Statistical learning bias predicts second-language reading efficiency

Luca Onnis (lucao@ntu.edu.sg)
Division of Linguistics and Multilingual Studies, Nanyang Technological University
Singapore, 637332

Stefan L. Frank (s.frank@let.ru.nl)
Centre for Language Studies, Radboud University
6525 HT Nijmegen, The Netherlands

Hongoak Yun (oaktreepark@gmail.com)
Department of Media and Communication, Konkuk University
Seoul, Korea, 143-701

Matthew Lou-Magnuson (matt0020@ntu.edu.sg)
Division of Linguistics and Multilingual Studies, Nanyang Technological University
Singapore, 637332

Abstract

Statistical learning (SL) is increasingly invoked as a set of general-purpose mechanisms upon which language learning is built during infancy and childhood. Here we investigated the extent to which SL is related to adult language processing. In particular, we asked whether SL proclivities towards relations that are more informative of English are related to efficiency in reading English sentences by native speakers of Korean. We found that individuals with a stronger statistical learning sensitivity showed a larger effect of conditional word probability on word reading times, indicating that they more efficiently incorporated statistical regularities of the language during reading. In contrast, L2 English proficiency was related to overall reading speed but not to the use of statistical regularities.

Keywords: statistical learning; sequential learning; reading; sentence processing; bilingualism.

Introduction

Human languages are learnt and processed in real time. Speech is the ultimate fleeting experience, as it dissipates as soon as it is produced. And while printed text is more stationary, proficient readers process words sequentially at a very fast pace, with relatively few gazes spent looking back to reread previous words. The inherent fleeting nature of language and the great efficiency that humans exhibit in learning and using languages suggest that the brain recruits mechanisms employed for processing sequential information. These mechanisms may involve the ability to unconsciously track and extract patterns of regularities across sensory modalities, and to abstract over these patterns (for reviews see Gomez & Gerken, 2000; Perruchet & Pacton, 2006).

Because of the probabilistic sequential nature of language processing, recent research has attempted to establish links between mechanisms for language learning and processing and of so-called statistical learning (SL), by relating individual variance in SL tasks with individual variance in tasks of language learning. The rationale of such approaches is to show that some measure of statistical learning ability, as assessed in tasks requiring implicitly learning relations among probabilistic sequences, is correlated with performance on one or more tasks involving language. Languages exhibit statistical properties at different levels of analysis, which make them potentially learnable from experience. In the early stages of language development infants and toddlers take in a considerable amount of this statistical structure. Infants exhibit individual differences in statistical learning skills that may modulate language development trajectories (e.g., Arciuli, & von Koss Torkildsen, 2012; Benasich et al., 2006; Kidd, 2012).

Here we take a similar approach in seeking evidence for a relation between statistical learning and second-language reading in adults. Arguably statistical language learning does not stop in the early years of childhood. Studies with older children have also linked poor implicit statistical skills with language and/or reading difficulties (Evans, Saffran, & Robe-Torres, 2009; Yim & Windsor, 2010) and adult native speakers are even sensitive to the particular statistical distribution of sentence structures within an experimental session, and adapt their processing preference accordingly (Fine, Jaeger, Farmer, & Quin, 2013). Thus, sensitivity to the statistical structure of language is likely to support not only children learning a language, but also adults using it daily. Indeed, direct predictive relations between statistical learning scores and online sentence processing and other linguistic tasks exist now both for children and adults (Yim & Windsor, 2010). In addition, neurophysiological data suggest that similar neural mechanisms appear to serve both syntactic processing of language and statistical learning of sequential patterns (Abla, Katahira, & Okanoya, 2008; Christiansen, Conway, & Onnis, 2012). Tracking implicit sequential regularities in linguistic and nonlinguistic stimuli seems to be independent of factors other than language performance, such as age, nonverbal IQ, and memory (Yim & Rudoy, 2010; Kaufman et al., 2010).

Here we are interested in capturing individual differences in statistical learning, language proficiency, and language comprehension, and we aim to correlate the three. Adult non-native speakers of English who show a stronger "English-

like" statistical learning bias in an artificial grammar task which is specifically designed as a litmus test to gauge the strength of preexisting experience with statistical regularities of English, are expected to be more sensitive to the statistics of English in an online reading task. Our study aims to achieve the following goals: 1) further support the view that statistical learning skills underlie not only language learning in childhood, but also language processing in adults; and 2) contrast the processing effect of individual differences in statistical learning vs. second-language (L2) proficiency on language reading.

Language-specific statistical learning

To measure individual statistical learning in second language speakers we used a task devised by Onnis & Thiessen (2013). The rationale for the task is to capture potential proclivities towards language-specific statistical relations. Natural languages differ in the statistical regularities available to learners. One such difference relates to the predominant directionality of conditional relations among words. For example, while "the" does not strongly predict "dog" (because many words can follow "the"), "dog" more strongly retrodicts "the." Learners are sensitive to informative relations in both directions (Jones & Pashler, 2007): 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). The predominant directionality of relations among elements of the input differs between natural languages. 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 (e.g., 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. English for example arranges prepositional phrases like "to school" such that the head "to" precedes the noun "school", while other languages favor postpositional organization (as in Korean 학교-에 *hakkjo-e* 'school to'). An intuitive prediction derived 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 smaller set of words that can precede "school." As these examples demonstrate, learners of different languages may experience different degrees of forward-going and backward-going cohesiveness. 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. Thus, language-specific information latent in the linear order of words partially predicts phrase structure in language.

Differences in statistical patterns between languages may, in turn, alter statistical learning itself. Sensitivity to backward-going regularities may be more adaptive for 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 auditory artificial grammar with conflicting forward and backward transitional probabilities, as in the training sample in a):

a) Training sample: .. *fushezirafunizitifugezibu* ..

Crucially, the artificial grammar was such that whenever forward transitional probability (fwd-TP) was low between any two adjacent syllables, backward transitional probability (back-TP) was high (e.g., fwd-TP($zi|she$) = .33) and back-TP($she|zi$) = 1), and vice versa (e.g., fwd-TP($ra|zi$) = 1 and back-TP($zi|ra$) = .33).

Two parses of sample a) into bisyllabic units are equally possible. One parse segments the signal such that the two syllables of a word have a high forward probability and a low backward probability (the HiLo pattern), while in the other parse the word-internal forward probabilities are low and the backward probabilities are high (the LoHi pattern).

b) Possible Parse I (HiLo): ..*fushe zira funi ziti fuge zibu* ..
c) Possible Parse II (LoHi): .. *shezi rafu nizi tifu gezi* ..

At test, two word groupings corresponding to the HiLo and LoHi patterns were pitted against each other in a two-alternative forced-choice task. A participant's *statistical learning bias* was defined as the proportion of LoHi choices over the sum of test trials presented to them. While both language groups had experienced the same grammar, native English speakers predominantly grouped syllables on the basis of high backward probabilities (as in example c above), while native Korean speakers preferred a grouping on the basis of high forward probabilities (as in example b). By contrast, with either visual or tonal non-linguistic 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 directionality of their native language, suggests that the difference is due to linguistic experience.

Thus, the findings of Onnis & Thiessen (2013) suggest that SL can adapt to the statistical structure of linguistic input in ways that lead learners to have different expectations about novel subsequent input. To further support this claim, we wanted to find out whether individuals' degree of SL bias is correlated with statistical sensitivity to natural language in an online language comprehension task.

Statistical learning and language processing

Statistical patterns play an important role in language learning. Language *processing*, too, has been shown to depend on statistical information. Because of the time-dependent and sequential nature of both speech and reading, language comprehension is an inherently sequential process that probabilistically anticipates upcoming material. Hence, the cognitive processing effort for a piece of language (e.g., a word) should depend on its occurrence probability. At its most fundamental, a word's occurrence probability is simply its frequency in the language. Indeed, this frequency predicts the time required to recognize the word. When the word forms part of a sentence or text, reading time on the word is logarithmically related to its occurrence probability given the preceding context (Smith & Levy, 2013).

Estimating occurrence probabilities of words in context requires a probabilistic language model that implements knowledge about the language's statistics. Possibly the simplest language model is the bigram model, which assumes that a word's probability depends only on its overall frequency and on the immediately preceding word. Hence, word probabilities under such a model equal forward transitional probabilities, which have been found to predict reading times (McDonald & Shillcock, 2003). More sophisticated models can capture language statistics more accurately, resulting in more accurate reading time predictions (e.g., Frank & Bod, 2011). For the current study, however, we limit ourselves to a bigram model because the SL bias in our artificial grammar learning study is defined in terms of transitional probabilities.

Method

Participants. Fifty-seven adult native speakers of Korean (45 women; age $M = 22.6$, $SD = 2.7$) were recruited at four universities in Seoul (Konkuk University, Ewha Womans University, Sogang University, and Seoul National University). To qualify for the study their TOEFL score of English as a Second Language score should be over 600 (old version), and they should have spent at least three years in English-speaking country or environment. They participated in a statistical learning task, a sentence reading task, and an English proficiency self-assessment task (in this order). They were tested individually in a quiet room at their own university and were paid 10,000 Korean Won for their participation.

Statistical Learning Task

Materials. For the artificial grammar, the same materials as Onnis & Thiessen's (2013) Experiment 1 were used. The grammar lexicon was composed of eight monosyllabic words (fu, zi, shae, ni, ge, ra, ti, bu). To generate the training materials these words were arranged in a seamless sequence according to the rules of a stochastic Markovian grammar chain. The process started by choosing one of the eight possible words at random, and then generating the next word according to two probabilistic sequence constraints:

whenever the forward probability between any two adjacent words was low (fwd-TP = .33), the backward probability was high (back-TP = 1), and vice versa. A sample of this template sequence is " …fu shae zi ra fu ni zi bu fu ge zi ra fu ni zi bu fu ge zi ti fu shae zi …". Frequencies of individual words, bigrams (two-word sequences), and their associated transition probabilities are summarized in Table 1.

Table 1: Summary of statistical relations among adjacent words in the grammar used in the the statistical learning task. Fwd-TP and Back-TP indicate forward and backward transitional probabilities among words in each bigram. Freq1 and Freq2 indicate frequency of occurrence of first and second word in each bigram, while Freq Bigram indicates how often each bigram occurred during training.

Bigram	Type	Fwd-TP	Back-TP	Freq 1	Freq 2	Freq Bigram
fu shae	LoHi	0.36	1.00	133	48	48
fu ni	LoHi	0.32	1.00	133	42	42
fu ge	LoHi	0.32	1.02	133	42	43
zi ra	LoHi	0.35	1.00	132	46	46
zi ti	LoHi	0.31	1.00	132	41	41
zi bu	LoHi	0.34	1.00	132	45	45
shae zi	HiLo	1.00	0.36	48	132	48
ni zi	HiLo	1.00	0.32	42	132	42
ge zi	HiLo	1.00	0.32	42	132	42
ra fu	HiLo	1.00	0.35	46	133	46
ti fu	HiLo	1.00	0.31	41	133	41
bu fu	HiLo	1.00	0.34	45	133	45

The actual sequence was realized using the speech synthesizer MBROLA, and concatenating the eight words to form a pauseless 3.5 minute speech stream of 711 words, with 80 ms for consonants and 260 ms for vowels. Because we were interested in the perception of grouping boundaries as driven by statistical biases alone, MBROLA did not use any prosodic or temporal cues to grouping boundaries. In addition, the sequence faded in and out for 5 s, giving the impression of an unbounded stream. The Italian diphone set in MBROLA was chosen to make the words dissimilar enough to Korean, but still clearly perceivable, and to engage participants in an "alien language" learning task. Finally, we ensured that all syllable sequences were phonotactically legal in Korean.

To verify whether participants preferred a specific pattern of transitional probabilities after exposure to the training phase, at test two types of bigrams were pitted one against the other in a forced-choice task, corresponding to a pattern of high fwd-TP and low back-TP (dubbed "HiLo" bigrams) versus the opposite "LoHi" bigrams. For example, the LoHi bigram 'fu shae' was presented against the HiLo bigram 'shae zi'. Six test pair trials were presented in random order, while the order within a pair was counterbalanced by repeating each test pair twice, for a total of 12 test trials. Note that HiLo and LoHi bigrams were composed of the same pseudowords and had been presented with an equal frequency at training. Hence, the only statistics useful to systematically choose one

type over the other would have to be a preference for the patterns of transition probabilities giving rise to the bigrams.

Procedure. Participants first listened to the training stream for 3.5 min, after which they were presented with the forced-choice task between pairs of LoHi and HiLo bigrams. For each pair they were asked to choose which sequence formed a grouping in the novel language they had just heard. We coded 1 for responses consistent with the English bias (preference for LoHi bigrams), and 0 for responses consistent with head-final languages such as Korean (namely, HiLo bigrams). We then defined a participant's *statistical learning bias* as the proportion of LoHi choices over the 12 test trials presented. The *strength* of the learning bias, computed as (learning bias – 0.5)^2, quantifies bias extremeness towards either LoHi or HiLo preference.

Sentence Reading Task

Materials. Sentences came from the 361-sentence UCL corpus (Frank, Monsalve, Thompson, & Vigliocco, 2013) explicitly created to evaluate language models on word-reading times. These sentences were drawn from original English narratives. Each participant was randomly assigned to one of ten groups, each containing 36 unique test sentences in English from the University College London UCL corpus, and five practice sentences. Test sentences were presented in random order. The words were displayed one at a time, progressing across the screen in their natural position with successive presses of the spacebar. Approximately half of the sentences were followed by a yes-no question regarding the content of what was just read in order to maintain the attention of the participants.

Proficiency assessment task

Participants self-assessed their proficiency in English listening, speaking, reading, and writing, as well as their accent, on a 7-point scale. All Pearson correlation coefficients between each pair of ratings was significantly positive (all *p* < .0005), and thus we took the average rating for each participant as a single measure of second language proficiency.

Results

A more efficient reader should adapt her reading times to the words' log-transformed occurrence probability, such that more probable words are read more quickly (Levy, 2008; Smith & Levy, 2008). Hence, we take the extent to which higher log-transformed forward transitional probability (as opposed to base word frequency) predicts shorter reading time as indicative of a reader's efficiency.

Naturally, we expect participants with higher English proficiency to read faster. In addition, they may also read more efficiently, in the sense that they display a more negative effect of forward probability on reading time. Alternative, non-exclusive possibilities are that participants read English faster and/or more efficiently if they have a more English-like (i.e., larger) learning bias or a stronger (i.e., more extreme) learning bias. We further expect any

effects of English proficiency and learning bias to be independent from each other, although proficiency and learning bias may themselves be correlated.

Learning bias and proficiency. The mean statistical learning bias (towards the English-like LoHi patterns), learning bias strength (towards either pattern), and L2 proficiency score were, respectively, 0.575 (*SD* = 0.197), 0.044 (*SD* = 0.051), and 5.144 (*SD* = 0.721). Proficiency was not significantly correlated with learning bias (*r* = .08; *p* > .5) nor with strength of learning bias (*r* = −.10; *p* > .4). We note that the mean SL bias preference is closer to the English-expected pattern than what Onnis & Thiessen (2012; OT dataset) found. A comparison of the data distributions in the two dataset indeed suggests a major difference, notably a bimodal distribution, with the two modes located at 0.4 and 0.6, i.e. on each side of the chance level value of 0.5 in the current data. Conversely, the OT dataset had a single mode at 0.3. The bimodal distribution in our data suggests that the absolute strength of learning bias may be a better measure reflecting sensitivity to SL.

Data preprocessing. Data on a complete sentence was excluded if any RT on a word was extreme (below 80 ms or above 3000 ms). Furthermore, we did not include data on sentence-initial and sentence-final words, words followed by a comma, and clitics. This left a total number of 23,640 data points for analysis.

Reading time analysis. To investigate how readers' sensitivity to language statistics is related to their learning bias and English proficiency, the collected data were analyzed by linear mixed-effects regression. Subject-specific predictors (fixed effects) were: statistical learning bias (SLBIAS), strength of the bias (SLBIAS^2), and English proficiency (PROFICIENCY). Item-specific predictors were: word position in the sentence (WORDPOS), number of letters of word (LENGTH), log-transformed word frequency (WORDFREQ), and the log-transformed forward transitional word probability (FORWPROB). Word frequency and forward transitional probability were computed from word and bigram counts in the written-text part of British National Corpus. Properties of the previous word (PREVLENGTH, PREVWORDFREQ, PREVFORWPROB) were also included to take into account potential spillover in the reading times. As trial-specific predictor, RT on the previous word (PREVRT) was included to factor out the auto-correlation between consecutive key presses (Baayen & Milin, 2010). In addition, the model included by-subject and by-item (i.e., word token) random intercepts and by-subject random slopes of all predictor variables except for the subject-specific ones. RTs were log-transformed and all independent variables were standardized.

The first model that was fitted included the two-way interactions between each of the three subject-specific predictors and each of the six item-specific predictors (PREV)LENGTH, (PREV)WORDFREQ, and (PREV)FORWPROB. Next, non-significant (|*t*| < 2) interactions were removed one at a time, starting with the least significant interaction. Table 3 shows the resulting model's fixed-effects coefficients with

corresponding *t*- and *p*-values (*p*-values are obtained by treating *t*-values as *z*-scores, which is justified by the very large amount of data).

Table 3: Regression model fitted to log-transformed RTs

Variable	Coeff.	*t*-value	*p*-value
(intercept)	6.192	339.9	< .0001
PREVRT	0.091	14.7	< .0001
WORDPOS	−0.015	−3.5	< .0005
LENGTH	0.041	9.8	< .0001
PREVLENGTH	−0.006	−2.2	< .03
WORDFREQ	−0.006	−1.3	.2
PREVWORDFREQ	−0.013	−2.5	< .02
FORWPROB	−0.025	−6.7	< .0001
PREVFORWPROB	−0.013	−3.1	< .002
PROFICIENCY	−0.087	−4.0	< .0001
SLBIAS	0.009	0.4	.7
SLBIAS^2	0.001	0.0	.9
PROFICIENCY × LENGTH	−0.021	−5.0	< .0001
SLBIAS^2 × FORWPROB	−0.007	−2.8	< .005

As expected, LENGTH is positively related to RT: Longer words take longer to read. In addition, there are reliable negative effects of PREVWORDFREQ and (PREV)FORWPROB on RT: words are read faster if they are more frequent or their occurrence is more likely given the previous word.

The strong negative effect of PROFICIENCY means that participants who self-assessed their level of English as higher read more quickly, which validates the proficiency measure. L2 proficiency also modulates the effect of LENGTH in that less proficient participants (relative to more proficient ones) display increased difficulty with longer words.

Crucially, statistical learning bias is not significantly related to RT nor does it play a role in any interaction effect. In contrast, the strength of learning bias modulates the effect of FORWPROB: The stronger the learning bias (i.e., higher SLBIAS^2), the more sensitive reading becomes to forward probability (the effect of FORWPROB is more negative)

Discussion

In this study we investigated the association of individual differences in second-language processing with individual differences in a probabilistic sequence learning task and in second-language proficiency. The rationale is that if SL subserves language, and learning languages implies the discovery of language-specific distributional relations, then SL biases that match the statistical structure of a specific language increase efficiency when processing – here: reading – that language.

Our study extends on recent literature relating SL and literacy development. Arciuli and Simpson (2012) found a correlation between a non-linguistic SL task and measures of reading abilities derived from standardized reading tests in both elementary school children and adult native speakers. In addition, Spencer, Kashak, Jones and Lonigan (2014) established correlations between SL measures and early skills related to literacy development, notably oral language abilities, vocabulary knowledge, and phonological processing. With respect to second language learning, Frost, Siegelman, Narkiss, and Afek (2013) provided evidence that SL predicts word decoding abilities in a second language.

Our results further suggest that the ability to track statistical relations in sequenced patterns may not only be useful in *learning* a language early in life – the focus of previous research – but is also significantly correlated with the ability to *process* natural language as adults. In addition, the above studies established relations between SL and either broad measures of literacy outcomes, such as scores of standardized reading tests, or measures of single-word orthographic knowledge. The present study allowed a finer-grained examination of the role of statistical learning in more naturalistic reading conditions. We examined how biases in statistical learning reflecting the optimization of language-specific knowledge are related to second language proficiency and efficiency in real-time reading. We found that participants whose learning bias more closely matched the English-like head-first pattern were better able to use word predictability (operationalized as forward transitional probability) in real-time sentence processing. Concurrently, those participants showed a weaker effect of the words' base frequency on RTs. This is consistent with our interpretation that having a more English-like SL bias makes one closer to an "ideal" expectation-based English-language processor, who is sensitive to the words' conditional probabilities rather than base frequencies (which are already incorporated in the conditional probability measure).

Although more proficient participants generally read faster, and particularly so on longer words, there was no interaction between L2 proficiency and word frequency or forward transitional probability, suggesting that –perhaps surprisingly– increased proficiency is not reflected in more accurate knowledge or use of English language statistics. Conversely, participants with more positive learning bias did not read more quickly. The absence of this main effect is hard to explain considering our claim that more positive learning bias correlates with more efficient reading, as one would expect more efficient readers (i.e., those who make more optimal use of language statistics) to be faster readers, too. Finally, our results show that L2 proficiency and statistical learning bias are independent factors: They did not correlate across participants nor did they significantly interact.

Conclusion

Finding correlations between artificial grammar learning proclivities and language processing abilities contributes to validating the statistical learning approach to language. Future work could help establish whether people who are more sensitive to statistical sequential information make better language learners, and ultimately language users. In addition, understanding exactly what type of statistical information is required to optimize language tasks such as reading would greatly expand our knowledge of the mechanisms required for language processing. This line of research is not only useful to inform theories of language in

the brain, but has potential practical applications. For example, it may be possible to assist inefficient second-language readers by helping them process statistical information more optimally. Furthermore, our statistical learning task could be used to predict delays in language development, in cases where direct assessment of language is difficult (toddlers, or multilinguals for which assessment of language delays is confounded with proficiency in a given language).

Acknowledgments

References

Abla, D., Katahira, K., & Okanoya, K. (2008). On-line assessment of statistical learning by event-related potentials. *Journal of Cognitive Neuroscience*, *20*, 952–964.

Arciuli, J., & Simpson, I. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science*, *36*, 286–304.

Arciuli, J., & von Koss Torkildsen, J. (2012). Advancing our understanding of the link between statistical learning and language acquisition: the need for longitudinal data. *Frontiers in Psychology*, *3*. 324.

Benasich, A. A., Choudhury, N., Friedman, J. T., Realpe-Bonilla, T., Chojnowska, C., & Gou, Z. (2006). The infant as a prelinguistic model for language learning impairments: predicting from event-related potentials to behavior. *Neuropsychologia*, *44*, 396–411.

Baayen, R. H. & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*, 12–28.

Christiansen, M. H., Conway, C. M., & Onnis, L. (2012). Similar neural correlates for language and sequential learning: evidence from event-related brain potentials. *Language and Cognitive Processes*, *27*(2), 231-256.

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*, 321–335.

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS ONE, 8*, e77661.

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science, 22*, 829–834.

Frank, S. L., Monsalve, I. F., Thompson, R. L., & Vigliocco, G. (2013). Reading time data for evaluating broad-coverage models of English sentence processing. *Behavior Research Methods*, *45*, 1182–1190.

Frost, R., Siegelman, N., Narkiss, A., & Afek, L. (2013). What predicts successful literacy acquisition in a second language? *Psychological Science*, *24*, 1243–1252.

Gomez, R. L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4(5), 178–186.

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

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321–340.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology*, *48*, 171–184.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*, 1126-1177.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research, 43*, 1735–1751.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*, 268–284.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward 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., & Pacton, S. (2006). Implicit learning and statistical learning: One phenomenon, two approaches. *Trends in Cognitive Sciences*, 10, 233–238.

Smith, N. J., & Levy, R. (2008). Optimal processing times in reading: a formal model and empirical investigation. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Meeting of the Cognitive Science Society* (pp. 595–600). Austin, TX: Cognitive Science Society.

Smith N. J., Levy R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302–319.

Yim, D., & Windsor, J. (2010). The roles of nonlinguistic statistical learning and memory in language skill. *Korean Journal of Communication Disorders*, *15*, 381–396.