

# Simultaneous Online Tracking of Adjacent and Non-adjacent Dependencies in Statistical Learning

Loan C. Vuong (loan.vuong@mpi.nl)

Antje S. Meyer (antje.meyer@mpi.nl)

Max Planck Institute for Psycholinguistics, Wundtlaan 1  
6525 XD Nijmegen, The Netherlands

Morten H. Christiansen (christiansen@cornell.edu)

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

## Abstract

When children learn their native language, they have to deal with a confusing array of dependencies between various elements in an utterance. Some of these dependencies may be adjacent to one another whereas others can be separated by considerable intervening material. Research on statistical learning has begun to explore how such adjacent and non-adjacent dependencies may be learned—but in separate studies. In this paper, we investigate whether both types of dependencies can be learned together, similarly to the task facing young children. Statistical learning of adjacent and non-adjacent dependencies was assessed using a modified serial-reaction-time task. The results showed (i) increasing online sensitivity to both dependency types during training, and (ii) non-adjacent dependency learning being highly correlated with adjacent dependency learning. These results suggest that adjacency and non-adjacency learning can occur in parallel and that they might be subserved by a common statistical learning mechanism.

**Keywords:** Statistical learning; Non-adjacent Dependencies; Adjacent Dependencies; Artificial Grammar Learning; Serial Reaction Time

## Introduction

It is generally assumed that statistical learning, a domain-general mechanism that encodes statistical regularities across space and time, plays a role in language acquisition (see Saffran, 2003, for a review). The extent to which statistical learning can support language acquisition is, however, a point of controversy (e.g., Marcus, Vijayan, Rao, & Vishton, 1999; Seidenberg & Elman, 1999; Yang, 2002). This is in part because in natural languages dependencies can concern adjacent elements (e.g., dependencies between verb stems and inflectional morphemes as in *learning*) or non-adjacent ones (e.g., dependencies between auxiliaries and inflectional morphemes as in *is learning*), but studies of artificial grammar learning have shown that adults and children are highly sensitive to the former type of regularity (e.g., Saffran, Aslin, & Newport, 1996) and considerably less sensitive to the latter (e.g., Gómez, 2002; Newport & Aslin, 2004).

Why are non-adjacent dependencies hard to learn? One proposal is that this might be due to a default bias towards adjacent dependencies (Gómez, 2002). Evidence for this hypothesis comes from studies demonstrating the *variability*

*effect*, which is the finding that high variability in the material embedded between the non-adjacent dependent elements facilitates non-adjacent dependency learning compared to low variability (Gómez, 2002; see also Onnis, Christiansen, Chater, & Gómez, 2003). In the relevant studies, participants were trained on an artificial grammar of the form  $aXb$ , in which the non-adjacent pairing between the first and third element ( $a-b$ ) was fixed at the perfect level (non-adjacent transitional probability = 1.0), while the adjacent pairing between the first and second element ( $a-X$ ) was variable. Across participant groups, the size of the adjacent set varied from 2 (yielding an adjacent transitional probability between the first and second element of .50) to 24 (yielding an adjacent transitional probability of .04). The participants listened to the sequences during training and later took a surprise grammaticality judgment test that assessed their non-adjacent dependency learning. Results showed significantly worse non-adjacent dependency learning at the smaller than at the larger adjacent set sizes.

In light of the above findings, Gómez (2002) proposed that learners might initially have a default focus on adjacent dependencies (henceforth, the *adjacency-first* view). When the adjacent set has relatively low variability (with a smaller adjacent set size), adjacency learning continues to predominate. By contrast, when the adjacent set has high variability (with a large adjacent set size), learners might soon “seek out alternative sources of predictability” (e.g., non-adjacent dependencies) (pg. 435). Statistical learning is therefore partially flexible, as the adjacency bias could be overridden given high variability in the adjacent set. However, a trade-off between adjacency and non-adjacency learning that favors adjacency learning might result when adjacent-set variability is low. This proposal thus elegantly accounts for the poorer non-adjacency learning performance at the smaller adjacent set sizes despite the still perfect non-adjacency statistics in those conditions. Note that not all observed effects of adjacent on non-adjacent dependency learning are negative. Later studies (Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007) have shown that if the non-adjacent dependency to be learned has been encountered earlier as an adjacent pair, then non-adjacent dependency learning is facilitated (see also, e.g., Newport & Aslin, 2004; Creel, Newport, & Aslin, 2004 for other facilitating

factors). In either case, existing evidence suggests a heavy influence of adjacent on non-adjacent dependency learning.

In the current study, we aimed to assess this interaction between adjacency and non-adjacency learning by re-examining the learning of non-adjacent dependencies under a small adjacent set size condition thought to be favorable to adjacent dependency learning, and less favorable to non-adjacent dependency learning (Gómez, 2002; Onnis et al., 2003). Previous studies used an offline judgment task to measure learning of non-adjacent dependencies (e.g., Gómez, 2002, Onnis et al., 2003), or they used an online task but measured learning of adjacent and non-adjacent dependencies in different experiments (Misyak & Christiansen, 2010; Misyak, Christiansen, & Tomblin, 2010), or they trained participants on adjacent and non-adjacent dependencies within the same experiment but in a serial manner (Lany & Gómez, 2008; Lany, Gómez, & Gerken, 2007). By contrast, we tracked *online* learning of both adjacent and non-adjacent dependencies *simultaneously* in the *same* task as learning progressed.

Training materials were similar in structure to those used in Lany et al. (2007, Experiment 3) and consisted of four non-adjacent dependency pairs with an adjacent set size of 6 (e.g.,  $aX_{1-6}b$ ). To track learning online, we adopted the artificial grammar learning-serial reaction time (AGL-SRT) paradigm (Misyak et al., 2010, see Figure 2). The participants listened to triplets of sound sequences, such as *jom - namie - mig*. On the computer screen, they saw the written versions of the stimuli along with three distractors. They listened to the strings one element at a time and made a mouse click response to select the corresponding element on the visual display. We measured the latencies from the presentation of each element in the string and tracked the participants' developing sensitivities to adjacent and non-adjacent dependency patterns by computing adjacency and non-adjacency facilitation scores. These scores were derived by taking the reaction time (RT) to the first, non-predictable targets as baseline and subtracting from it the RTs to the predictable second and third targets, respectively. In this paradigm, increasing sensitivity manifests itself in increasing facilitation scores, averaged across each training block.

In order to provide an opportunity for non-adjacent dependency learning to develop, we used more training trials than usual in this kind of study (two sessions on successive days totaling over 1000 trials compared to typical training regimes consisting of a single session under 500 trials). As in Misyak et al. (2010), the penultimate block of each session featured only ungrammatical sequences in which both previously trained adjacent and non-adjacent pairings were violated. Because the first element in the ungrammatical block was no longer predictive of the second and third elements, the prediction-based benefits in RT should be eliminated, thus leading to a drop in the facilitation scores towards baseline in this block. The final block of each session was a recovery block in which the original pairings were reinstated. If participants had learned

the dependency patterns well, there should be a rebound in the facilitation scores because the first element was again predictive in this block. After completing the AGL-SRT task, participants were asked to complete a prediction task as well as a standard grammaticality judgment task.

Given that the probabilities used in this study favored non-adjacent dependencies, one may perhaps expect better non-adjacency than adjacency learning eventually. However, the critical prediction concerns the learning time course of the two patterns. As the current adjacent set size was relatively small (and would not fall into the category of "high variability" in Gómez's (2002) study), the *adjacency-first* view would predict that learners first focus on adjacent dependencies and then on non-adjacent dependencies (if at all). Therefore, according to this view, we should find an early emergence of sensitivity to adjacent and a later emergence of sensitivity to non-adjacent dependencies.

## Methods

### Participants

Twenty-two Dutch native speakers recruited from local universities in Nijmegen, the Netherlands (mean age = 20.8,  $SD = 2.2$ , 5 men) participated in exchange for monetary compensation (8 Euros/hour).

### Materials

Training materials consisted of spoken three-element strings of Dutch pseudowords. In keeping with previous materials for comparison purposes (Gómez, 2002), the first and third element of the strings were monosyllabic (*lin, pes, zol, taf, bur, jom, mig, vun*), whereas the second one was disyllabic (*namie, hufel, hagix, dazan, fenar, gunis, dosef, witus, sluro, kapek, worat, ruxot*). Assignment of tokens to target elements was randomized across participants. The stimuli were recorded by a female Dutch native speaker. The average durations were 373 ms ( $SD = 48$ ) for the monosyllabic items and 546 ms ( $SD = 46$ ) for the disyllabic items.

The first and third element of the training strings formed four fixed non-adjacent dependency pairs ( $a-b, c-d, e-f, g-h$ ; non-adjacent transitional probability = 1.0). Similar to Lany et al. (2007; Experiment 3), the non-adjacent dependency pairs were divided into two subsets, each sharing the same set of six intervening items in the second position ( $aX_{1-6}b$  and  $cX_{1-6}d$  vs.  $eY_{1-6}f$  and  $gY_{1-6}h$ ; adjacent transitional probability = .17, see Figure 1). Therefore, for a given participant, two non-adjacent dependency pairs (e.g., *jom-mig* and *zol-vun*) were only combined with one second-element subset (e.g., *namie, sluro, hagix, worat, ruxot*, and *kapek*), whereas the other two non-adjacent dependency pairs (e.g., *taf-bur* and *pes-lin*) were only combined with the remaining subset of second elements (e.g., *hufel, fenar, dosef, witus, dazan*, and *gunis*).

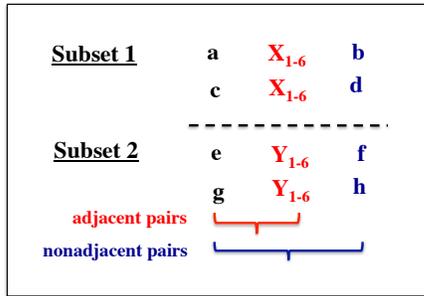


Figure 1: Structure of the training materials.

On the computer screen, the participants saw written versions of the targets along with three distractors drawn from items at the same element position of the other stimulus subset (e.g., only the second-element, but not the first- or third-element items, from the other subset could serve as distractors for a given second-element target). Each pseudoword appeared equally often as a target and as a distractor. For the ungrammatical block, both adjacent and non-adjacent pairs were violated such that the second targets were switched between subsets while the third targets were switched between pairs within the same subset (e.g., *\*aYd: jom-fenar-vun, \*eXh: taf-ruxot-lin*).

For the prediction task, correct target elements were presented together with distractors as in training (see *Procedure*). For the grammaticality judgment task, half of the strings contained intact adjacent and non-adjacent pairs (e.g., *aXb: jom-namie-mig*), while the other half contained intact adjacent but violated non-adjacent pairs with the third elements being switched between subsets (e.g., *\*aXh: jom-namie-lin*).

### Procedure

Participants were tested in two sessions, with the second session taking place within 24 +/-3 hours after the first. In each session, participants went through six training blocks of 96 trials each (4 non-local pairs x 6 local pairs x 4 repetitions), followed by an ungrammatical block of 24 trials (4 non-local pairs x 6 local pairs), and a recovery block of 96 trials (4 non-local pairs x 6 local pairs x 4 repetitions). In the second session, they were also administered two surprise tasks, a prediction task and a grammaticality judgment task (24 trials each).

Similar to previous AGL-SRT studies (see Misyak et al., 2010 for details), target items were presented along with distractors in a rectangular grid display on the computer screen (see Figure 2). The first, second, and third target appeared in either the upper or lower row of the first, second, and third column, respectively. Target positions were counterbalanced such that each target was as likely to occur in the upper as in the lower row. Each training trial started with the presentation of a fixation cross in the center of the computer screen for 750 ms. Then the visual display was shown until the end of the trial. 250 ms after the onset of the display, the first target was played. The participants were asked to make a mouse click inside the rectangular

area containing the target as quickly and accurately as possible. Immediately following the participants' response, the second target was played and the participants made a second mouse click response. The same procedure applied to the third target, after which the current trial was brought to an end. A different random order of trials was used for each participant.

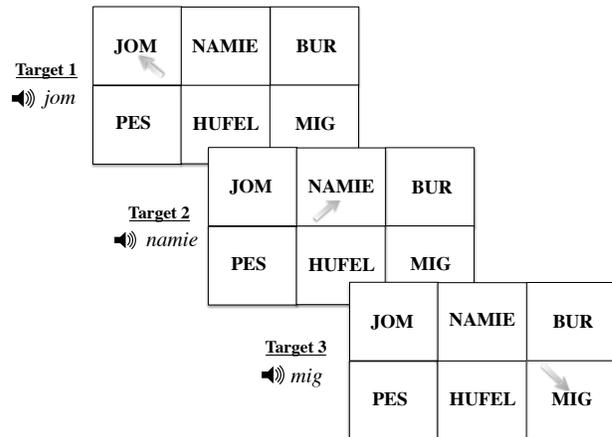


Figure 2: Example of a series of events in a training trial. The gray arrows indicate the targets.

Following the training trials in the second session, participants were told that the triplets heard on training trials had followed certain patterns on which they would be tested in two short tasks. In the prediction task, the same procedure was followed as during training for the first two elements, but for the third element the participants were asked to select the appropriate target without any auditory information. In the grammaticality judgment task, they heard a stimulus triplet and responded by pressing one of two keys to indicate whether or not it followed the previously trained patterns. The materials were presented in a different randomized order to each participant.

### Results

Participants were 98% accurate on average ( $SD = 1.17$ ) in making mouse click responses during training. Correct reaction times (RT) beyond 2000 ms were excluded from analysis (0.20% of the data). As the auditory targets differed in length, we corrected for length by performing a linear regression predicting each participant's correct RTs from the targets' spoken durations (Ferreira & Clifton, 1986). We then computed each participant's average residual RT per block. Adjacency and non-adjacency facilitation scores were computed by subtracting average second- and third-target residual RTs, respectively, from first-target residual RTs for each block and participant. Positive adjacency facilitation scores indicate faster responses to the second than to the first target. Likewise, positive non-adjacency facilitation scores indicate faster responses to the third than to the first target. A summary of the facilitation scores is shown in Figure 3.

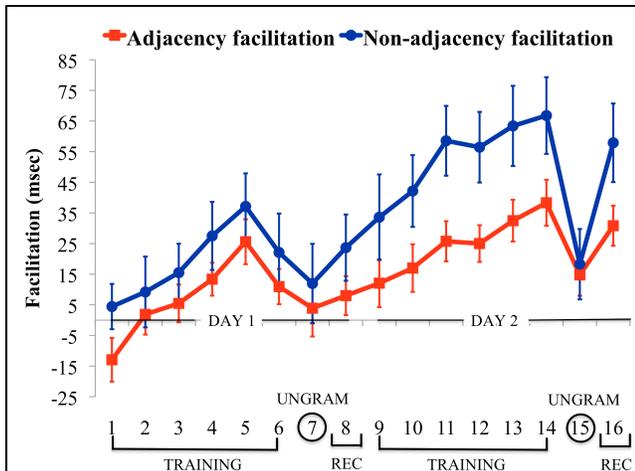


Figure 3: Mean adjacency and non-adjacency facilitation (msec) across training blocks and days (ungram = ungrammatical block, rec = recovery block).

There were two clear patterns in the data. First, the two facilitation curves showed remarkably similar increasing trends across the grammatical training blocks (except for block 6 in session 1, featuring an unexpected drop in facilitation perhaps due to fatigue). Second, there was more non-adjacency than adjacency facilitation, especially in the second session. Results from a 2 (Type: adjacency vs. non-adjacency facilitation) x 2 (Day: 1 vs. 2) x 8 (Block: 1 to 8) within-subject ANOVA on mean facilitation scores per block confirmed that, overall, facilitation scores changed as a function of blocks in both the training sessions (main effect of Block, Greenhouse-Geisser corrected for violation of sphericity assumption:  $F(4.72, 99.02) = 8.63, p < .001$ ). The overall amount of non-adjacency facilitation indeed exceeded that of adjacency facilitation (19 ms difference, main effect of Type:  $F(1, 21) = 10.49, p = .004$ ), and this difference was significantly larger in the second than in the first training session (Type x Day interaction was significant:  $F(1, 21) = 7.73, p = .01$ ). Analyses of simple effects showed that the difference between the adjacency and non-adjacency facilitation scores was only marginally significant in the first session (12 ms difference,  $p = .07$ ), but highly significant in the second session (25 ms difference,  $p = .001$ ). No other interaction effects were significant.

To assess learning further, we focused on the final three training blocks of each session and performed two planned comparisons (last training vs. ungrammatical block, and recovery vs. ungrammatical block) on the adjacency and non-adjacency facilitation scores. Significantly greater facilitation in the last training and recovery block compared to that in the ungrammatical block would indicate that the participants had acquired the trained patterns. In the first session, neither the adjacency nor the non-adjacency facilitation scores in the last training or recovery block differed significantly from the score in the ungrammatical block ( $t_s \leq 1$ ), indicating that participants had not learned

the adjacent or non-adjacent dependencies. However in the second session, clear evidence for non-adjacent dependency learning emerged (last training vs. ungrammatical block: 49 ms facilitation;  $t(21) = 6.19$ ; recovery vs. ungrammatical block: 40 ms facilitation.  $t(21) = 3.99$ , both  $p_s < .001$ ). The facilitations scores for adjacent dependency learning were likewise significant (last training vs. ungrammatical block: 24 ms,  $t(21) = 3.08, p < .01$ ; recovery vs. ungrammatical block: 16 ms,  $t(21) = 2.37, p < .03$ ).

Although an earlier emergence of sensitivity to adjacent dependencies was not evident in the online data averaged across blocks, such a pattern might have emerged and then faded extremely early during training, so that the *adjacency-first* pattern could be revealed only if the earliest training trials were examined. To assess that possibility, we divided the first training block into four sub-blocks (24 trials each) and tracked adjacency and non-adjacency facilitation scores across the sub-blocks. As Figure 4 shows, the data did not support the *adjacency-first* hypothesis, as the adjacency facilitation scores were overall lower than the non-adjacency facilitation scores (however, none of the facilitation scores were significantly above the first-target baseline in any of the sub-blocks).

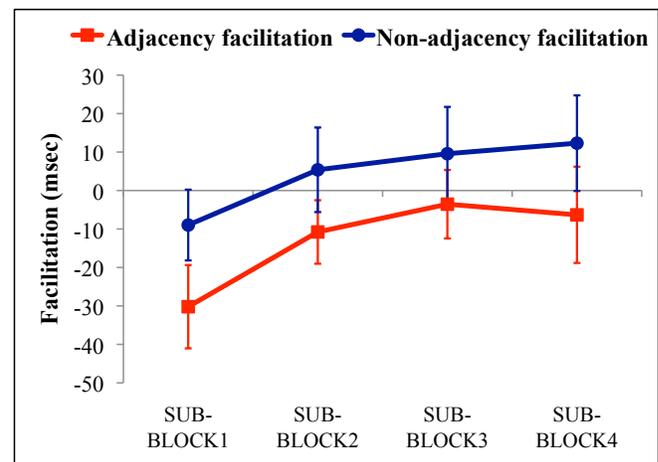


Figure 4: Mean adjacency and non-adjacency facilitation (msec) across sub-blocks of block 1.

Participants scored an average of 64% correct ( $SD = 20.2$ ) on the prediction task and 62% correct ( $SD = 20.6$ ) on the judgment task. Consistent with the positive RT non-adjacent dependency learning results in the second session, performances were significantly above chance level (i.e., 50%) on both tasks ( $t(21) = 3.34, p = .003$ , and  $2.81, p = .01$ , respectively).

## Discussion

The participants acquired the adjacent dependencies, as evidenced by the significantly greater facilitation in the last training and recovery block compared to that in the ungrammatical block in the second session. They also acquired the non-adjacent dependencies, as evidenced by the significant facilitation and accuracy results in the same

session. Consistent with prior studies that trained participants in one session only (e.g., Gómez, 2002), we found no evidence of non-adjacent dependency learning in the first session. The significant learning found in the second session demonstrates that extended training is worthwhile if the goal is to track gradual acquisition of complex patterns.

In accordance with the higher informativeness of the non-adjacent statistics compared to the adjacent statistics used in this study, the non-adjacent dependencies were learned better than the adjacent ones. Importantly, at no point during training was sensitivity to adjacent dependencies greater than sensitivity to non-adjacent dependencies. Despite the relatively small adjacent set size, participants did not first focus on adjacent dependencies and then on non-adjacent dependencies during learning. The current results therefore do not support the *adjacency-first* view of non-adjacent dependency learning (Gómez, 2002). Instead, they suggest that adjacent and non-adjacent dependency learning processes can run in parallel as learners are exposed to strings containing learnable adjacent and non-adjacent dependencies. It should, however, be noted that we used quite different probabilities for the two dependency types, much higher for non-adjacent than for adjacent dependencies (in keeping with the smaller adjacent set size conditions in Gómez, 2002). It remains to be seen whether an adjacency-first bias would arise when the relative strengths of the two dependency types are more comparable.

An issue that is yet to be settled is whether adjacent and non-adjacent dependency learning involves the same or different mechanisms. In a brain-imaging AGL study, Friederici, Bahlmann, Heim, Schubotz, and Anwander (2006) found additional recruitment of Broca's area (BA 44/45) in the processing of trained sequences that contained hierarchically nested non-adjacent dependencies (e.g.,  $A_1A_2B_2B_1$ ), as compared to the brain activation pattern for the processing of trained sequences containing adjacent dependencies (e.g.,  $A_1B_1A_2B_2$ ). This suggests a specific brain area, and presumably a separate mechanism, for the computation of hierarchical non-adjacent dependencies (see also Bahlmann, Schubotz, & Friederici, 2008, but see de Vries, Monaghan, Knecht, & Zwitserlood, 2008). Weakening that claim, however, are convergent findings from a number of recent studies showing that Broca's area is also engaged in artificial grammar learning that does not involve hierarchical non-adjacent dependencies (Christiansen, Kelly, Schillcock, & Greenfield, 2010; Petersson, Folia, & Hagoort, in press; Uddén et al., 2008).

The time course data in our study do not reveal whether learning the adjacent and non-adjacent dependencies engaged one learning mechanism, which simultaneously extracted both patterns, or whether distinct learning mechanisms were involved. To explore the relationships between adjacent and non-adjacent dependency learning in a different way, we computed by-subject online learning scores using the RT facilitation in the first training block as baseline and subtracting it from the facilitation obtained in

the last training block of the second session (block 14; see also Misyak et al., 2010). We found a strong positive correlation between the adjacency and non-adjacency online learning scores across participants ( $r = .78, p < .001$ , see Figure 5)<sup>1</sup>. While the absence of a correlation or a negative correlation would have pointed towards distinct learning mechanisms, the presence of correlation cannot be unambiguously interpreted. It certainly suggests a shared learning mechanism, but it could also be the case that the acquisition of adjacent and of non-adjacent dependencies both depends on a trait or ability that was not assessed in our study (see, e.g., Kaufman, DeYoung, Gray, Jiménez, Brown, & Mackintosh, 2010).

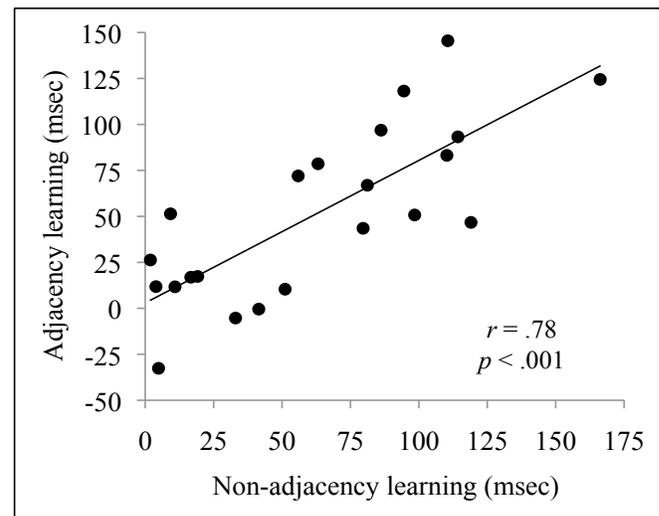


Figure 5: Correlation between adjacency and non-adjacency learning.

Finally, we found better learning in the second than in the first training session. This is somewhat reminiscent of consolidation effects found in other verbal learning tasks (e.g., Davies & Gaskell, 2009; Dumay & Gaskell, 2007), though we can, of course, not separate the effects of increased practice and memory consolidation over time nor do we know whether sleep was essential for obtaining any consolidation effects. The finding that learning increments in the second session were larger for non-adjacent than for adjacent dependencies may point towards the involvement of different learning or consolidation mechanisms. Alternatively, the divergence in learning rates in the second session might simply reflect cumulative differences with extended exposure that stemmed directly from the difference in statistical strength for the adjacent vs. non-

<sup>1</sup> A similar correlation obtained ( $r = .74, p < .001$ ) when adjacency and non-adjacency learning scores were computed without involving the common first element in the baseline, by subtracting second-element residual RTs at the final training block from those at the first training block (adjacency learning), and by subtracting third-element residual RTs at the final from those at the first block (non-adjacency learning).

adjacent dependencies. To assess these options future studies are required, where one could, for instance, systematically vary the retention intervals between training sessions, the timing of the sessions (with/without intervening sleep), and the adjacent and non-adjacent dependencies in the training sequences.

In sum, the current study demonstrates the usefulness of obtaining online data for assessment of statistical learning (see also Misyak & Christiansen, 2010; Misyak et al., 2010). Through this approach, we obtained compelling evidence for simultaneous adjacent and non-adjacent dependency learning in an extended artificial grammar learning task. Thus, our results suggest that statistical learning is both more powerful, making it possible to learn multiple types of dependencies simultaneously, and more robust, allowing non-adjacent dependencies to be learned without high adjacent-set variability, than previously thought. This provides further support for the hypothesis that statistical learning may play a crucial role in the acquisition of long-distance dependencies in natural language.

### Acknowledgments

We would like to thank Katrin Scheibe and Ceci Verbaatschoot for assisting with the preparation of the materials and with participant testing. We also thank three anonymous reviewers for their helpful comments on an earlier version of this paper.

### References

- Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage*, *42*, 525–534.
- Christiansen, M. H., Kelly, L.M., Shillock, R., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, *116*(3), 382-393.
- Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of non-adjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(5), 1119–1130.
- Davis, M. H. & Gaskell, M. G. (2009). A complementary systems account of word learning: neural and behavioural evidence. *Philosophical Transactions of the Royal Society B*, *364*, 3773-3800.
- De Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition*, *107*, 763–774.
- Dumay N., Gaskell M. G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science*, *18*, 35–39.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348-368.
- Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, 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.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.
- Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jimenez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition*, *116*, 321-340.
- Lany, J., & Gómez, R.L. (2008). 12-Month-Olds Benefit from Prior Experience in Statistical Learning. *Psychological Science*, *19*, 1274-1252.
- Lany, J., & Gómez, R.L., & Gerken, L.A. (2007). The role of prior experience in language acquisition. *Cognitive Science*, *31*, 481-507.
- Marcus, G. F., Vijayan, S., Rao, S. B., & Vishton, P. M. (1999). Rule learning by seven-month-old infants. *Science*, *283*(5398), 77-80.
- Misyak, J. B., & Christiansen, M. H. (2010). When 'more' in statistical learning means 'less' in language: individual differences in predictive processing of adjacent dependencies. *Proceedings of The 32<sup>nd</sup> Annual Conference of the Cognitive Science Society* (pp. 2686 – 2691).
- Misyak, J.B., Christiansen, M.H. & Tomblin, J.B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*. doi:10.3389/fpsyg.2010.00031.
- Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. *Cognitive Psychology*, *48*(2), 127–162.
- Onnis, L. Christiansen, M., Chater, N. & Gómez, R. (2003) Reduction of uncertainty in human sequential learning: Evidence from artificial language learning. *Proceedings of The 25th Annual Conference of the Cognitive Science Society*. (pp. 886-891). Mahwah, NJ: Lawrence Erlbaum.
- Petersson, K. M., Folia, V., Hagoort, P. (in press). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*.
- Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, *12*, 110–114.
- Saffran, J.R., Aslin, R. N., & Newport, E.L. (1996). Statistical learning by 8-month old infants. *Science*, *274*, 1926-1928.
- Seidenberg, M. S., & Elman, J. L. (1999). Do infants learn grammar with algebra or statistics? *Science*, *284*, 434-435.
- Uddén, J., Folia, V., Forkstam, C., Ingvar, M., Fernández, G., Overeem, S., Van Elswijk, G., Hagoort, P., & Petersson, K. M. (2008). The inferior frontal cortex in artificial syntax processing: An rTMS study. *Brain Research*, *1224*, 69-78.
- Yang, C.D. (2004). Universal grammar, statistics or both? *Trends in Cognitive Sciences*, *8*, 451–456.