Language-induced Biases on Human Sequential Learning

Luca Onnis (lucao@hawaii.edu)
Department of Second Language Studies & Center for Second Language Research
University of Hawai‘i at Manoa, 1890 East-West Road, Honolulu, Hawaii 96822 USA

Erik Thiessen (thiessen@andrew.cmu.edu)
Department of Psychology
Carnegie Mellon University, Baker Hall, Pittsburgh, Pennsylvania 15213 USA

Abstract

What are the effects of experience on subsequent learning? We explored the effects of language-specific word order knowledge on the acquisition of sequential conditional information. Korean and English adults were engaged in a sequence learning task involving three different sets of stimuli: auditory linguistic (nonsense syllables), visual non-linguistic (nonsense shapes), and auditory non-linguistic (pure tones). The forward and backward probabilities between adjacent elements generated two equally probable and orthogonal perceptual parses of the elements, such that any significant preference at test must be due to either general cognitive biases, or prior language-induced biases. We found that language modulated parsing preferences with the linguistic stimuli only. Intriguingly, these preferences are congruent with the dominant *word order* patterns of each language, as corroborated by corpus analyses. These findings suggest that mechanisms of statistical sequential learning are implicated in language, and experience with language may affect cognitive processes and later learning.

Keywords: corpus analyses; experience-dependent learning; implicit learning; prediction; retrodiction; second language acquisition; sensitive periods; sequential learning; statistical learning; transitional probabilities; word order; linguistic typology.

Introduction

Statistical information has been argued to be an important cue to linguistic structure. For example, sounds within a word are more predictable than sounds across boundaries, which may help infants discover words in fluent speech. Because this type of statistical information is present in all languages, statistical information may be a particularly important cue early in development, one that can be used without requiring prior experience with the native language (e.g., Thiessen & Saffran, 2003).

But while statistical learning may be a universal cue to linguistic structure, it is also the case that the statistical structure across languages differs. If statistical learning fails to adapt to these differences, it is unlikely to be an optimal learning strategy. While much research has examined how statistical learning helps learners adapt to the structure of their native language (e.g., Maye, Werker, & Gerken, 2002; Thiessen & Saffran, 2007), it is unknown whether statistical learning itself adapts to the characteristics of the native language. In this series of experiments, we ask whether experience with language alters the kinds of statistical regularities that learners detect in ways that are consistent with the predominant statistical structure in their native language.

Prediction and retrodiction in sequential learning

Recent studies have shown that learners can exploit both predictive and retrodictive relations, operationalized as forward and backward transitional probabilities respectively. For instance, Jones & Pashler (2007) showed participants sequences of shapes governed by probabilistic relations, and then asked them to choose which shape reliably came after a probe shape (prediction test) or before a probe shape (retrodiction test). In experiments where forward and backward probabilities were made informative, they found that both prediction and retrodiction were used effectively for recalling memories. In a similar experiment using a continuous sequence of nonsense syllables, Perruchet & Desaulty (2008) found that participants perceived word boundaries based on backward transitional probabilities as well as forward probabilities equally well. Likewise, Pelucchi, Hay, & Saffran (2009) provided evidence that infants can track backward statistics in speech.

The three studies above tested cases in which forward and backward probabilities were never in conflict. Rather, each cue was made maximally informative in a given experiment, while the other was made uninformative. Yet in naturalistic circumstances, prediction and retrodiction may need to be effectively combined, as when learning the word order of a language. In this respect, a comparison between English and Korean seems particularly appropriate because several word order relations of English are reversed in Korean.

Prediction and retrodiction in natural languages: Typology and word order tendencies.

In English, the head elements in a phrase come first, while in Korean the head follows the phrase (e.g., ‘[Door-OBJECT] [close-IMPERATIVE]’ = ‘[You close] [the door]’, where square brackets indicate phrase groupings). The English sentence “I saw him go there” is glossed as “I him there go saw”. Likewise “Give me the ball” is glossed as “Ball me give”; “Let’s go get some food” is glossed as “Food get go let’s”. Thus, frequent constructions such as transitives, imperatives, and exortatives in English have a reversed word-order in Korean.

English is also prepositional (‘to school’), while Korean is postpositional (‘school to’). We conjectured that since the linear word order relations in English and Korean are often specular, different sets of expectancies for predictive and retrodictive dependencies may emerge during learning each specific language: for example, in English the predictive probability of the noun *school* following the preposition *to* (*p(school|to)*) is lower than the retrodictive probability of *to* preceding *school* (*p(to|school)*), and vice-versa in Korean.

To corroborate these intuitions, we first conducted large-scale corpus studies of the two languages. Then an ambiguous artificial grammar containing different conflicting cues was presented to Korean and American speakers in a sequence learning task. This grammar is unlearnable from surface statistical regularities alone unless previous biases that favor predictive and retrodictive cues are in place in the system before learners start the experiment. Thus, the grammar was used as a litmus test for assessing potential prior biases on learning. To establish whether the biases should be attributable to experience with language or general sequential biases, we tested sequential learning across speakers of the two languages with opposite word order, and in three different modalities: auditory linguistic (speech), visual non-linguistic (abstract shapes), and auditory non-linguistic (pure tones).

We reasoned that if sequential learning mechanisms are directly involved in language acquisition and processing, as it has been put forth, these mechanisms should show language-specific biases effects when adults are engaged in sequence-learning tasks with speech-like stimuli. We further conjectured that if the bias is due to language experience – and not to some more general temporal processing bias – adult participants engaging in the same sequence task using non-linguistic stimuli (visual shapes and auditory tones), should behave consistently *irrespective* of their language background. Another possibility is that sequential learning mechanisms are shared among perceptual modalities and exhibit inherent a priori biases for sequences of stimuli, for example for predictive relations. In this latter case, we would expect a consistent pattern of preference across languages and modalities. Finally, there may be patterns of preference consistent across languages but differing by modality, in which case any effect may be attributable to modality-specific biases.

Corpus Analyses

We quantified the hypothesis that word order tendencies in Korean and English generate *opposite* patterns of predictive and retrodictive conditional probabilities that signal phrase cohesiveness and syntactic information.

Corpora

For English we used the SUSANNE Corpus, consisting of 130,000 words of published American English annotated with part-of-speech (POS) and syntactic information (Treebank)[1]. For Korean we sampled the freely available Sejong Corpus, with a syntactically annotated subcomponent containing 800,000 words[2]. For each sentence in the corpus, we derived unigram and bigram frequency counts as well as forward and backward transitional probability statistics between any two words. Ngram frequencies in English were sampled from the Google Ngram database for the year 2000 (~4 million unigrams and 60 million bigrams). Korean ngram token frequencies were summed over three different corpora: Sejong, HC Korean (55 million unigrams), and KAIST (70 million unigrams) in order to obtain reliable frequency counts. Finally, for each word pair in a sentence we derived the level of syntactic boundary inherent in the syntactic annotation.

Corpus Measures of phrase cohesiveness

Independent Variable I: Ngram frequencies Because several psycholinguistic studies have shown that humans are sensitive to the logarithm of event frequencies as opposed to raw frequencies, it is customary to consider log-frequencies as opposed to raw frequencies. The logfrequency of a sequence of two words (logBigram) can be taken as an approximation of phrase cohesiveness, following Tremblay et al., 2011. The logfrequency of each individual word can also be useful in predicting headedness, as higher frequency words tend to be heads of phrase constituents (Gervain et al., 2008).

Independent Variable II: Conditional probability Another way to measure how likely two words are to occur together is to look at a word and estimate what words are likely to follow it. The likelihood of a given word following is the forward probability of the word pair. For example, for the sequence ‘in Sapporo’:

$$\text{fwdTP(Sapporo|in)}= \frac{freq(in\_Sapporo)}{freq(in)}$$

The calculation can also be computed in the opposite direction. That is, examine a word and estimate what words are likely to precede it. The likelihood of a given word preceding is known as the backward probability between the two words.

$$\text{bckTP(in|Sapporo)}= \frac{freq(in\_Sapporo)}{freq(Sapporo)}$$

For example, suppose the word “in” occurs 2853 times in the corpus, but the word “Sapporo” occurs only 9 times and the sequence of words “in Sapporo” occurs 3 times. Since the word “in” occurs 2853 times, and only 3 of those times with the word Sapporo, this pair of words has a very low forward probability (3/2853), However, if we examine the pair from the opposite direction, we see that three out of the nine times the word “Sapporo” appears, it is preceded by the word “in.” Thus, the backward probability is 3/9, or .33.

Dependent Variable: Phrase structure cohesiveness To estimate the informativeness of transitional probabilities and frequencies in parsing at the constituent level, we followed Johnson (1965): Sentences from the tree-tagged Susanne corpus and Sejong corpus were divided up into phrasal constituents. For every word pair transition considered

[1] Available at http://www.grsampson.net/Resources.html

[2] Available at http://rocker.snu.ac.kr:8080/search

linearly from left to right, it is possible to rank-order the level at which a constituent node for that transition occurs. For example, in "[[[The house] [across [the street]]] [is burning]]" the highest node is at transition 5 (street_is), followed in rank by transition 2 (house_across), then transition 3 (across_the). Finally, transitions 1, 4, and 6 are tied on the same rank. Using the syntactically annotated corpora, every bigram in each sentence can be assigned a syntactic rank, following the example above.

Typological studies of the world languages have uncovered important correlations in the linear order of the constituents in different subdomains. For instance, the constituent order of a clause (the relative order of subject, object, and verb); the order of modifiers (adjectives, numerals, demonstratives, possessives, and adjuncts) in a noun phrase; and the order of adverbials and adpositions. Most languages appear to have a preferred word order that is usually most frequent. This ordering of constituents is often represented as a tree where branches can be divided into other minor branches, which may also branch in turn. English is often described as a right-branching language, because it tends to place dependents after the head words. Adjectives follow nouns, direct objects follow verbs, and adpositions are prepositional. This type of branching is also known as head-first order. Left-branching languages, like Korean and Japanese, exhibit the opposite tendency, that is, they tend to place the head element of a phrase to the left. Objects appear to the left of verbs, sentences appear to the left of subordinating conjunctions and noun phrases appear to the left of prepositions (which, for this reason, are often called postpositions in these languages). Since postpositions come after the noun in left-branching languages, our example phrase, "in Sapporo," would actually be in the opposite order, "Sapporo in."

We were interested in assessing whether regularities in the word order of English and Korean engender language-specific probabilistic expectations between sequences of adjacent words. For example, considering the sequence "in Sapporo" the forward probability is expected to be low, arguably because many words can follow "in" (Rome, New York, summer, me, the, lovely, etc.). Conversely, the backward probability should be high, because only a few words are expected before "Sapporo" (to, in). Thus a pattern of "forward low-backward high" probability (*LoHi* for short) is expected to run as a sentence unfolds in English. If we express this combined pattern in mathematical terms as the difference between the forward and the backward probability (TPdiff), for any word pair in a sentence we should expect a positive larger TPdiff to indicate a more cohesive phrase unit in the language. Thus, the TP difference could be taken to predict the level of syntactic constituency between any word pair in the syntactically tagged corpora. Notably, for Korean the pattern of transition probabilities is expected to be reversed. For "Sapporo in", the forward probability should be high relative to the backward probability. Thus, *HiLo* patterns are expected to run as a sentence unfolds in Korean. Using the same differential measure (TPdiff) between forward and backward probability, this time we can expect a large negative TPdiff to predict more cohesive phrase units in the syntactically-tagged Korean corpus. These predictions "by example" are by no means granted for the whole language. In the syntactic literature it has long been noted that the right-branching/left-branching dichotomy may not hold for an entire language, and in the case of English it is not fully consistent even at the phrasal level (for instance for the word ordering within a Noun Phrase, see Cook and Newson 2007). Thus it is important to evaluate whether these probabilistic biases are significantly and robustly correlated with word order across the two language corpora.

Results

Ordinal logistic regressions were run to predict the syntactic tree level between any two members of word pairs (word1, word2, e.g., "in Sapporo") in a sentence. The syntactic tree level was obtained from the syntactic parsing provided in the Susanne and Sejong corpora (henceforth English and Korean corpus respectively). Therefore, the tree level (henceforth *tree*) was the dependent variable to be predicted by the regression models. The following predictors were considered in the following order: log frequency of each bigram, log-frequency of first word, log-frequency of second word, forward probability, backward probability. Because node levels above 6 were very infrequent in both corpora, we considered the first six node levels, accounting for 99.5% of bigrams in the English corpus (109,861 bigrams entered in the analyses) and for 98.1% of bigrams in the Korean corpus (22,382 bigrams entered).

Model fit For each corpus, ordinal logistic regression models with different complexity were fitted and compared for goodness of fit. The null model contained no predictors, then increasingly complex models added logBigram, LogFrequency of first word, LogFrequency of second word, Forward Probability, and Backward Probability as predictors. Analyses of deviance between each increasingly complex model indicated that including all predictors except LogFrequency of second word increased the fit of each regression model significantly with respect to the previous less complex model by reducing deviance. This result held for both corpora. Thus, in the following analyses the log-frequency of the second word was excluded as a predictor. All other variables contributed significantly to predict the level of syntactic node between any two adjacent words in a sentence in the corpus. Using the lrm function in R we were also able to assess the goodness of fit of the models. As the p-values of the G test statistics are 0 in both language models, the null hypothesis can be rejected that there is no overall significant relationship between the dependent variable tree and the independent variables. The predictive ability of the model can also be measured using C, an index of concordance between the predicted probability and the observed response. (if C=0.5 the predictors are random, when it is 1, prediction is perfect). Since C=0.71 for English and C=0.64 for the Korean corpus, we have confidence that both models have moderate to strong predictive capacity. Somer's Dxy is a rank correlation between predicted probabilities and observed responses. It ranges between 0 (randomness) and 1 (perfect prediction). Since Dxy=0.43 (English) and Dxy=0.30 (Korean) we have again confidence that both language models have moderate predictive capacity. Kendall's Tau-a rank correlations also assess the correlations between all predicted probabilities and observed response. Here Tau-a=0.26 (English) and Tau-a=0.16 (Korean).

English Corpus We were particularly interested in the *coefficient sign* of the independent variables across the two corpora. For English, the coefficients for *logBigram* were consistently negative, indicating that the more frequent bigrams are associated with tighter phrase boundaries. Thus, the frequency of a bigram can be used to partially predict phrase cohesiveness, in accord with the psycholinguistic evidence obtained by Tremblay et al. (2011). Most importantly for the hypothesis being tested in this study, lower forward probabilities *and concurrently* higher backward probabilities (a *LoHi* pattern) were associated with higher phrase cohesiveness, as indicated by a positively valued coefficient for TPdiff. Remember that low syntactic levels indicate that the word pair tends to be occurring within the same phrase, or a across a transition that is at a lower level up the syntactic tree. In addition, higher frequency of first words was associated with more phrasal cohesiveness, in accord with Gervain et al. (2008).

Korean Corpus LogBigram frequency was negatively associated with syntactic depth, indicating again that more cohesive phrases tend to be more frequent. Crucially, the coefficient for TPdiff was now positive (and reversed with respect to English), indicated that higher forward probabilities and lower backward probabilities (a *HiLo* pattern) were associated with higher phrase cohesiveness.

Summary When comparing English and Korean, the patterns of probability that support syntactic parsing are clearly reversed in the two languages, as originally predicted. In particular, phrase cohesiveness correlates with a LoHi pattern of transition probabilities in English, and with a HiLo pattern in Korean. Below we ask whether these language-specific patterns of probabilities are a source of experience-induced bias when learners group novel stimuli in a sequence learning task.

Experiment 1

The corpus analyses above provide a rationale for our main hypothesis, namely that the predictive regularities most consistently experienced in one's native language, may impose processing biases on human sequential learning (Table 2). A speech-synthesized stream of syllables was constructed so that two mutually exclusive sets of syllable groupings could possibly be extracted, according to either a bias for a *LoHi* probability pattern (as in English '*to school*', Table 2), or a *HiLo* probability pattern (as in Korean '*school to*', Table 2). Because the two sets were equally frequent (HiLo grouping, M=59.2, SD=2.9; LoHi grouping, M=59.3, SD=3.2; difference *ns*), a consistent preference for either of them would be indicative of a statistical learning bias developed prior to the experiment.

Method

Participants Thirty-seven English monolingual and 36 native Korean students participated. Korean participants were enrolled in graduate programs at the University of Hawaii, and their scores in the TOEFL test of English as Second Language was high (M=252.14, out of 300, SD=16).

Stimuli A seamless 5-min speech sequence of 8 synthesized syllables (*bu, fu, ra, ti, she, zi, ge, ni*) was generated following the forward and backward transitional probabilities in Table 3, with 80 ms for consonants and 260 ms for vowels, and 5 s fade-in and out. We used the Italian diphone set in order to engage participants in a foreign language learning task for both groups. The Italian phonemes we chose had equivalent phonemic realizations in English and Korean, and all syllable sequences were phonotactically legal. No participant knew Italian. No cue to word boundary was present other than the patterns of predictive and retrodictive dependencies (Table 1). Importantly, whenever forward probability was low between any two adjacent syllables, (fwdTP(zi|she=.33)), backward probability was high (backTP(she|zi=1)), and vice versa (Table 1). At test, two groupings corresponding to a pattern of HiLo probability and LoHi probability were pitted one against the other in a forced-choice task. None of the possible groupings was an actual word in either language. The six test pair trials were presented in random sequential order, while the order within a pair was counterbalanced by repeating each test pair twice.

Procedure Participants in each language group were randomly assigned either to the experimental condition that included Training and Test or to a control condition that included the Test phase only (18 English native speakers, 21 Korean native speakers). Participants in the experimental condition listened to the training stream for 5 minutes. Test consisted in 12 two forced-choice task trials between pairs of LoHi and HiLo groupings. For each pair they were asked to choose which sound sequence formed a grouping in the language they had just heard. All instructions were administered in the native language of participants.

Table 1. The forward [square brackets] and backward (no brackets) transition probabilities associated with any two adjacent stimuli in the training sequence of the three experiments (in Experiments 2 and 3 the syllables were replaced by abstract shapes and tones respectively). For example, given *bu* only *fu* could follow (fwd-TP(fu|bu)=[1]), while given *fu* there was a .33 probability that *bu* preceded it (back-TP(bu|fu)=(.33)). Last three rows: A sample of the training sequence of Experiment 1. Perceptual grouping boundaries could emerge either when the forward transitional probability between adjacent syllables was high and the backward probability was low (HiLo groupings), or viceversa (LoHi groupings).

	To							
From	bu	fu	ra	Ti	she	zi	ge	ni
bu	0	[1]	0	0	0	1	0	0
fu	.33	0	.33	.33	[.33]	0	[.33]	[.33]
ra	0	[1]	0	0	0	1	0	0
ti	0	[1]	0	0	0	1	0	0
she	0	1	0	0	0	1	0	0
zi	[.33]	0	[.33]	[.33]	[.33]	0	.33	.33
ge	0	1	0	0	0	[1]	0	0
ni	0	1	0	0	0	[1]	0	0
Training	… fushezirafunizitifugezibu ...							
HiLo groupings	… fushe zira funi ziti fuge zibu …							
LoHi groupings	...fu shezi rafu nizi tifu gezi …							

Results

In all three experiments presented participants responses were coded in terms of proportion endorsements for HiLo groupings. Consequently, low endorsement rates for HiLo indicate preferences for LoHi groupings. A 2 (Language: Korean, English) x 2 (Condition: Experimental, Control) ANOVA revealed a main effect of Language ($F(1,72)=11.22$, $p<0.01$), and a Language by Condition interaction ($F(1,72)=10.67$, $p<0.01$). In particular, English native speakers exposed to training consistently preferred the LoHi groupings 62% of times (SD=14%), which was reliably different than chance ($p < .001$). Conversely, Korean native speakers exposed to training preferred HiLo groupings 59% of times (SD=13.6%), which was reliably different than chance ($p < .025$). Thus, English and Korean participants attended to the transitional probabilities that were most predictive of the canonical word order of their native language, as predicted by the corpus analyses. When presented with the Test items alone, no preference emerged for either groupings above or below change (English, $t(17) = 0.44$, $p = 0.67$; Korean, $t(20) = 1.03$, $p = 0.32$), ensuring that the bias in the experimental condition was not due to inherent preferences for certain sound combinations of the test items.

Experiment 2

In order to further ascertain that the different scores in Experiment 1 were due to language-specific biases, in Experiment 2 we tested whether learning biases would arise when the same miniature grammar was implemented with non-linguistic stimuli in the visual modality. As discussed previously, sensitivity to forward and backward probabilities has been demonstrated in non-language domains, including visual processing (Fiser & Aslin, 2002; Jones & Pashler, 2007). However, the two cues were not pitted against each other in those experiments, but rather one or the other was maximally informative in the input. Here, at any stimulus transition the two cues are equally informative, but pitted against each other. Therefore we expected one of two scenarios. A null result would obtain if learners as a group weigh each cues equally, as indeed Jones & Pashler (2007)'s data suggest. Alternatively, if a priori visual preferences are attested in participants' responses, we expect them to be general visual sequential processing biases not influenced by language experience. Thus, regardless of scenario, we expected no differential preferences based on the language of our participants.

Method

Participants 15 new English and 15 Korean native speakers from the same population as Experiment 1 participated.

Stimuli A continuous sequence was generated that had exactly the same structure and length as Experiment 1, with the only exception that the 8 synthesized syllables were now replaced by 8 abstract shapes. Shapes appeared in succession on the screen for 340ms. The sequence had the same statistical properties as the language in Experiment 1.

Procedure The same learning and test procedure as Experiment 1 applied. At test participants received a two forced-choice task between pairs of LoHi and HiLo shapes. For each pair they were asked to choose which one formed a grouping in the sequence they had just seen. All instructions were administered in the native language of participants.

Results

A one-way ANOVA revealed no significant main effect of Language ($F(1,28)=0.025$, $p=0.87$). Moreover, mean test items endorsements did not differ from chance (Korean, $M = 0.51$, $SD=0.17$, $t(14) = 0.26$, $p = 0.8$; English, $M = 0.51$, $SD=0.16$, $t(13) = 0.03$, $p = 0.97$).

Experiment 3

The results of Experiment 2 indicate that the difference in directional preference between English and Korean speakers does not extend to visual sequences. This is consistent with the hypothesis that what drives the difference between language speakers in Experiment 1 is their experience with language. However, prior research indicates that there are important differences between processing visual stimuli and audio stimuli, due in part to the more transient nature of audio information (e.g., Conway & Christiansen, 2005). As such, it may be the case that the differences between English and Korean speakers are not specific to language, but rather arise from more general differences in auditory processing. To assess this possibility, we created a tonal analog of the input from Experiment 1.

Method

Participants Both English monolinguals (N=15) and Korean/English bilingual (N=15) who reported Korean as their dominant language participated in this experiment. All English monolinguals were undergraduates at Carnegie Mellon University, as were six of the Korean English bilinguals. The other 9 bilingual participants were recruited via advertising in Pittsburgh churches.

Stimuli Each of the syllables in the language used in Experiment 1 was replaced by a unique 330 ms tone (bu = A4, ti = B, ge = C#, zi = D, fu = E, ni = F#, she = G#, ra = A5) in the key of A major. The resulting tonal sequence thus had an identical statistical structure as in Experiment 1.

Procedure Participants listened to the tone sequence over headphones. Next, participants were given 12 forced choice questions and asked to indicate, on a response sheet, which of two items sounded "more like" the tone sequence they had just heard. On each of the 12 questions, a tone item with a high-forward, low-backward transitional probability was paired with an item with low-forward, high-backward transitional probability.

Results

A one-way ANOVA revealed no effect of Language ($F(1,28)=0.025$, $p=0.87$). Both English ($M = 0.56$, $SD = 0.09$, $t(14) = 2.6$, $p = .01$) and Korean-English bilinguals ($M = 0.58$, $SD = 1.14$, $t(14) = 2.1$, $p = .05$) selected test items with high forward transitional probabilities (HiLo items) at a rate above chance. The fact that they performed equivalently in this experiment is consistent with the hypothesis that the differences between these two populations in Experiment 1 are language-specific, and strengthen the claim that they arise from linguistic experience. Unlike adults' lack of preference for shape test items in Experiment 2, participants did have a consistent preference for test items with high forward transitional probabilities, perhaps because the preference for forward-going items is a domain-general auditory preference, similar to the Iambic-Trochaic Law (e.g., Hay & Diehl, 2007). This preference may be early-developing, and then modified by linguistic experience: strengthened for Korean learners, and contravened by English learners. Alternatively, experience with music may inculcate a bias in both English and Korean listeners; on this account, musical experience is more consistent cross-culturally than linguistic experience.

Discussion

We tested the hypothesis that adult English and Korean speakers come to the lab having already developed opposite statistical preferences for parsing continuous speech. The results of Experiment 1 supported that hypothesis: where English speakers preferred items with high backward probabilities, Korean speakers preferred items with high forward probabilities. Experiments 2 and 3 suggest that this preference is limited to linguistic materials. This limitation is consistent with the possibility that the preference arises from experience with language. Corpus analyses of English and Korean are consistent with this possibility, as the predominant word order of both languages mirrors the direction preference of English and Korean speakers.

Our findings have implications for understanding processes of second language acquisition. Our Korean participants were advanced second language speakers of English, and were enrolled in graduate programs in the United States. Their sensitivity to the learning bias opposite to that of the English participants suggests that they implicitly used their solidified L1's learning biases when learning the novel artificial language. Because patterns that conform to those initially learned are further promoted, interference can be most severe with the learning of patterns that do not conform to those initially learned. We propose that lower-order kinds of transfer involving basic sequential processing biases are at play in second language acquisition, and may have a ripple effect on encoding higher-order processes such as word order structure.

The results suggest that statistical learning changes throughout development by adapting to the characteristics of the native language. This opens many avenues for subsequent research, including understanding the mechanisms and developmental time course through which experience with native language alters subsequent learning.

Reference

Conway, C.M., & Christiansen, M.H. (2005). Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 24-39.

Cook, V.J., & Newson, M. (2007). *Chomsky's universal grammar: An introduction* (3rd ed.). Wiley-Blackwell.

Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, *99*, 15822-15826.

Gervain, J. Nespor, M., Mazuka, R., Horie, R., Mehler J. (2008) Bootstrapping word order in prelexical infants: a Japanese-Italian cross-linguistic study. *Cognitive Psychology*, *57*(1), 56-74.

Hay, J.F., & Diehl, R.L. (2007). Perception of rhythmic grouping: Testing the Iambic/Trochaic Law. *Perception and Psychophysics*, *69*, 113-122.

Johnson, N.F. (1965) The psychological reality of phrase structure rules. *Journal of Verbal Learning and Verbal Behavior*, *4*, 469-475.

Jones, J., & Pashler, H. (2007). Is the mind inherently forward looking? Comparing prediction and retrodiction. *Psychonomic Bulletin and Review*, *14*(2), 295-300.

Maye, J., Werker, J. F. & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*(3), B101- B111.

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.

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

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

Tremblay, A., Derwing, B., Libben, G. and Westbury, C. (2011), Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks. *Language Learning*, *61*: 569–613.