Early developing syntactic knowledge influences sequential statistical learning in infancy

Erik D. Thiessen a,⇑, Luca Onnis b, Soo-Jong Hong c, Kyung-Sook Lee d

a Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA
b Program in Linguistics and Multilingual Studies, Nanyang Technological University, Singapore 639798, Singapore
c Asan Medical Center, University of Ulsan College of Medicine, Seoul 05505, South Korea
d Department of Rehabilitation, Hanshin University, Seoul, South Korea

Article history:
Received 10 July 2017
Revised 15 April 2018

Keywords:
Statistical learning
Linguistics
Head
Parameter
Cross-linguistic differences
Phrase structure

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. In the experiments, 7-month-old infants showed no bias for forward-going or backward-going regularities. By 13 months, however, English-learning infants favored 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.

© 2018 Published by Elsevier Inc.

Introduction

Statistical information has been argued to play an important role in language development (e.g., Romberg & Saffran, 2010; Saffran, Aslin, & Newport, 1996; Thiessen & Erickson, 2014). 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, 2013). 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).

However, accounts of language development and processing that rely on statistical information are often critiqued as relying on knowledge that is local and strongly lexically based (e.g., Tomasello, 2000). Detractors of these statistical approaches claim that many crucial linguistic properties, such as word order, are abstract and represented independently of individual words (e.g., Gervain, Nespòr, Mazuka, Horie, & Mehler, 2008). This perspective emphasizes the generativity of linguistic structures; once infants have identified an abstract property of linguistic structure, they can generalize the property to novel input and produce novel constructions. This perspective often describes language learning in terms of discovering abstract symbolic rules. For example, once infants have learned that the past tense of the English verb involves adding \textit{ed} (e.g., kick transforms to kicked), they can apply this rule to any verb (Pinker & Ullman, 2002).

We have proposed an alternative to this perspective, which is that statistical learning itself adapts to the structure of linguistic input in ways that lead learners to have expectations about novel subsequent input (Onnis & Thiessen, 2013). From this perspective, what infants are learning as they acquire a language is not a set of symbolic rules. Rather, they are absorbing examples from the input and generalizing on the basis of the statistical structure of those examples (e.g., Thiessen & Saffran, 2003). This allows statistical learning to “specialize” to the structure of the input, for example, by focusing attention on cues that have been informative in past experience. That is, although statistical learning is an early developing and potentially universal cue to linguistic structure, different languages are characterized by different statistical regularities. Statistical learning may adapt to these regularities in ways that make learners better prepared for subsequent learning in that language. Note that this contrasts with the view that statistical learning (and perhaps implicit learning more generally) is a stable trait that does not show substantial change over time (e.g., Arciuli & Simpson, 2012).

The claim that statistical learning adapts to the structure of the input suggests that it should be possible—even likely—to observe differences in statistical learning as a function of the kinds of regularities that are informative across different languages. One difference in the statistical regularities across languages occurs as a consequence of the predominant directionality of phrase structure 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, whereas 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 which shapes reliably occurred both 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 (e.g., 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), whereas 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), whereas 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 (2013) 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 statistical learning itself. Sensitivity to backward-going regularities may be more adaptive to learners in an English environment than to 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, whereas 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 is 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 over 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 early in life, such that statistical learning can help infants to better discover the structural properties of the linguistic input. An alternative possibility, however, is that this kind of directionality bias emerges relatively late as a byproduct of established and potentially abstract syntactic knowledge rather than as an integral adaptation necessary to acquire that knowledge. In these experiments, we attempted to track the origins of sensitivity to syntactic structure by assessing the grouping biases of infants at 7 and 13 months. At 7 months, infants 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).

Both theoretical perspectives suggest that 7-month-old infants should not have a directionality bias. However, their predictions for 13-month-old infants are potentially divergent. If infants are acquiring symbolic rules, we should not expect them to show evidence that they process linguistic input differently as a function of these rules until infants have learned the rules—that is, until they show evidence of phrase structure knowledge near the end of the second year of life (e.g., Jolly & Plunkett, 2008). By contrast, if infants are learning from examples, and generalizing on the basis of those examples, we might expect to see evidence of sequential word order knowledge as soon as infants have learned enough words in the native language to discover that there are predictable regularities in the order in which those words occur. As such, the statistical learning perspective predicts a somewhat earlier emergence of order-based expectations.
Experiment 1: English-learning infants

Method

Participants

Data from 25 7-month-old infants and 25 13-month-old infants were included in the final sample. An additional 14 infants (6 7-month-olds and 8 13-month-olds) were excluded from the analysis for the following reasons: fussing or crying (n = 8), failure to look at the test trials for an average of at least 3 s (n = 3), and experiment error (n = 3). All participants were recruited from a database of infant births within the state of Pennsylvania and whose parents reported that their children were exposed to at least 90% English, had no history of hearing problems, and did not show evidence of language delay.

Stimuli

The stimuli used in this experiment were based on those used in Onnis and Thiessen (2013). Infants were familiarized with a sequence of consonant–vowel (CV) syllables based on a Markovian grammar chain organized around eight 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 Fig. 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, whereas 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.

Fig. 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 number above each arrow represents the forward-going transitional probability between two elements (e.g., a 33% chance that X is followed by D), whereas the number below each arrow represents the backward-going transitional probability (e.g., a 100% chance that D is preceded by X).
The eight symbols underlying the grammar were mapped onto eight different synthesized CV syllables that occur in both English and Korean (fu, shae, zi, ra, ni, bu, ti, and ge). To control for item biases, each infant heard a familiarization with a different randomized mapping (i.e., for one infant syllable A = fu, whereas for a different infant syllable A = ti). These syllables were synthesized by the MBROLA synthesizer (Dutoit, 1997). Each syllable lasted 340 ms (with a 260-ms vowel), and there were no pauses between syllables. The familiarization stream lasted for 3 min 11 s (191 s), during which time infants heard 350 syllables (see Fig. 2).

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 test 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 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 and were 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 Habit X software (Cohen, Atkison, & 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 infant’s 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 checkboard and the artificial language began to play. The checkboard remained on the screen, and the language continued to play, for 3 min 11 s. At the end of this time, the attention-getting movie reappeared on the screen.

Once the infant reoriented to the central monitor, the test phase began. During this phase, 12 test trials were presented. In each test trial, a bisyllabic item was repeated, with pauses of 1.4 s between repetitions. For six of the test items, the bisyllabic item was characterized by high forward transitional

![Fig. 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 errors.](image-url)
probabilities and low backward transitional probabilities (HiLo test items). For the other six test items, the item was characterized by low forward transitional probabilities and high backward transitional probabilities (LoHi test items). All LoHi and HiLo test items were synthesized using identical parameters as were used to create the familiarization stream. A test trial began with the attention-getting movie playing on the central monitor drawing the infant’s 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 his or her gaze on the central monitor, the test trial continued up to a maximum of 20 s. When the infant looked away for more than a consecutive 2 s, indicating a lack of interest in that specific test item, 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 with their listening times for test items exemplifying non-English (HiLo) groupings. We modeled infant looking times using mixed-effect models with the open-source statistical package R (R Core Team, 2015). The full model included looking times per test item (LoHi or HiLo) as dependent variable and participants as random factors; we added age (7 or 13 months) and test item type (LoHi or HiLo) as interacted predictor variables. The experimental logic is as follows. If infants have failed to learn any words from the input, they should find LoHi and HiLo test items to be equally compelling and listen to them for an equal amount of time. By contrast, if infants have segmented words from the input, they should listen differentially to HiLo versus LoHi items.

We found no significant main effect of age, $\beta = 0.14, t = 0.13, p = .89$, indicating that the overall looking times (averaged across test items) for 7- and 13-month-olds were of similar duration. There was, however, a significant main effect of test item, $\beta = 1.62, t(48) = 2.94, p < .01$, Cohen’s $d = 0.22$. 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, $\beta = -1.58, t = 2.02, p < .05$.

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-olds 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, $t(24) = 0.07, p = .94$, Cohen’s $d = 0.02$. These results indicate that 7-month-olds do not show 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 listened to the LoHi items for 14.9 s ($SD = 3.3$) and to the HiLo items for 13.4 s ($SD = 3.6$). This difference was significant, $t(24) = 2.50, p < .05$, Cohen’s $d = 0.43$. Unlike the 7-month-old, 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 interaction term of the analysis of variance (ANOVA), suggesting that 13-month-olds perform the task of grouping the fluent speech into smaller chunks differently than 7-month-old English-learning infants.

Of note is the fact that English-learning infants showed a familiarity preference: that is, they listened longer to test items that are consistent with their native language. This is not something we could have predicted a priori because infants in this type of experiment sometimes show familiarity preferences and sometimes show novelty preferences (e.g., Thiessen, Hill, & Saffran, 2005). However, one trend that emerges across the literature on direction of preference is that infants are more likely to show a familiarity preference in difficult learning tasks and more likely to show a novelty preference in easier learning tasks (for more extensive discussion, see Hunter & Ames, 1988). One possibility is that the presence of conflicting potential segmentations in the input (LoHi vs. HiLo) caused the task to be somewhat difficult. We return to this idea of conflict, and how it may be shaping responses across developmental time, in the General Discussion.
Experiment 2: Korean-learning infants

Whereas 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 responses of Korean-learning infants. If the preference that 13-month-olds 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 a different pattern of preference because Korean phrase headedness is predominantly opposite the typical English ordering.

Method

Participants

Data from 26 13-month-old Korean-learning infants were 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 (n = 5), failure to look at the test trials for an average of at least 3 s (n = 2), and experiment error (n = 2).

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-olds did not show a preference for LoHi test items. On average, infants looked at HiLo test trials for 10.55 s (SD = 4.32) and looked at LoHi trials for 9.69 s (SD = 3.74). This difference was not statistically significant, t(25) = 0.90, p = .36, Cohen’s d = 0.18. However, a mixed-effect model with looking times as dependent variable, participant as random variable, and experiment and test type as predictor variables revealed two significant effects. The first was a significant effect of experiment, β = −2.94, t = 2.81, p < .01. This is due to the fact that Korean-learning infants in Experiment 2 had shorter overall looking times than English-learning infants in Experiment 1. The reason for this is not clear, but it may be due to individual differences in research assistants or procedure 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 important, there was a significant interaction between experiment and test item, β = −2.50, t = 2.26, p < .05. This indicates that English-learning infants preference (for LoHi) test items in Experiment 1 was significantly different from the Korean-learning infants (lack of) 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 was dependent on their linguistic experience. If it were due to maturational change, we would expect that 13-month-olds would show the same pattern of preference regardless of their linguistic experience.

One possible alternative explanation for these results is that there are differences between the demographics of our English and Korean samples or differences in the way in which the procedure is implemented in the two languages/laboratories. Although we cannot rule this out on the basis of our current results, it is intriguing to note that the Korean infants’ preference was numerically (but not significantly) in the direction of test items that follow the expected word order in their native language. As indicated by the lack of significance, this bias toward test items showing Korean-
typical directionality (HiLo items) was not as strong as that of the English-learning infants' bias toward test items showing English-typical directionality (LoHi items). This is consistent with our prior work with adults, where we have found that English speakers show a stronger directionality bias 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 because the input to which they are exposed does not show as regular a pattern of directionality. Regardless, the differences between English-learning and Korean-learning infants' preference for LoHi and HiLo test 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 like 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 coherent 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 biases may emerge at different rates as a function of the characteristics of the native language.

A crucial feature of the grammar used in these experiments is that both HiLo and LoHi sequences are equally frequent and coherent, making the grammar purposefully ambiguous as to which grouping should be favored. 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, whereas 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 for English-learning infants and potentially somewhat later for Korean-learning infants. The fact that this bias emerges relatively early in development suggests that statistical learning rapidly adapts 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).

Accounts of language that rely on statistical information are often described as relying on knowledge that is local and lexically based (e.g., Tomasello, 2000). Detractors of these statistical 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 because 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 directionality bias for statistical groupings cannot be linked to familiarity with specific lexical items. 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 is 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 between forward-going and backward-going test items at 7 months suggests that directionality biases are primarily learned as the result of experience, as opposed to experience-independent biases that are subsequently modified over the course of experience. Furthermore, the fact that experience alters grouping biases by 13 months 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 because 13-month-olds are unlikely to have such knowledge (e.g., Jolly & Plunkett, 2008; Kedar et al., 2006). Instead, it is likely that a directionality bias emerges as a function of infants’ experience with the word order of their native language because 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 related to infants who are raised in a bilingual environment. One possibility is that bilingual infants 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 (cf. 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 adapts. That is, how is statistical learning 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 reviews, see Frank, Goldwater, Griffiths, & Tenenbaum, 2010; Thiessen et al., 2013). One potential explanation for the age-related change in directionality bias that we observed in English-learning infants is that the process of statistical learning changes over time, such that the learning mechanism itself no longer behaves in an identical way across speakers of different languages. However, we believe that the answer is likely to involve not changes in the mechanism itself but rather changes in the aspects of the input to which learners attend. Attention plays 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 suggesting that items can be grouped together only when they are held simultaneously in attention (e.g., Baker et al., 2004; Perruchet & Vinter, 1998). Learners exposed to a language like English may be biased to attend to a prior item when grouping one element with another, whereas learners exposed to a language like Korean may be biased to attend to a subsequent item. That is, when presented with a stimulus like A–B–C, where B is (for whatever reason) a salient element, English-learning infants may create a chunk like A–B, whereas Korean-learning infants may create a chunk like B–C.

An alternative possible explanation is that what changes about statistical learning is not the uptake of information (gated by attention) but rather the response to that information (perhaps mediated by executive control). That is, both HiLo and LoHi segmentations are equally supported by the input, but they are in conflict with each other. Older infants may be better able to resolve this conflict than younger infants, leading to more consistent performance at test. This would be consistent with prior work suggesting that response inhibition plays an important role in at least some aspects of language learning that involve conflicting cues to structure (e.g., Weiss, Gerfen, & Mitchell, 2009). An important avenue for future research, one that may help to clarify the mechanistic underpinnings of these biases, will be to assess how pliable learners are in response to input that is inconsistent with their prior biases. 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 & Desauty,
The extent to which learners can identify structure that contradicts these biases, and the conditions under which they do so, may provide some clarity about the origin and nature of these biases. Although 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, the outcomes of statistical learning change 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 because 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, 2013) 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 in which 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, South Korea, and Jae-un Kim and Young-In Lee for running Experiment 2. Support came from a Start Up Grant and a Singapore Ministry of Education Tier 1 grant (RG81/14) to L.O.

References