 13 out of 14 days due to their caregiver forgetting about the story one day. We determined one less exposure was still acceptable but no more than one exposure could be missed. Infants were excluded if they fussed out and could not complete 8 blocks of trials (2), had chronic ear infections which reduced their ability to hear (1), or their caregiver was not able to come in on the fifteenth day (3).

Materials

After recruitment, an experimenter visited parents at their homes to give them the materials for the training portion of the study which consisted of the storybook, a sleep log for logging naps and nighttime sleep, four CDs with the storybook recorded on them, and a CD player. Each recording of the storybook contained 31 exposures to "Lola" meaning that after 14 exposures, each infant heard "Lola" 434 times. Each of the four CDs contained a recording of

one of four female lab members reading the storybook word for word. We used CDs to ensure that the speed and pronunciation of the words remained consistent throughout the homes of the participants. We also felt it necessary to have four different women's voices speaking the story because real world learning in infants involves many sources of language exposure.

Procedure

Training Phase. See Figure 2 for a depiction of the procedure in time. Infants' caregivers were instructed to play one of the four CD's containing the storybook recording once a day for fourteen days within an hour of the infant's sleep. The storybook log randomized and listed the dates for the upcoming two weeks with the preferred order of CD's (labeled A, B, C, and D). Sleep could be either a daytime nap or nighttime sleep. We were particularly interested in having the caregivers record sleep patterns and play the recording within an hour of sleep because previous literature has shown that sleep is important in learning especially for infants and children (Gomez, Bootzin &Nadel, 2006; Gómez & Edgin, 2015; Simon et al., 2017). Sleep in young infants and toddlers supports generalization of information after learning and can also enhance retention. If the audio sounded too loud or too quiet in the participants' homes, they were to adjust the volume accordingly but to note they had changed the volume level. While the storybook played, caregivers were to have the infant sit on their lap and hold out the book in front of the infant like a typical storybook experience. Caregivers were also told to note any odd events that happened during the playing of the storybook. After fourteen days, guardians and their infants came to the lab and participated in the test portion of the experiment on the 15th day.

Test phase. The test portion occurred in the Child Cognition Lab at the University of Arizona. Caregivers were consented and instructed to take a seat in the head-turn preference booth facing a video camera.

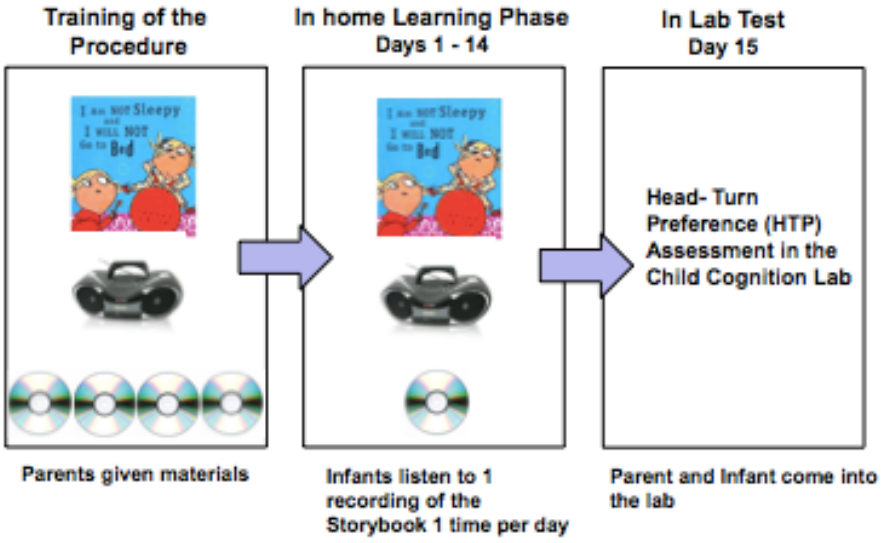

Figure 2: Diagram of Training Procedure

Stimuli

Familiarization Stimuli

During familiarization in the head-turn preference booth, the infants first heard passages narrated by an English speaking, female played at a stable volume that contained two Iambic words prefaced with "Lola". The four possible iambic words chosen were, "guitar, device, beret, and surprise". A typical type of statement for familiarization would sound like "Lola's beret is a pretty color", or "Lola's guitar has a nice sound". After familiarization, the infant proceeded to the test.

Test Stimuli

The test phase was the repetition of all four iambic words in isolation randomized for twelve trials with some variation in intonation. Suring testing the words repeated multiple times until the child looked away from the according light for 2 seconds or had listened to the entire trial (see details of the test booth and procedure below). Once the trial had ended, the next trial

began with the infant having to look at the center light. The process continued until the 12th, and final, trial ended.

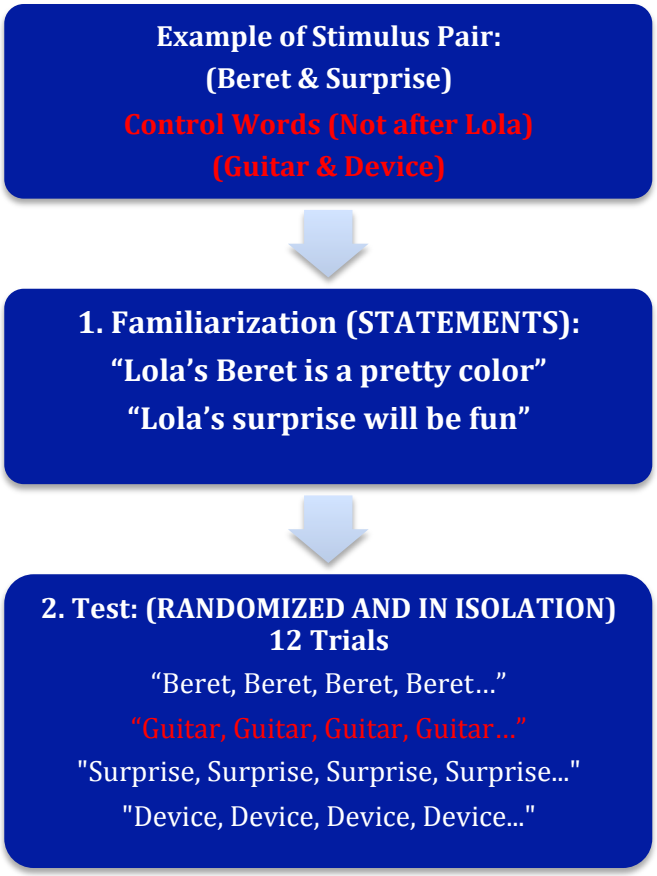

Figure 3: Diagram showing familiarization and test stimuli

Design

Participants were assigned to a pair of words. For example, this could be "surprise" and "beret" or "guitar" and "device". During familiarization infants would hear two of these words in streams of speech prefaced with "Lola". The paired word conditions were counterbalanced between male and female participants.

During test, infants heard 12 trials of repetitions of all four of the iambic words spoken in isolation by the same female voice. The two words after "Lola" from familiarization were referred to as the target words, and the two words not after "Lola" were the control words.

Test procedure and apparatus

The testing took place in the head-turn preference booth. The booth itself is dimly lit and contains a chair with a middle red light and two orange sidelights that flash in front of the infant on each side. Initially, a center light would flash in front of the infant. In a booth next to the test booth, the experimenter would start the familiarization trial by pressing a button. Once the infant looked at the center, a light would flash on either the left or right while the familiarization statements played. The infants sat on the caregivers' laps. Caregivers wore headphones with music playing to mask the audio their infants listened to so they would not be as tempted to interfere or direct their infant. Caregivers were also instructed to not point or try to direct their infant in any way towards the stimuli or lights. If an infant showed signs of crying or fussiness, the experimenter took a break from playing the stimuli and put on a brief puppet show to regain the infants' attention. The experimenter in a separate room observes the infant through the camera and determined when the infant would fixate on the center and sidelights.

Results

Figure 4 shows the mean listening times to the words that occurred after Lola versus the control words.

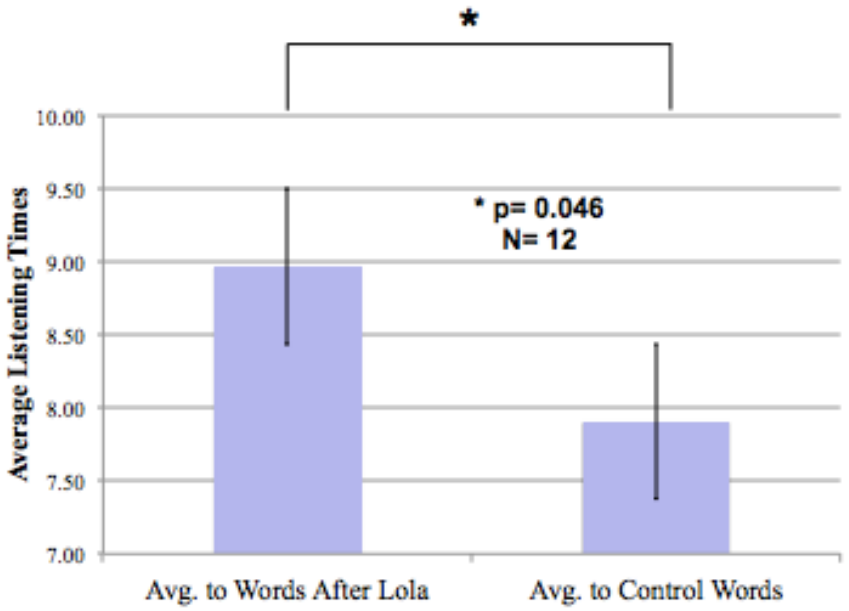

Figure 4: Average Listening times

As shown in Figure 4, a paired, two-tailed T-test was performed with the average listening times for words after Lola and control words. The value of p= 0.046 indicates that average listening times to words following Lola were significantly greater than that of the control iambic words. This demonstrates that in streams of speech, infants were able to use their familiarity with Lola to correctly pay attention to and segment target iambic words.

Discussion

The present study found a significant difference in the listening time for familiarity to Lola versus listening time to control words. Therefore, after 14 days of exposure to Lola in the storybook (434 exposures total), infants became familiar *enough* with the name to segment the iambic words. This corresponds to Sandoval and Gomez's (2016) finding that infants can segment an iambic word after "Mommy" or "Mama". However it cannot be determined whether this much exposure is necessary and whether or not the we can decrease the amount of exposure. A future study could examine the results after ten days, and even less (such as a week) and see

how the exposures compare. Another issue that this study did not address is whether or not the infants used their exposure to Lola from only the past 24 hours to assist them in segmentation. In a follow-up study infants listen to the storybook once a day for seven days and then the infants have a seven day period where they have no exposure before the test. This will allow us to ask whether 7 days (or half the exposure) is sufficient to set down the memory infants can retain the memory for 7 days. This type of design will contribute to our aim of understanding how much exposure infants need to become familiar with a word and remember it. The idea is to create a high frequency, familiar experience that would be similar to when infants hear "Mommy" or "Mama" in their daily lives, not an uncommon experience that only happens once but is memorable to an infant because it was novel. If infants do not retain Lola at a level that will allow them to segment a novel word, we will test two more conditions. In one condition we will test infants on Day 8. In the other condition we will expose infants for 14 days and test them on Day 21. Our current design does not inform exactly how much exposure infants need to form a memory of a word strong enough to aid in segmentation and how long they can retain it. Our proposed studies will help converge on this aim.

Implications

The main implication of the present study is the understanding and establishment of a baseline amount of exposure required for an infant to become familiar to a word. The discovery made by Sandoval and Gómez (2016) that points to the familiarity of a word as a driver of segmentation is important and needs to be pursued by future research. The more that researchers can find out about the exposure young infants require to segment the less frequent stress patterns making up iambic words, the richer the literature will be in terms of understanding how infants weight statistical learning strategies and ultimately how they begin to lay down a foundation for

their native language. While statistical learning and language acquisition are becoming increasingly popular to study, research in young infants outside of a laboratory is scarce, and research regarding retention in this area is scarcer.

Works Cited

Gómez, R. L. (2017). Do infants retain the statistics of a statistical learning experience? Insights from a developmental cognitive neuroscience perspective. *Phil. Trans. R. Soc. B*, *372*(1711), 20160054.

Gómez, R. L., Bootzin, R., & Nadel, L. (2006). Naps promote abstraction in language learning infants. *Psychological Science, 17,* 670-674.

Gómez, R. L., & Edgin, J. O. (2015). Sleep as a window into early neural development: Shifts in sleep-dependent learning effects across early childhood. *Child development perspectives*, *9*(3), 183-189.

Jusczyk, P. W., Cutler, A. and Redanz, N. J. (1993), Infants' Preference for the Predominant Stress Patterns of English Words. Child Development, 64: 675–687. doi:10.1111/j.1467 8624.1993.tb02935.x

Jusczyk, P. W., & Hohne, E. A. (1997). Infants' memory for spoken words. *Science*, *277*(5334), 1984-1986.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. Cognitive Psychology, 39 (3–4), 159–207. doi:10.1006/cogp.1999.0716

Saffran JR, Aslin RN, Newport EL. 1996 Statistical learning by 8-month-old infants. Science 274, 1926 – 1928. (doi:10.1126/science.274.5294.1926)

Sanders, L. D., & Neville, H. J. (2000). Lexical, Syntactic, and Stress-Pattern Cues for Speech Segmentation. *Journal of Speech, Language, and Hearing Research : JSLHR*, *43*(6), 1301–1321.

Sandoval, M., & Gómez, R. L. (2016). Overriding the metrical bias with lexical information: English-learning 7.5-month-olds use mommy to segment iambic words. *Language Learning and Development*, *12*(4), 398-412.

Shi, R., Cutler, A., Werker, J., & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. The Journal of the Acoustical Society of America, 119, EL61–EL66. doi:10.1121/1.2198947

Simon, K., Werchan, D., Goldstein, M., Sweeney, L., Nadel, L., Bootzin, R. R., & Gómez, R. L., (2017). Sleep and statistical language learning: Insights into cortical development and retention in very young infants. *Brain and Language*. Available online for early view. http://dx.doi.org/10.1016/j.bandl.2016.05.002