

When ‘More’ in Statistical Learning Means ‘Less’ in Language: Individual Differences in Predictive Processing of Adjacent Dependencies

Jennifer B. Misyak (jbm36@cornell.edu)

Morten H. Christiansen (christiansen@cornell.edu)

Department of Psychology, Cornell University, Ithaca, NY 14853 USA

Abstract

Although statistical learning (SL) is widely assumed to play a key role in language, few empirical studies aim to directly and systematically link variation across SL and language. In this study, we build on prior work linking differences in nonadjacent SL to on-line language, by examining individual-differences in *adjacent* SL. Experiment 1 documents the trajectory of adjacency learning and establishes an individual-differences index for statistical bigram learning. Experiment 2 probes for within-subjects associations between adjacent SL and on-line sentence processing in three different contexts (involving embedded subject-object relative-clauses, thematic fit constraints in reduced relative-clause ambiguities, and subject-verb agreement). The findings support the notion that proficient adjacency skills can lead to an over-attunement towards computing local statistics to the detriment of more efficient processing patterns for nonlocal language dependencies. Finally, the results are discussed in terms of questions regarding the proper relationship between adjacent and nonadjacent SL mechanisms.

Keywords: Predictive Dependencies; Sentence Processing; Bigrams; Serial Reaction Time; Artificial Grammar

Introduction

With the expansion of studies on statistical learning (SL) over the past decades, focus has intensified towards probing the potential role for probabilistic sequence learning capabilities in acquiring and using linguistic structure (e.g., Gómez, 2002; Saffran, 2001). A clearer understanding has in turn begun to crystallize about the ways in which SL mechanisms may underpin language across various levels of organization—phonetic, lexical, semantic, syntactic—and across differing timescales—phylogenetic, ontogenetic, and microsecond unfoldings. Largely missing from this picture, however, is empirical evidence that directly links language and SL abilities within the typical population.

There are, though, a few recent studies that address the issue of whether better statistical learners are indeed better processors of language. In a small-scale study of individual differences, Misyak and Christiansen (2007) observed that standard measures of SL performance are positively associated with comprehension accuracy for various sentence-types in natural language. Conway, Bauernschmidt, Huang and Pisoni (2010) reported that better SL performance correlates with better processing of perceptually-degraded speech in highly-predictive lexical contexts. Misyak, Christiansen and Tomblin (2010) found that more-skilled statistical learners of nonadjacent structure were also more adept at the on-line processing of long-distance dependencies in natural language. Thus far, these

results would support the general assumption that SL and language processes are systematically interrelated, with positive correspondence in intraindividual variation across them. But is it always the case that greater SL is associated with better language functioning? Or, may excelling at one of these implicate poorer performance at the other?

Such ability-linked reversals in performances within a cognitive domain would not be unprecedented. As an example, bilingual individuals appear to possess more efficient ‘inhibitory control’ processes than their monolingual peers across a number of studies, which has usually been imputed in some manner to bilinguals’ greater experience with ‘control’ processes for suppressing irrelevant information in the course of successfully using two languages (see Bialystok et al., 2004). However, in a negative priming paradigm where distractor locations that were supposed to be previously ignored became relevant for facilitating responses to a current trial (as they do for monolinguals), bilinguals are at a disadvantage in the cognitive control task, with decreases from a neutral baseline in performance accuracy (Trecanni et al., 2009). Analogously then, might there be natural language contexts in which superior SL skill also becomes disadvantageous?

One possibility is that a statistical learner may focus too much on computing certain statistics, while ignoring others, with repercussions for their linguistic processing. For example, language embodies predictive dependencies that can be broadly characterized as involving either *adjacent* or *nonadjacent* temporal relationships. Thus, a good adjacency learner might perform poorly on nonadjacent dependencies in language. Introducing a new task for documenting micro-level trajectories and individual differences in SL, Misyak et al. (2010) were able to link variation in *nonadjacent* SL positively to signature differences in reading time patterns for the complex *nonlocal* dependency structure of center-embedded object-relative clause sentences. However, this study raises a new set of questions, including ones that directly bear on the above hypothetical, namely: Does the timecourse of *adjacent* SL differ from that of nonadjacent SL? Can substantial differences in adjacent SL also be empirically related to on-line sentence processing? And if so, might this differ from the kinds of positive correlations observed for nonadjacency processing?

We investigated these questions by adapting the AGL-SRT paradigm from Misyak et al. (2010) to isolate the learning of adjacent dependencies. The task implements an artificial grammar (AG) within a modified two-choice serial reaction-time (SRT) layout, using auditory-visual sequence-

strings as input. Experiment 1 thus documents the group trajectory and range of individual differences for adjacency learning obtained from this task. A ‘bigram index’ reflecting individual differences in adjacency learning is then used to probe relationships to the processing patterns observed in our subsequent natural language experiment (Experiment 2).

Experiment 1: Statistical Learning of Adjacencies in the AGL-SRT Paradigm

The ability of humans to use adjacent statistical information has been demonstrated across various studies. As early as two months of age, humans can identify bigrams, or first-order adjacent pairs, from the co-occurrence frequencies of elements within a constrained temporal sequence (Kirkham, Slemmer & Johnson, 2002). Throughout later development and adulthood, humans can also use adjacent conditional probabilities to locate relevant constituent-boundaries in a continuous stream composed of nonwords, tones, visual elements, or nonlinguistic sounds (see Gebhart, Newport & Aslin, 2009, for a review). And further, both children and adults can learn adjacent predictive dependencies that signal the underlying phrase structure of an artificial language (Saffran, 2001).

Below, we adapt the biconditional grammar of Jamieson and Mewhort (2005) to examine adults’ SL of bigrams. This grammar was chosen since it is defined by first-order transitions only, imposes no positional constraints on element placement, and generates strings of equal length. These merits thereby permit us to effectively isolate the learning of predictive adjacencies by our participants.

Method

Participants Thirty native English speakers from the Cornell undergraduate population (15 females; age: $M=19.4$, $SD=0.8$) were recruited for course credit.

Materials Participants observed sequences of auditory-visual strings generated by an eight-element grammar in which every element could be followed by one of only two other elements, with equal probability. Each string consisted of 4 elements, with adjacent probabilities between them as shown in Table 1. The nonwords (*jux*, *tam*, *hep*, *sig*, *nib*, *cav*, *biff*, and *lum*) were randomly assigned to the stimulus tokens (*a*, *b*, *c*, *d*, *e*, *f*, *g*, *h*) for each participant to avoid

Table 1: Transition probabilities for elements at positions n and $n + 1$ of a string, with n as an integer from (0, 4).

Element at n	Element at position $n + 1$ of string							
	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>	<i>h</i>
<i>a</i>	0	.5	.5	0	0	0	0	0
<i>b</i>	0	0	.5	.5	0	0	0	0
<i>c</i>	0	0	0	.5	.5	0	0	0
<i>d</i>	0	0	0	0	.5	.5	0	0
<i>e</i>	0	0	0	0	0	.5	.5	0
<i>f</i>	0	0	0	0	0	0	.5	.5
<i>g</i>	.5	0	0	0	0	0	0	.5
<i>h</i>	.5	.5	0	0	0	0	0	0

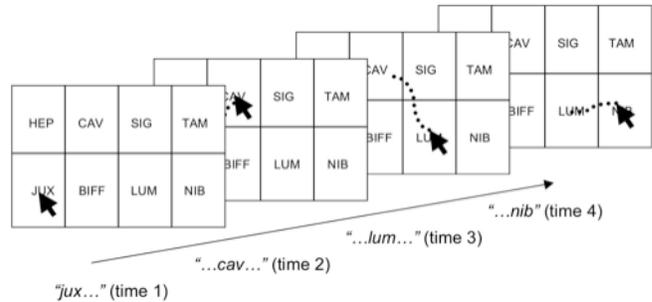


Figure 1: The pattern of mouse clicks for a single trial with the auditory target string “*jux cav lum nib*.”

potential learning biases due to specific sound properties of words. Auditory versions of the nonwords were recorded from a female native English speaker and length-edited to 550 ms. Written versions of nonwords were presented with standard spelling in Arial font (all caps) and appeared within the rectangles of a 2 x 4 computer grid (see Figure 1). Each of the 4 columns of the computer grid, from left to right, displayed the nonword options corresponding to the 1st thru 4th respective elements of a string. Ungrammatical strings were created by introducing an incorrect element at the 2nd or 3rd string position, with the next element being one that legally followed the incorrect one (e.g., as in “*a *d e g*”).

Procedure Each trial corresponded to a different configuration of the grid, with each of the eight written nonwords centered in one of the rectangles. Every column contained a nonword (target) from a stimulus string, as well as a foil. The first column contained the selection for the first element of a string, the second column contained the selection for the second element, and so on. For example, a trial with the stimulus string *jux cav lum nib*, as shown in Figure 1, might contain the target *jux* and the foil *hep* in the first column; the target *cav* and the foil *biff* in the second column; the target *lum* and the foil *sig* in the third column; and the target *nib* and the foil *tam* in the fourth column. Each nonword appeared equally often as target and as foil within and across the columns. The top/bottom locations of targets and foils were randomized and counterbalanced.

Participants were informed that the purpose of the grid was to display their selections and that a computer program randomly determines a target’s location within either the top or bottom rectangle. On every trial, participants heard an auditory stimulus string composed of four nonwords and were instructed to respond to each nonword in the sequence as soon and as accurately as possible by using the computer mouse to select the rectangles displaying the correct targets.

Thus for any given trial, after 250 ms of familiarization to the visually presented nonwords, the first nonword of a string (the target) was played over headphones. Next, the second, third, and fourth words of a given string were each played after a participant had responded in turn to the prior nonword. For example, on a trial with the stimulus string *jux cav lum nib*, the participant should first click the rectangle containing JUX upon hearing *jux* (Fig. 1, left), CAV upon next hearing *cav* (Fig. 1, center-left), LUM upon hearing *lum* (Fig.

l, center-right), and NIB upon hearing *nib* (Fig. 1, right). After a participant had responded to the last nonword, the screen cleared for 750 ms before a new trial began.

An intended consequence of this design is that, for any given trial, the first element of a string cannot be anticipated in advance of hearing the auditory target. However, all subsequent string transitions might be reliably anticipated using statistical knowledge of the bigram structure. Thus, as participants become sensitive to the bigrams, they should be able to anticipate the string transitions, which should be evidenced by faster response times (following standard SRT rationale). Accordingly, our dependent measure on each trial was the reaction time (RT) for a predictive target, subtracted from the RT for the non-predictive initial-column target (which serves as a baseline and controls for practice effects). The predictive target used in this calculation was equally distributed across all non-initial columns across trials. Analogously, for an ungrammatical string trial, if participants are sensitive to the bigrams, then their RTs for incorrect, or violated, elements should be slower; thus, the DV for ungrammatical trials was the RT for the illegal target subtracted from the initial-target RT.

There are 64 unique strings ($8 \times 2 \times 2 \times 2$) defined by the grammar; these were all randomly presented once each for each grammatical block of trials. Training consisted of six grammatical blocks, followed by an ungrammatical block of 16 trials and then a single grammatical ('recovery') block. Transitions across blocks were seamless and unannounced.

After these eight blocks, participants were informed that the strings had been generated according to rules specifying the ordering of nonwords and were asked to complete two tasks involving prediction and bigram recognition, respectively. The prediction task consisted of 16 trials that were procedurally similar to the trials observed during training, but with the omission of the auditory target for the final column.¹ Instead, participants were told to select that nonword in the final column that they believed best completed the sequence.

In the bigram task, participants were randomly presented with 32 test items of auditory nonword-pairs. They were requested to judge whether each pair followed the rules of the grammar by pressing 'yes'/'no' computer keys. Half of the test items were the 16 bigrams licensed by the grammar (e.g., *a b*); the remaining half were illegal pairings formed by reversing each bigram (e.g., *b a*). Thus, successful discrimination reflects knowledge of the conditional bigrams, rather than only sensitivity to co-occurrences.

Results and Discussion

Analyses were performed on only 'good' trials—that is, accurate string-trials with only one selection for each target.

¹ Instructing participants to complete string *endings* allows for maximal procedural similarity to the speeded training trials without introducing additional cue prompts that would be needed if the aurally-omitted element varied across non-initial columns. It also avoids any indirect feedback effects from presenting the next element after a participant's correct/incorrect medial selection.

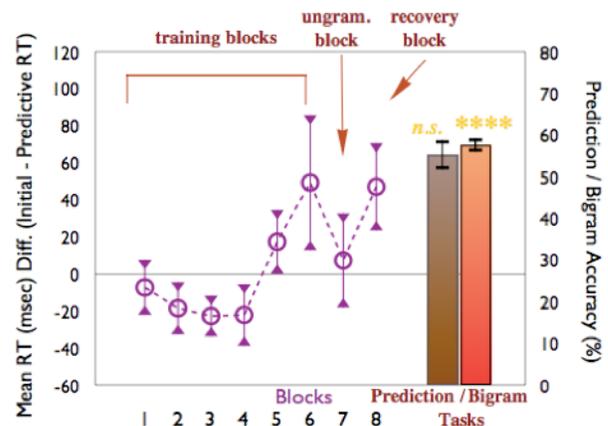


Figure 2: Group learning trajectory (mean RT difference scores per block) and accuracy for prediction (left bar) and bigram (right bar) tasks.

Prior to analysis, the data from five participants were omitted (2 for withdrawing participation; 2 for improperly performing the task, with less than 40% good trials; and 1 for abnormally elevated RTs, averaging in excess of 1470 ms per single response). For remaining participants, good trials averaged 88.2% ($SD=5.9$) of training block trials.

Mean RT difference scores, as described above (i.e., for grammatical trials: initial-target minus predictive-target RT; for ungrammatical trials: initial-target minus illegal-target RT) were computed for each block and submitted to a one-way repeated-measures analysis of variance (ANOVA) with block as the within-subjects factor. Since the assumption of sphericity was violated ($\chi^2(27) = 113.27, p < .001$), degrees of freedom were corrected using Greenhouse-Geisser estimates ($\epsilon = .33$). Results indicated a main effect of block on RT difference scores, $F(2.31, 55.36) = 3.82, p = .02$. As seen in Figure 2, mean RT difference scores appear to increase by the final training block, decrease in the ungrammatical block, and increase once again in the recovery block. As RT difference scores measure the amount of facilitation from the predictive targets, an improvement in scores across blocks (as seen here) reflects sensitivity to the adjacent dependencies.

Planned contrasts between the ungrammatical block and preceding/succeeding grammatical blocks confirmed a performance decline for the ungrammatical trials (Block 6 minus Block 7: $M = -42.0$ ms, $SE = 19.6$, $t(24) = 2.14, p = .04$; Block 8 minus Block 7: $M = 39.8$ ms, $SE = 17.8$ ms, $t(24) = 2.23, p = .04$). This provides evidence for participants' learning of the sequential dependencies, consistent with standard interpretations in the sequence learning literature for comparing RTs to structured versus unstructured material (e.g., Thomas and Nelson, 2001).

Since the amount of exposure to the dependencies during training is equivalent to that which a similar number of participants ($n=30$) received in the Misyak et al. (2010) study of nonadjacent SL, this invites a comparison of group learning trajectories. The RT timecourse pattern documented here for adjacent SL is very similar to that observed for nonadjacent SL, but with greater variance in

performance for the final training block and with ostensibly more modest (albeit not statistically different) performance in the recovery block. In both cases, sensitivity to the statistical structure does not show signs of emerging until after considerable exposure (the 5th block of training).

Mean accuracy on the prediction task was 55.3% ($SD=17$), which was not above chance ($t(24) = 1.51, p = .14$)—despite 20% of participants scoring at or above 75%. However, accuracy on the bigram task reflected adjacency learning ($t(24) = 4.66, p < .0001$), with a mean of 57.6%. This performance level is consistent with participants' judgment accuracy in an AGL study with manipulations of this same type of grammar when participants are tested with ungrammatical items containing few rule violations (Jamieson & Mewhort, 2009). Bigram scores further ranged from 37.5 – 71.9%, but with less variance ($SD=8$) than that observed in the prediction task. In post-study questioning, only four participants disclosed that they had noticed any general pattern in the sequence but were unable to verbalize at least one instance of a bigram, suggesting that their performance in the bigram task was not the product of explicit recall or well-formulated meta-knowledge. Next, we use scores on this bigram index to assess whether and how variation in adjacent SL may be associated with differences in processing local and nonlocal language dependencies.

Experiment 2: Individual Differences in Language Processing and Statistical Learning

Sensitivity to both local and long-distance relationships is indispensable to processing natural language, and pervades basic aspects of our everyday sentence comprehension and production—such as those involved in relating the modified subject/object of a described action or state to the main event of a sentence (embedded relative clauses), in identifying whether someone is the recipient or doer of an action (agent-patient thematic roles), and in correctly linking subjects with their verbs (number agreement). The aim of Experiment 2 is to investigate whether predictive processing as exemplified by adjacent SL is empirically related to the on-line processing of such natural language contexts. Consider the following examples of the sentence-types that constitute the focus of the current experiment.

- (1a-b) The reporter [that attacked the senator / that the senator attacked] admitted the error.
- (2a-b) The [crook/cop] arrested by the detective was guilty of taking bribes.
- (3a-b) The key to the [cabinet/cabinets] was rusty from many years of disuse.

In the first sentence example, the subject-relative (SR; 1a) and the object-relative (OR; 1b) versions differ with respect to the manner in which the embedded verb *attacked* relates to its object. This involves a more complex, backwards-tracking long-distance dependency (to the head-noun) for ORs. In prior studies using materials resembling those in (1a-b), greater processing difficulty is elicited at the main verb of ORs compared to that of SRs, with considerable

individual differences in the magnitude of this effect (e.g., Wells, Christiansen, Race, Acheson & MacDonald, 2009).

Next, consider the sentence pair (2a-b), which is temporarily ambiguous between a main verb (MV) and a reduced relative (RR) clause interpretation. Its resolution is influenced by the constraint of thematic fit—the fit between the head noun phrase (*the crook* or *the cop*) and the verb-specific roles of the verb (*arrested*). Given verb-specific conceptual knowledge, the reader knows that *cop* is a typical agent of *arrested*, whereas *crook* is a typical patient. Controlling for animacy, thematic fit functions as an immediately integrated constraint computed over the noun and adjacent verb—with its effect on RTs occurring in the subsequent agent NP region (McRae, Spivey-Knowlton & Tanenhaus, 1998). Thus, the second condition (2b) in which the initial noun is a typical agent for the adjacent verb will elicit greater processing difficulty for the RR interpretation than that for the corresponding patient condition (2a). For our purposes, this provides an example of sensitivity to a local relation relevant for on-line sentence processing.

Lastly, (3a-b) illustrate subject-verb number agreement. In English, it is required that a number-marked subject (*key*) agrees with the number-marking of its verb (*was*). This is the case irrespective of the numerical marking of any intervening material (e.g., *to the cabinet/s*), and individuals are sensitive to this fact during reading. When a sentence's head noun is singular, individuals read longer at the MV in a condition where the 'distracting' local noun (*cabinets*) mismatches in number (i.e., is plural) than in a condition where the local noun matches the head noun's number (i.e., is singular); shorter reading latencies are also found for the word after the verb in the match condition (Pearlmutter, Garnsey & Bock, 1999). Although subject-verb agreement may occur locally between adjacent constituents, materials in the literature (and here) have involved a nonlocal dependency created from interposing a prepositional phrase.

Method

Participants The same participants from Exp. 1 participated directly afterwards in this experiment for additional credit. Because the analyses reported below involve correlations with the bigram index from Exp. 1, data was omitted for those participants already excluded in Exp. 1 analyses and from three others (2 for bilingual status and 1 for declining to participate in the second task).

Materials There were four sentence lists, each consisting of 9 practice items, 60 experimental items, and 50 filler items. The experimental items were sentences drawn from previous studies of sentence processing: 20 subject-object relative clauses (SOR; Wells et al., 2009), 20 reduced relative ambiguities influenced by thematic fit (TF; McRae et al., 1998), and 20 subject-verb agreement transitives (S-V; Pearlmutter et al., 1999). A yes/no comprehension probe followed each item. Item conditions within sentence sets were counterbalanced across lists.

Procedure Each participant was randomly assigned to a list, whose items were presented in random order using a

a standard word-by-word, moving window, self-paced reading paradigm. Millisecond reading times (RTs) per word and accuracy were recorded for analyses.

Results and Discussion

Overall comprehension accuracy across participants was high, $M= 87.4\%$, $SD=7.6$. RTs in excess of 2500 ms (0.2% of data) were removed, and remaining RTs were then length-adjusted for the number of characters in a word using a standard procedure (Ferreira & Clifton, 1986). Unless otherwise noted then, all RTs reported below for each of the sentence sets have been length-adjusted, with the same sentence regions examined as those in the original studies. RTs connected with relevant effects for each of the sets were then used to probe for associations with individuals' bigram scores from Experiment 1, as summarized below.

Subject-Object Relatives. Results replicated the main effect for clause-type at the MV from Wells et al. (2009), $F(1, 21) = 5.55$, $p= .03$. OR MVs were read reliably longer (91 ms) than SR MVs. However, there was no significant correlation between bigram scores and MV RTs for either SR ($r = .04$, $p= .85$) or OR ($r = -.16$, $p= .47$) sentences. Thus, differences in adjacent SL did not appear to directly map onto differences in processing long-distance dependencies in these relative clauses.

Thematic Fit. The influence of TF was replicated at the 2-word MV region (e.g., *was guilty*), $F(1, 21) = 6.42$, $p = .02$, albeit not at the directly preceding agent NP region.² Agent conditions were read 39 ms longer than patient conditions at the MV region. The correlation between bigram scores and unadjusted RTs at the MV of the 'congruent' patient condition was not significant ($r = .29$, $p= .19$); but for the 'incongruent' agent condition, the correlation reached marginal significance ($r = .40$, $p= .06$), with better adjacent statistical learners taking longer to read the disambiguating verb phrase. This suggests a tendency for greater bigram sensitivity (in adjacent SL) to negatively correspond with resolving nonlocal ambiguity when the local TF constraint provides an opposing bias to the RR clause interpretation.

Subject-Verb Agreement. A 34 ms effect of match (i.e., the difference between match and mismatch conditions) was obtained at the verb, $F(1, 21) = 31.28$, $p < .0001$, which replicated Pearlmutter et al.'s (1999) findings. There was a smaller effect of match (23 ms) at the post-verb region, $F(1, 21) = 4.48$, $p= .05$, which was also numerically present but not reliable in Pearlmutter et al. Additionally, the correlation between bigram scores and RTs was significant for the effect at the verb ($r = .51$, $p= .02$), with better bigram learning corresponding to a larger effect of match condition. To further examine differences in processing patterns according to SL status, a median-split was performed on bigram scores, establishing 57.8% as the cut-off for defining membership in either a "high" bigram ($n=11$, $M= 63.9\%$,

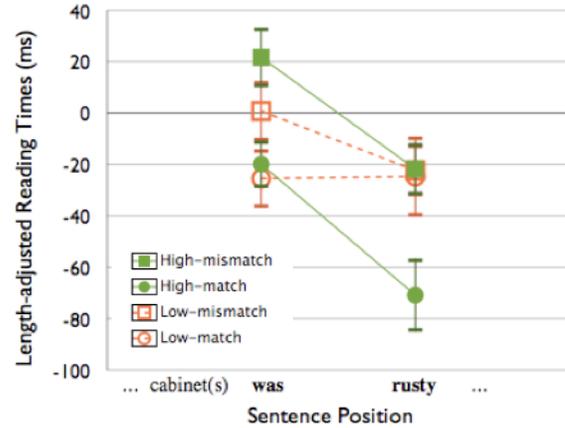


Figure 3: RT patterns on the S-V agreement sentences by bigram group (high/low) and condition (match/mismatch).

$SD=4.0$) or "low" bigram group ($n=11$, $M= 51.4\%$, $SD=5.8$). Significant bigram-group differences emerged for the effect of match condition across regions (as shown in Figure 3). While the low-bigram group did not elicit a significant effect of match condition at either the verb or post-verb region ($p= .13$ and $p= .91$, respectively), the high-bigram group showed a clear effect in both regions (both p 's $< .001$). As apparent in Fig. 3, the high-bigram group demonstrated greater sensitivity to the interference created by the locally mismatched marking of the noun in the prepositional phrase (which was irrelevant for computing agreement). Thus, the better adjacent SL of the high-bigram group was related to generally less efficient processing than that by their low-bigram peers of the long-distance dependency entailed by the initial noun and verb. Since bigram groups did not differ in comprehension accuracy for any sentence-types in the experimental sets (all p 's $> .15$), nor fillers ($p= .83$), these RT patterns were not the result of a speed-accuracy tradeoff.

Our findings suggest that adjacent SL skill may not directly tap into the processes most relevant for handling long-distance dependencies in natural language—even though nonadjacent SL abilities appear to do so. Thus, while Misyak et al. (2010) reported a positive association between differences in nonadjacent SL and processing for the same SOR clauses as used here, no correlation was detected for adjacent SL. More generally, this is consistent with the lack of within-subjects correlation found between adjacent and nonadjacent SL in Misyak and Christiansen (2007).

However, while 'high' bigram learners may not differ from 'low' learners on processing long-distance relations as such, their increased sensitivity to local relations might interfere with the processing of the longer-distance elements within the sentence. This tendency is seen in the TF set, where above-average bigram tracking abilities seem to have a negative effect for processing the MV—the site where the initial, nonlocal ambiguity must be resolved. Similarly, too much sensitivity to local information is clearly evidenced within the last sentence set, where the irrelevant marking of an adjacent noun negatively affects better bigram learners' resolutions of S-V agreement, with protracted RTs also at the MV site of integrating the long-distance dependency.

² The later-occurring but nonetheless reliable effect of thematic fit is likely due to differences in the length of the moving window used in this study (1-word) and that by McRae et al. (2-word).

General Discussion

This study investigated the processing of adjacent predictive dependencies to address questions related to the timecourse of adjacent SL and the nature of any empirical association to natural language variation. While a learning trajectory similar to nonadjacent SL was documented in Exp. 1, findings from Exp. 2 indicated that above-average gains in adjacent SL performance do not necessarily translate to gains in language processing. Notably, those individuals who were strongly attuned to tracking statistical bigrams exhibited a *negative* pattern of correlations to tracking longer-distance aspects of language when either countervailing adjacent constraints or nearby distractive elements were present. This inverse pattern was not evidenced, though, when processing long-distance relations without conflicting local information (in the SOR clauses).

Instances where better bigram learners were worse language processors (or tended towards less efficient RT patterns) occurred when the integration of adjacent information (between a head-noun and part-participle verb) induced greater difficulty for resolving an ambiguity as a RR (the TF constraint in Exp. 2)—or when locally irrelevant information disrupted agreement computations between a nonlocal subject and verb (S-V agreement in Exp. 2). It would appear in these situations that those better in adjacent SL, although excelling at bigram pattern recognition in the SL task, are overly attuned to adjacency patterns and become more susceptible to local ‘garden-paths’; in such cases, it may be the ‘over-focus,’ rather than any preexisting weakness in processing long-distance dependencies (as evidenced by parallel performance of groups in the SOR set) that hinders efficient resolution of nonlocal relationships.

This interpretation of our findings suggests that intraindividual differences in processing biases for the integration of competing constraints among adjacent- and nonadjacent dependencies may contribute to variation across SL-linked language processing skills. As such, it speaks to an open issue regarding whether different systems or different processing biases may be entailed by adjacent and nonadjacent processing capabilities in humans. It has been proposed, for instance, that the two forms of processing may be subserved by separate brain areas (Friederici et al., 2006), or that the two types of SL are only nominally distinct as the outcome of task-specific attention processes that may selectively hone in on adjacent or nonadjacent statistics (cf. Pacton & Perruchet, 2008). The findings here, of negative and specific associations between adjacent SL and aspects of language processing, suggest that future individual differences research incorporating careful attention to a diversity of natural dependency-structures may be needed to help establish the proper relation between these two manifestations of SL and the extent to which they may ‘tap’ into the same underlying mechanisms.

Acknowledgments

Thanks to Parry Cadwallader, Becky Fortgang and Stephan Spilkowitz for assistance with running participants.

References

- Bialystok, E., Craik, F.I.M., Klein, R. & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290-303.
- Conway, C.M., Bauernschmidt, A., Huang, S.S. & Pisoni, D.B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition, 114*, 356-371.
- Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25*, 348-368.
- Friederici, A.D., Bahlmann, J., Heim, S., Schibotz, R.I. & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences, 103*, 2458-2463.
- Gebhart, A.L., Newport, E.L. & Aslin, R.N. (2009). Statistical learning of adjacent and nonadjacent dependencies among nonlinguistic sounds. *Psychonomic Bulletin & Review, 16*, 486-490.
- Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*, 431-436.
- Jamieson, R.K. & Mewhort, D.J.K. (2005). The influence of grammatical, local, and organizational redundancy on implicit learning: An analysis using information theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 9-23.
- Jamieson, R.K. & Mewhort, D.J.K. (2009). Applying an exemplar model to the artificial-grammar task: Inferring grammaticality from similarity. *Quarterly Journal of Experimental Psychology, 62*, 550-575.
- Kirkham, N.Z., Slemmer, J.A. & Johnson, S.P. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition, 83*, B35-B42.
- McRae, K., Spivey-Knowlton, M.J. & Tanenhaus, M.K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38*, 283-312.
- Misyak, J.B. & Christiansen, M.H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language. In *Proceedings of the 29th Annual Cognitive Science Society* (pp. 1307-1312). Austin, TX: Cognitive Science Society.
- Misyak, J.B., Christiansen, M.H. & Tomblin, J.B. (2010). Sequential expectations: The role of prediction-based learning in language. *Topics in Cognitive Science, 2*, 138-153.
- Pacton, S. & Perruchet, P. (2008). An attention-based associative account of adjacent and nonadjacent dependency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 80-96.
- Pearlmutter, N.J., Garnsey, S.M. & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*, 427-456.
- Saffran, J.R. (2001). The use of predictive dependencies in language learning. *Journal of Memory and Language, 44*, 493-515.
- Thomas, K.M. & Nelson, C.A. (2001). Serial reaction time learning in preschool- and school-age children. *Journal of Experimental Child Psychology, 79*, 364-387.
- Treccani, B., Argyri, E., Sorace, A. & Della Sala, S. (2009). Spatial negative priming in bilingualism. *Psychonomic Bulletin & Review, 16*, 320-327.
- Wells, J.B., Christiansen, M.H., Race, D.S., Acheson, D.J. & MacDonald, M.C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250-271.