Early developing syntactic knowledge influences sequential statistical learning in infancy

Erik D. Thiessen

Department of Psychology

Carnegie Mellon University

Luca Onnis

Program in Linguistics and Multilingual Studies

Nanyang Technological University

Soo-Jong Hong

Kyung-Sook Lee

Asan Medical Center

Address for Correspondence:

Erik D. Thiessen

Department of Psychology

Carnegie Mellon University

5000 Forbest Avenue

Pittsburgh, PA 15213

e-mail: thiessen@andrew.cmu.edu

Abstract

Adults' linguistic background influences their sequential statistical learning of an artificial language characterized by conflicting forward-going and backward-going transitional probabilities. English-speaking adults favor backward-going transitional probabilities, consistent with the head-initial structure of English. Korean-speaking adults favor forward-going transitional probabilities, consistent with the head-final structure of Korean. These experiments assess when infants develop this directional bias. Seven-month-old infants showed no preference for forward-going or backward-going regularities. By 13 months, though, English-learning infants favor backward-going transitional probabilities over forward-going transitional probabilities, consistent with English-speaking adults. This indicates that statistical learning rapidly adapts to the predominant syntactic structure of the native language. Such adaptation may facilitate subsequent learning by highlighting statistical structures that are likely to be informative in the native linguistic environment.

Word Count: 4,411

Statistical information has been argued to play an important role in language development (e.g., Saffran, Aslin, & Newport, 1996; Romberg & Saffran, 2010; Thiessen & Erickson, in press). One aspect of statistical learning has been studied especially closely in this regard: the use of conditional statistical information to group linguistic elements into units (Perruchet & Vinter, 1998; Thiessen, Kronstein, & Hufnagle, in press). For example, sounds within words predict each other better than sounds across word boundaries, and both infants and adults can use this information to group sounds together into words (e.g., Aslin, Saffran, & Newport, 1998; Graf-Estes, Evans, Alibali, & Saffran, 2007). Similarly, learners are sensitive to the likelihood of words occurring together, and can use this information to identify phrasal clusters in a string of words (Saffran, 2001; Thompson & Newport, 2007). Sensitivity to the predictable relation among elements of the input may play an especially important role early in language acquisition, because unlike many acoustic cues to linguistic structure it does not require infants to have language-specific biases or expectations (e.g., Thiessen & Saffran, 2003).

Although statistical learning is an early-developing and potentially universal cue to linguistic structure, different languages are characterized by different statistical regularities. How do learners move from universal preferences to idyosincratic ones that best fit the ambient language? In this paper we propose that statistical learning biases to language-specific regularities emerge early on in ways that can make young learners better prepared for subsequent learning in that language. The literature relating statistical learning to language development has not thoroughly explored the possibility that mechanisms of statistical learning may start off as universally open to the input, but progressively be subject to developmental trajectories in ways that optimize the input of each language.

One such difference in the statistical regularities available in different languages relates to the predominant directionality of conditional relations in the input. Languages contain both predictive (forward-going) and retrodictive (backward-going) relations among elements of the input. These relations are not necessarily identical; for example, while “the” does not strongly predict “dog” (because many words can follow “the”), “dog” strongly retrodicts “the.” Recent experimental work using artificial input has demonstrated that learners are sensitive to informative relations in both directions. For instance, Jones and Pashler (2007) showed participants sequences of shapes governed by probabilistic relations, and found that participants were able to recall both which shapes reliably occurred after a probe shape in the input (prediction test) and before a probe shape (retrodiction test). Similarly, both infants and adults are able to segment fluent speech into words on the basis of either forward-going relations among syllables, or backward-going relations (Pelucchi, Hay, & Saffran, 2009; Perruchet & Desaulty, 2008).

In natural languages, the predominant directionality of relations among elements of the input can differ. One example of this is described in linguistic terms as the “headedness” of a language. The head of a phrase is the word that defines the syntactic function of the phrase (i.e., the verb in a verb phrase). Some languages (such as English) are classified linguistically as head-initial, meaning that the head of the phrase tends to occur before complement items (e.g., “going” in “going home”), while other languages are head-final and show the opposite word-order tendency (Haspelmath, Dryer, Gil, & Comrie, 2005). A related regularity describes the use of adpositions in the language.

English is a prepositional language (e.g., “to school”), while other languages favor postpositional organization (“school to”, as in Korean and Japanese). An intuitive prediction deriving from the linear organization of the input is that English word clusters are more syntactically cohesive in a backward-going direction. For example, in a phrase like “to school,” “to” does not strongly predict any word – because many nouns can follow “to” – but “school” more strongly retrodicts “to” because there is a relatively small set of words that can precede “school.” More generally, as these examples demonstrate, learners of different languages may experience different degrees of forward-going and backward-going cohesiveness as a function of the structure of their linguistic input. To assess this possibility, Onnis and Thiessen (in press) performed a corpus analysis of English (a predominantly head-initial and prepositional language) and Korean (a predominantly head-final and postpositional language). The results indicated that in English, high backward transitional probabilities and low forward transitional probabilities were a better indicator of phrase cohesiveness than high forward transitional and low backward probabilities; in Korean, the opposite pattern held true.

These differences in linguistic input, in turn, may lead to changes in implicit statistical learning biases. Sensitivity to backward-going regularities may be more adaptive to learners in an English environment than for learners in a Korean environment. Consistent with this hypothesis, Onnis and Thiessen (2013) found differences between English and Korean speakers when they were exposed to an artificial grammar in which forward and backward transitional probabilities were in conflict. Native English speakers grouped the syllables together on the basis of backward transitional probabilities, while native Korean speakers grouped the syllables together on the basis of forward transitional probabilities. By contrast, with either visual or tonal stimuli, English and Korean speakers performed equivalently. The fact that the difference in performance between English and Korean speakers is limited to linguistic input, and consistent with the predominant emphasis in their native language on forward-going or backward-going regularities, suggests that the difference is due to experience with the native language.

The finding that adult Korean and English native speakers favor opposite directional cues to cohesiveness when parsing an artificial grammar begs a developmental question: when did these language-specific learning biases emerge? It is often argued in the statistical learning literature that statistical learning mechanisms contribute to the development of higher order syntactic knowledge (e.g., Saffran, 2003; Thompson & Newport, 2007). This perspective suggests that an adaptation to the structure of the native language occurs in preverbal infants and toddlers, such that statistical learning can help infants better discover the structural properties of the linguistic input. An alternative possibility, though, is that this kind of directionality bias emerges relatively late, as a by-product of established and potentially abstract syntactic knowledge rather than as an integral adaptation necessary to acquire that knowledge.

In two experiments, we attempt to track the origins of sensitivity to syntactic structure. We first assess the grouping biases of English exposed infants at 7 and 13 months. At 7 months, infants appear sensitive to statistical relations between novel syllable sequences (Aslin, Saffran, & Newport, 1998), and thus this group represents a plausible starting age to test language-specific biases in an artificial language learning task. However, they are just beginning to develop a vocabulary, so we would not expect them to have the opportunity to develop a systematic bias. By 13 months, however, infants have learned several words, even though they have not discovered higher-order linguistic structure such as phrasal regularities, a sensitivity that is typically first documented in infants between 19 and 24 months (e.g., Jolly & Plunkett, 2008; Kedar, Casasola, & Lust, 2006; Noble, Rowland, & Pine, 2011).

Experiment 1: English-learning infants

Method

Participants.

Data from 25 7-month-old and 25 13-month-olds was included in the final sample. An additional 14 infants (six 7-month-olds and eight 13-month-olds) were excluded from the analysis for the following reasons: fussing or crying (7), failure to look at the test trials for an average of at least three seconds (3), and experiment error (3). All participants lived in the Pittsburgh area and were being raised in monolingual English speaking homes.

Stimuli.

The stimuli used in this experiment were based on those used in Onnis and Thiessen (2013). Infants were familiarized with a sequence of syllables based on a Markovian grammar chain organized around 8 symbols (X, Y, A, B, C, D, E, and F) organized into a sequence. To generate the sequence, the first symbol was chosen at random, and then each subsequent symbol was generated according to the grammar's probabilistic sequence rules (see Figure 1). For example, given the symbol X, three possible symbols (D, E, or F) could follow with equal probability. In this case, the forward transitional probability between X and any of those symbols is .33, while the backward probability was 1 because D, E, and F were always preceded by X. At the next transition, given any of the symbols D, E, or F, only one syllable (Y) could follow. This yields a forward probability of 1, and a backward probability of .33. The crucial feature of this grammar is that it features conflicting forward-going and backward-going transitional probabilities. Whenever a syllable pair has a low forward-going transitional probability, it has a high backward-going transitional probability, and vice versa. If infants prioritize forward-going transitional probabilities, XD is not a statistically coherent grouping. But if infants prioritize backward-going transitional probabilities, XD is statistically coherent.

The 8 symbols underlying the grammar were mapped onto 8 different syllables (foo, sae, zee, ra, nee, boo, tee, gae). To control for item biases, each infant heard a familiarization with a different randomized mapping (i.e., for one infant syllable A = foo, while for a different infant syllable A = tee). The 8 syllables were synthesized and concatenated by the MBROLA (Dutoit, 1997) synthesizer using the Italian male voice. This voice and language combination were chosen for their novelty to both English learning and Korean learning infants (see Experiment 2). At the same time, English and Korean adult native speakers independently confirmed that all syllables were distinctly perceptible and pronounceable in both English and Korean. Each syllable lasted 340 msec (with a 260 msec vowel), and there were no pauses between syllables. The familiarization stream lasted for 3 minutes and 11 sec, during which time infants heard 350 syllables.

At test, infants were presented with two different kinds of bisyllables: items with high forward transitional probabilities and low backward transitional probabilities (HiLo test items, e.g. DY), and items with low forward transitional probabilities, and high backward transitional probabilities (LoHi items, e.g. XD). The test items differed for each infant, as a function of the mapping between the underlying grammar and the surface realization of the syllable. But for each infant, both the HiLo and the LoHi items occurred equally often during familiarization; the only difference between them was the directionality of statistical coherence.

Procedure.

Infants were tested individually in a sound-attenuated testing room, seated on a caregiver's lap 150 cm away from a 32 inch LCD monitor. An experimenter outside the testing room observed the infant over closed-circuit video and recorded the duration of his or her gaze at the central monitor using the Habit X software (Cohen, Atkinson, & Chaput, 2004). To eliminate bias, parents were asked to wear headphones, and the experimenter was blind to the nature of the stimuli being presented. Two speakers situated next to the central LCD monitor were used to present the audio stimuli.

At the beginning of the experiment, the infants' attention was attracted to the central LCD monitor by the presentation of a colorful Winnie the Pooh video, accompanied by an attention-getting phrase. Once the infant looked at the central monitor, the video was replaced by a static image of a checkerboard, and the artificial language began to play. The checkerboard remained on screen, and the language continued to play, for two minutes. At the end of this time, the attention-getting movie reappeared on the screen.

Once infants reoriented to the central monitor, the test phase begin. During this phase, 12 test trials were presented. In each test trial, a bisyllabic item was repeated, with pauses of 1.4 seconds between repetitions. For six of the test items, the bisyllabic item was characterized by high forward transitional probabilities, and low backward transitional probabilities. For the other six test items, the item was characterized by low forward transitional probabilities, and high backward transitional probabilities. A test trial began with the attention-getting movie playing on the central monitor drawing the infants' gaze forward. When the observing experimenter pressed a key indicating that the infant had fixated, the monitor displayed a video of a looming green ball on a black background, while the speakers began to play the test item. For as long as the infant maintained their gaze on the central monitor, the test trial continued, up to a maximum of 20 seconds. When the infant looked away for more than two consecutive seconds, the test trial ended and the attention-getting video reappeared on the central monitor.

Results

We analyzed infants' listening times for test items exemplifying an English-like (LoHi) grouping compared to their listening times for test items exemplifying non-English (HiLo) groupings. We modelled infant looking times using mixed-effect models using the open-source statistical package R (R Core Team, 2015). The full model included looking times per test item (LoHi and HiLo) as dependent variable and participants as random factors; we added age (7 and 13 month) and test item type (LoHi and HiLo) as interacted predictor variables. We found no significant main effect of age, ß = 0.14, $t = 0.13$, $p = 0.89$, indicating that the overall looking times (averaged across test items) for 7- and 13-month-old infants were of similar duration. There was, however, a significant main effect of test item, ß = 1.62, $t = 2.94$, $p < 0.01$. Collapsing across age, infants listened to LoHi test items for 14.4 s (*SD* = 3.3), and to HiLo test items for 13.6 s

(*SD* = 4.0). Importantly, this main effect of test item was qualified by a significant interaction between age and test item, ß = -1.58, $t = -2.02$, $p < 0.05$.

To characterize the interaction between age and test item, we performed planned t-tests assessing infants listening time to the test items at the two different ages. As predicted, 7-month-olds showed no preference between low-forward, high-backward (LoHi) test items and high-forward, low-backward (HiLo) transitional probability test items. The 7-month-old infants looked at LoHi test items for 13.7 s (*SD* = 3.4), and at HiLo test items for 13.6 s (*SD* = 4.4). A two-tailed t-test indicated that this difference was not significant, $t(24) = 0.07$, $p = .94$. These results suggest that 7-month-olds have not yet developed a preference for the test items (the LoHi items) that most closely match the predominant word order pattern in English.

By contrast, 13-month-olds did have a preference for the LoHi test items. These infants listen to the LoHi items for 15.13 s (*SD* = 3.1), and to the HiLo items for 13.5 sec (*SD* = 3.7). This difference was significant, $t(24) = 2.5$, $p < .05$. Unlike the 7-month-old infants, the 13-month-olds preferred those test items that are consistent with the predominant word order regularity in English. The difference between these two patterns of performance is significant, as indicated by the ANOVA's interaction term, suggesting that 13-month-olds are performing the task of grouping the fluent speech into smaller chunks differently than 7-month-old English-learning infants.

Experiment 2: Korean-learning infants

While the results from Experiment 1 are consistent with the suggestion that English-learning infants have learned about the predominant word order of their native language between 7 and 13 months, there is an alternative possibility. It may be the case that maturation causes the older infants to respond differently to the stimuli than younger infants, due to factors not related to linguistic experience. For example, many experiments have demonstrated age-related changes in infants' preferential listening to the same stimuli (e.g., Hunter & Ames, 1988). To assess this possibility, it is necessary to look at a group of infants exposed to a language with a different word order pattern than English; based on our previous work (Onnis & Thiessen, 2013), we chose to examine the preferences of Korean-learning infants. If the preference that 13-month-old infants demonstrated in Experiment 1 was due to some maturational change, 13-month-old Korean-learning infants should show the same pattern of preference. However, if their preference was driven by linguistic experience, Korean-learning infants should show the opposite pattern of preference, as Korean phrase headedness is predominantly opposite the typical English ordering.

Method

Participants.

Data from 26 13-month-old Korean-learning infants was included in the final sample. These infants participated in a research space in the Asan Medical Center in Seoul, South Korea. An additional 9 infants were excluded from the analysis for the following reasons: fussing or crying (5), failure to look at the test trials for an average of at least three seconds (2), and experiment error (2). All participants lived in the Seoul metropolitan area and were being raised in monolingual Korean speaking homes.

Stimuli and Procedure

The stimuli and procedure used in this experiment were identical to those used in Experiment 1.

Results

Unlike the English-learning infants in Experiment 1, Korean-learning 13-month-old infants did not show a preference for LoHi test items. On average, infants looked at HiLo test trials for 10.55 sec (SD = 4.32), and looked at LoHi trials for 9.69 sec (SD = 3.74). This difference was not statistically significant, $t(25) = 0.9$, $p = 0.36$. However, a mixed-effect model with looking times as dependent variable, participant as random variable, and experiment and test type as interacted predictor variables revealed two significant effects. The first was a significant effect of experiment, ß = -2.94, $t = -2.81$, $p < 0.01$. This is due to the fact that infants had shorter overall looking times in Experiment 2. The reason for this is not clear, but it may be due to individual differences in Research Assistant across the two laboratories, or due to the fact that the MBROLA phoneme set (Italian) used to create the language was differentially novel or interesting to English-learning and Korean-learning infants. More importantly, there was a significant interaction between Experiment and Test Item, ß = -2.50, $t = -2.26$, $p < 0.05$. This indicates that the English-learning infants preference in Experiment for LoHi test items was significantly different than the Korean-learning infants preference for test items in Experiment 2. This is consistent with the hypothesis that the preference that emerged for 13-month-old English-learning infants was not due to a maturational change, but instead dependent on their linguistic experience.

Intriguingly, while the Korean infants' preference was consistent with the expected word order in their native language, this preference was not as strong as that of the English-learning infants. This is consistent with our prior work with adults, where we have found that English speakers show a stronger directionality preference than Korean speakers. This, in turn, may be related to differences in the typicality of the word order patterns in English and Korean, with English showing a somewhat more predominant word order patterning (Onnis & Thiessen, 2013). If this is correct, it might take Korean-learning infants a somewhat longer time to develop a directionality bias. Regardless, the differences in English-learning and Korean-learning infants' preference for LoHi and HiLo word items suggests that the developmental pattern found in Experiment 1, in which a language-consistent word order preference emerged by 13 months, is related to English-learning infants' linguistic exposure.

General Discussion

All languages can be characterized, at least in part, in terms of their statistical regularities. Because of this, statistical learning may serve as a language-general avenue via which infants learn about the structure of their native language (e.g., Thiessen & Saffran, 2003). For example, sensitivity to statistical coherence allows infants to extract lexical items from linguistic input even in the absence of language-specific acoustic cues to word boundaries (e.g., Thiessen & Saffran, 2007). However, it is also the case that the statistical structure of the input differs across linguistic environments. In English, phrases are statistically coherent in a backward-going direction (as in "to school," where "to" does not strongly predict the upcoming element, but "school" more strongly retrodicts "to"), whereas in a language such as Korean, the opposite directionality of statistical coherence predominates (Onnis & Thiessen, 2013). The current results indicate that by 13 months, English-learning infants exposed to a continuous stream of syllables group those syllables together in ways that are consistent with the predominant statistical structure of their native language: they prefer groupings that are coherent in a backward-going (LoHi) direction over groupings that are consistent in a forward-going (HiLo) direction. Korean-learning infants do not show such a preference at 13 months, consistent with the hypothesis that the English-learning infants' preference is due to their experience with their native language, and suggesting that directionality preferences emerge at different rates as a function of the characteristics of the native language.

A crucial feature of the grammar used in this experiment is that both HiLo and LoHi sequences are equally frequent and coherent, making the grammar purposefully ambiguous as to which grouping should be preferred. Thus, the grammar functions as a litmus test for the learners' bias in attending to a particular directionality of statistical coherence. Our prior work has demonstrated that adults' linguistic background influences this bias. English-speaking adults are biased to group syllables in ways that respect backward-going statistical coherence, while Korean-speaking adults are biased to group syllables in ways that respect forward-going statistical coherence, consistent with the predominant phrase-level regularities in their respective native languages (Onnis & Thiessen, 2013). The results of the current experiment indicate that this adaptation to the structure of the native language occurs at some point between 7 and 13 months of age for English-learning infants, and perhaps somewhat later for Korean-learning infants. The fact that this bias emerges relatively early in development dovetails with an intriguing adaptive account whereby statistical learning preferences adapt to the characteristics of the native language in ways that should be beneficial for subsequent learning (e.g., Lew-Williams & Saffran, 2012; Thiessen & Saffran, 2007). This adaptive account of statistical learning does not necessarily imply a qualitative change in the cognitive mechanisms of statistical learning, but indicates that statistical learning minimally undergoes adaptation to the input of the environment.

Accounts of language that rely on statistical information are often described as relying on knowledge that is local and strongly lexically-based (e.g., Tomasello, 2000). Detractors of these usage-based approaches claim that properties such as the word order regularity investigated here are an abstract and general property, learned and represented via processes that are not based on individual words (e.g., Gervain, et al., 2008). Under this view, it would seem that generativist theories are more powerful, as they posit innate processes that operate over abstract representations. For example, once a representation related to headedness is triggered by the input, it would generalize to all linguistic input because the representation of headedness is not linked to any specific lexical item. Our results suggest an alternative account for infants' successful generalization: that statistical learning itself adapts to the characteristics of the linguistic input. Because infants in this experiment were exposed to novel sequences of nonsense syllables, their emergent preference for statistical groupings cannot be linked to mere familiarity with specific lexical items they experience previously. Instead, these results suggest that statistical learning adapts to the input broadly at a systemic level. Thus, by adapting to the environment, statistical learning mechanisms may be responsible for biasing toward general tendencies in the input that learners can capitalize on to arrange and parse novel word order sequences.

As such, the developmental pattern seen in the current results are not only descriptively useful, but also informative about the processes that lead to infants' adaptation to the statistical structure of the native language. The lack of a preference at 7 months suggests that these grouping biases are primarily learned as the result of experience, as opposed to experience-independent biases that are subsequently modified over the course of experience. Further, the fact that experience alters grouping biases by 13 months of age for English-learning infants limits the kinds of experience that can plausibly be suggested to play a role. For example, it is implausible that these biases arise as a result of explicit syntactic knowledge, as 13-month-olds are unlikely to have such knowledge (e.g., Jolly & Plunkett, 2008; Kedar et al., 2006). Instead, it is likely that this knowledge emerges as a function of infants' experience with the word order of their native language, as infants are capable of learning about serial order from a young age (e.g., Dominey & Ramus, 2000; Saffran et al., 1996). An intriguing avenue for subsequent investigation is what this account predicts about infants who are raised in a bilingual environment. One possibility is that they favor the grouping strategy appropriate for one language, and apply it to both languages (cf. Cutler, Mehler, Norris, & Segui, 1992). Another possibility is that infants develop biases appropriate to both languages, and deploy them selectively in the appropriate linguistic environment (c.f. Werker, Byers-Heinlein, & Fennell, 2009). This is a prediction that we are currently investigating.

One question that cannot be directly addressed by these results is how statistical learning is altered by prior experience. A definitive answer to this question is impossible without a mechanistic account of the process of statistical learning, of which there are several (for review, see Frank et al., 2010; Thiessen, Kronstein, & Hufnagle, 2013). However, we believe that the answer is likely to involve attention, which appears to play an important role in discovering conditional statistical relations (e.g., Baker, Olson, & Behrmann, 2004; Toro, Sinnett, & Soto-Faraco, 2005). After exposure to word order regularities that exemplify a dominant direction of statistical coherence, infants may be sensitized to that directionality in subsequent learning. This is consistent with accounts of statistical learning that suggest that items can only be grouped together when they are held simultaneously in attention (e.g., Perruchet & Vinter, 1997; Baker et al., 2004). Learners exposed to a language like English may be biased to attend to a *prior* item when grouping one element with another, while learners exposed to a language like Korean may be biased to attend to a *subsequent* item. These biases are not absolute; for example, adults and infants can discover both forward-going and backward-going statistical regularities (e.g., Pelucchi et al., 2009; Perruchet & Desaulty, 2008). An important avenue for future research, one which may help to clarify the mechanistic underpinnings of these biases, will be to assess how pliable they are in response to input that is inconsistent with the learner's bias.

Future work should address a number of shortcomings in the current study. For example, the lack of preference in 7 month old infants exposed to English is could reflect infants extracting both word types and thus fail to prefer one group systematically. This could be tested by contrasting HiLo and LoHi groupings with novel unheard combinations composed of heard syllables. Likewise, that the developmental trajectory for Korean-learning infants appears to be more protracted would have to be tested with a sample of older Korean-speaking infants, who would have to show a clear preference for high forward transitional probability sequences.

While several open questions remain, the current results provide an important insight into the contribution of statistical learning to language development. Several prior results have demonstrated that statistical learning plays a role in infants' adaptation to the structure of their native language, such as the discovery of a language's phonemic inventory or the predominant phonological form of lexical items (Maye, Werker, & Gerken, 2002; Thiessen & Saffran, 2007). As our results demonstrate, statistical learning *itself* changes as a function of experience, as learners become biased to identify statistical clusters that follow the predominant structure of their native language. This adaptation is likely to be beneficial for subsequent learning in the same linguistic environment, as it highlights statistical coherence that is consistent with the structure of the native language. Conversely, this adaptation may impair learning in linguistic environments that are characterized by different distributional regularities. As such, these results are consistent with theoretical accounts suggesting that prior experience plays an important role in subsequent language learning outcomes. In relation to other existing accounts, our results are consistent with rational constructivist approaches (e.g., Xu & Kushnir, in press), as well as developmental systems approaches (e.g., Spencer et al., 2009) whereby general inductive biases may emerge as a function of the interaction between experience with the environment (linguistic in this case) and learning mechanisms. These approaches point to ways to overcome the traditional nativist-empiricist debate.

Acknowledgments

We are grateful to Sook Whan Cho and Hongoak Yun for facilitating data collection at Asan Hospital in Seoul, and Jae-un Kim and Young-In Lee for running Experiment 2. Support comes from a Start Up Grant and a Singapore Ministry of Education Tier 1 grant RG81/14 to L.O.

References

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

Baker, C. I., Olson, C. R. & Behrmann, M. (2004). Role of attention and perceptual grouping in visual statistical learning. *Psychological Science, 15,* 460-466.

Cohen, L.B., Atkison, D.J., & Chaput, H.H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology, 24*, 381-410.

Dominey, P.F., & Ramus, F. (2000). Neural network processing of natural language: I. Sensitivity to serial, temporal, and abstract structure of language in the infant. *Language and Cognitive Processes, 15*, 87-127.

Dutoit, T. (1997). An Introduction to Text-To-Speech Synthesis, Kluwer Academic Publishers, Dordrecht Hardbound, ISBN 0-7923-4498-7, 312 pp.

Frank, M.C., Goldwater, S., Griffiths, T., & Tenenbaum, J.B. (2010). Modeling human performance in statistical word segmentation. *Cognition, 117,* 107-125.

Graf Estes, K., Evans, J. L., Alibali, M. W. & Saffran, J. R. (2007.). Can infants map meaning to newly segmented words? *Psychological Science, 18*, 254-260.

Gervain, J., Nespor, M., Mazuka, R., Horie, R., & Mehler, J. (2008). Bootstrapping word order in prelexical infants. *Cognitive Psychology, 57,* 56-74.

Haspelmath, M., Dryer, M.S., Gil, D., & Comrie, B. (2005). *World Atlas of Language Structures*. Oxford, Oxford University Press.

Jolly, H.R., & Plunkett, K. (2008). Inflectional bootstrapping in 2-year-olds. *Language and Speech, 51,* 45-59.

Jones, J., & Pashler, H. (2007). Is the mind inherently forward looking? Comparing prediction and retrodiction. *Psychonomic Bulletin and Review*, *14,* 295-300.

Kedar, Y., Casasola, M., & Lust, B. (2006). Getting there faster: 18- and 24-month-old infants' use of function words to determine reference. *Child Development, 77*, 325-338.

Lew-Williams, C., & Saffran, J.R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition, 116,* 241-246.

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

Noble, C.F., Rowland, C.F., & Pine, J.M. (2011). Comprehension of argument structure and semantic roles: Evidence from infants and the forced-choice pointing paradigm. *Cognitive Science, 35*, 963-982.

Onnis, L., & Thiessen, E.D. (2013). Language experience changes subsequent learning. *Cognition,* http://dx.doi.org/10.1016/j.cognition.2012.10.008.

Pelucchi, B., Hay, J.F., Saffran, J.R. (2009). Learning in reverse: Eight-month-old infants track backwards transitional probabilities. *Cognition*, *113*, 244-247.

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, *36*, 1299-1305.

Perruchet, P. & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*, 246-263.

Team, R. C. (2015). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2015.

Romberg, A.R., & Saffran, J.R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*.

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition, 81*, 149-169.

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

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

Spencer, J.P., Blumberg, M.S., McMurray, B., Robinson, S.R., Samuelson, L.K., & Tomblin, J.B. (2009). Short arms and talking eggs: Why we should no longer abide the nativist-empiricist debate. *Child Development Perspectives*, 3:2, 79-87.

Thiessen, E.D., & Erickson, L.C. (in press). The statistical learning approach to word segmentation. To appear in T. Mintz (Ed.), *Statistical Approaches to Language Acquisition*.

Thiessen, E.D., Kronstein, A.T., & Hufnagle, D.G. (in press). The extraction and integration framework: A two-process account of statistical learning. *Psychological Bulletin.*

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

Thiessen, E. D. & Saffran, J. R. (2007). Learning to learn: Infants' acquisition of stress-based strategies for word segmentation. *Language Learning and Development, 3,* 73-100.

Thompson, S.P., & Newport, E.L. (2007) Statistical learning of syntax: the role of transitional probability. *Language Learning and Development*, *3*,1–42.

Toro, J.M., Sinnett, S., & Soto-Faraco, S. (2005). Speech segmentation by statistical learning depends on attention. *Cognition, 97*, B25-B34.

Werker, J.F., Byers-Heinlein, K., & Fennell, C.T. (2009). Bilingual beginnings to learning words. *Philosophical Transactions of the Royal Society B, 364*, 3649-3663.

Xu, F. & Kushnir, T. (in press). Infants are rational constructive learners. *Current Directions in Psychological Science*.

Figure Captions

Figure 1. Description of the Markovian grammar chain used to generate the familiarization stream heard by infants. Arrows represent the transitional probabilities between elements of the grammar. The numbers above each arrow represent the forward-going transitional probability between two elements (e.g., a 33% chance that X is followed by D), while the numbers below each arrow represent the backward-going transitional probability (e.g., a 100% chance that D is preceded by X).

Figure 2. Looking times to HiLo (Korean-like) and LoHi (English-like) test items for 7- and 13-month-old infants in English-exposed and Korean-exposed infants. Error bars indicate standard error.

Figure 1

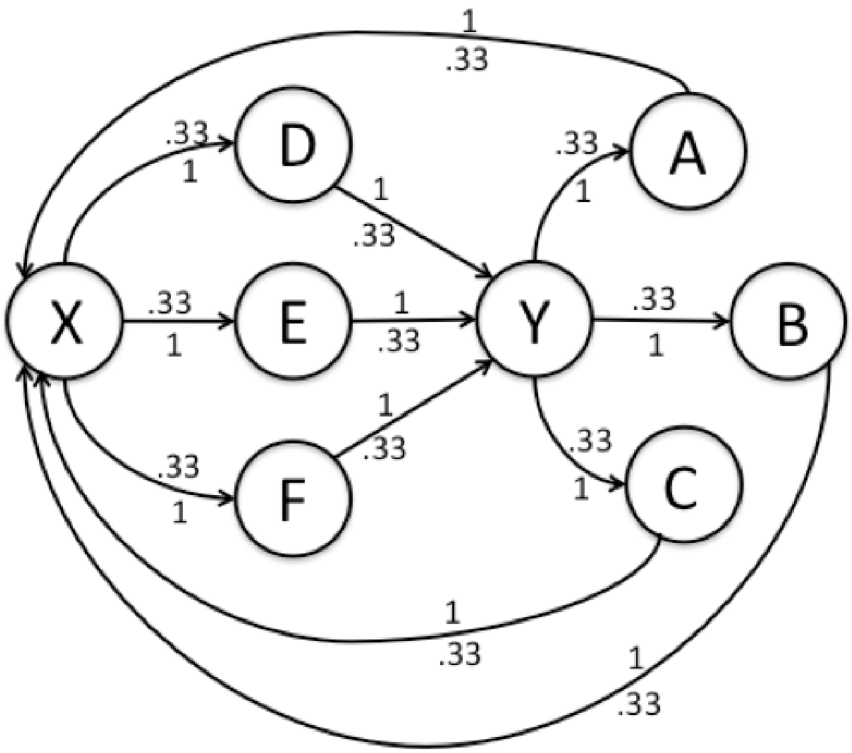
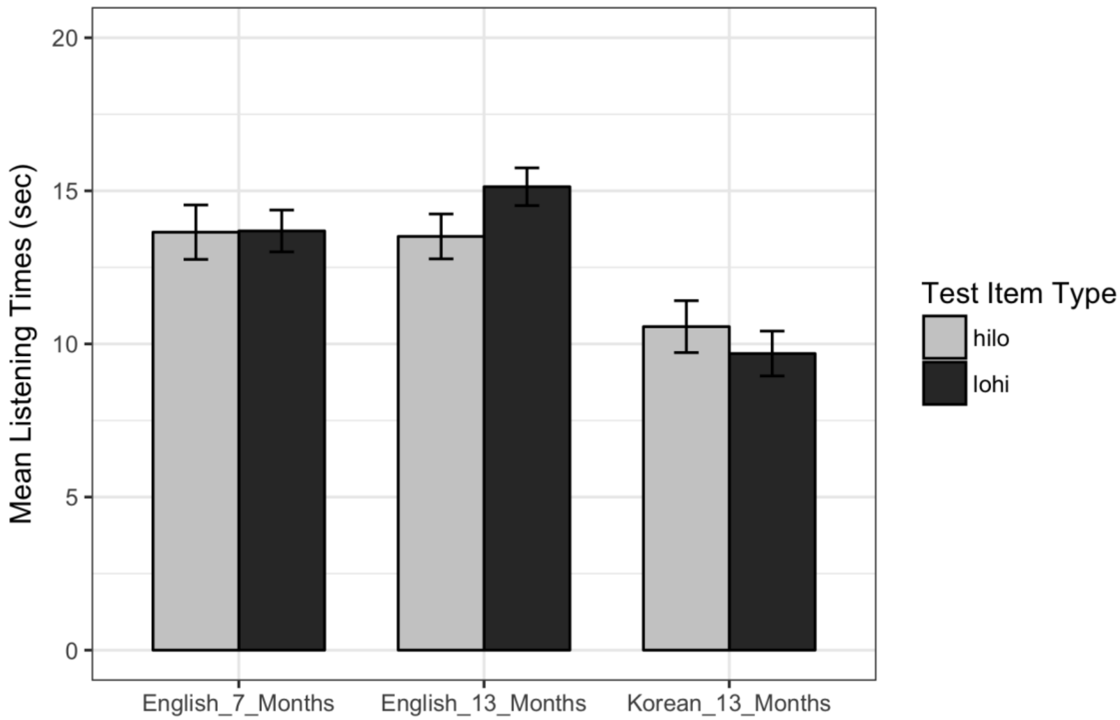


Figure 2


Mean Listening Times (sec)
0
5
10
15
20
English_7_Months
English_13_Months
Korean_13_Months
Test Item Type
hilo
lohi