Constraints on Non-Adjacent Dependency-Learning: Distance Matters
An Artificial Grammar Learning Study with Adults

Ileana Grama, Frank Wijnen & Annemarie Kerkhoff

Utrecht University, Utrecht Institute of Linguistics OTS, the Netherlands

Introduction

Natural languages display a variety of structures where grammatical morphemes instantiate non-adjacent dependencies:

(1) a. **Nous** part**ons** demain. (French)
We leave.**1stpl**. tomorrow
b. l**a** bambin**a** / **il** bambin**o** (Italian)
the.**fem** child.**fem** / the.**masc** child.**masc**
c. Vandaag **heb** ik de dokter **ge**beld (Dutch)
Today **have** I the doctor **PART.**called.

In the examples above, the morphemes in bold (pronouns, auxiliaries, determiners, inflections) predict each other with a high degree of probability: for instance, in (1a) the form of the pronoun determines the exact form of the inflection on the verb stem. When either one of them is changed, the sentence becomes ungrammatical: ****Nous** part**ez** demain*, ****Vous** part**ons** demain*. This one-to-one correspondence is indicative of a more abstract relationship between any verb and its subject, whether or not it is overtly marked: agreement. Agreement not only functions at sentence level (1a), but also within smaller phrases, like nominal phrases (1b). Other formal relationships besides agreement can be overtly marked, for instance the one between the functional domain of auxiliaries and the lexical domain of verbal predicates, as in (1c).

In short, languages possess means to overtly mark important formal syntactic relationships[1]: this marking presupposes a correspondence between two (non-adjacent) functional morphemes, so that a specific morphological form of one morpheme will predict the specific form of the other morpheme. If a child were in possession of a cognitive mechanism that allowed it to detect and retain patterns of co-occurrence between functional markers, this mechanism could mediate (at least in part) the inference of syntactic rules or patterns from surface (distributional) properties of the input.

The existence of such a mechanism has been extensively investigated in the literature to date: artificial grammar studies have confirmed that both infants and adults are capable of extracting non-

[1] Note that the existence of morpho-syntactic dependencies does not merely signal the presence of a syntactic relationship, but also carries information on the specific nature of that relationship: thus, agreement in (1a) depends on number and person, and does not vary as a function of gender, whereas in (1b) the relationship depends on gender (and number), but not person.

adjacent dependencies (NADs) from highly simplified strings of nonsense words (…aXb…), where the presence of one token *a* predicts the (non-adjacent) occurrence of a specific distinct token *b* with 100% certainty (e.g. Gómez, 2002, Peña et al., 2002, Onnis et al., 2004, Newport & Aslin, 2004, Gómez & Maye, 2005, Endress & Bonatti, 2007, Endress & Mehler, 2009).

Gómez (2002) showed that when exposed to a simple language composed of aXb strings such as 'pel wadim rud' (presented in sequences separated by 750ms pauses), adult participants learned the dependencies between monosyllabic a and b elements (three distinct dependencies), but only when the set of intervening bisyllabic Xs was sufficiently large. After familiarization, subjects performed above chance in discriminating correct $a_i\_b_i$ from incorrect $*a_i\_b_j$ dependencies when X could be instantiated by 24 different nonce words, but not 12, showing that the high variability of elements spanned by a dependency facilitates NAD-learning. Gómez & Maye (2005) replicated this finding with infants of 18 months, but found that 12-month-olds did not perform above chance even in the high variability condition, suggesting that (perhaps due to a working memory limitation) it is only around 15 months that infants begin tracking non-adjacent co-occurrence patterns in spoken input.

Studies on natural language acquisition have suggested that, around the same age that infants show learning of non-adjacent dependencies in artificial grammar paradigms, they also become sensitive to non-adjacent dependencies in their native language. Santelmann & Jusczyk (1998) confirmed that English-learning infants of 18, but not 15 months were able to discriminate between sentences like the grammatical (2a) and the minimally distinct, ungrammatical (2b):

(2) a. At the bakery, everybody **is** bak**ing** bread.
b. *At the bakery, everybody **can** bak**ing** bread.

Subsequent studies have confirmed this early sensitivity to morpho-syntactic dependencies in German (Höhle et al., 2006) French (Heugten & Shi, 2010) and Dutch (Wilsenach & Wijnen, 2004; van Heugten & Johnson, 2010), and have attempted to correlate acquisition of these dependencies with their distributional properties, showing that frequency, probability of co-occurrence or relative proximity (average length spanned) of some dependencies can correlate with the order of their acquisition (van Heugten & Johnson, 2010, but see Tincoff, Santelmann, & Jusczyk, 2000).

Positional Factors in NAD-learning

Research with artificial grammars has shown that NAD-learning is a highly constrained process (Gómez, 2002; Peña et al., 2002; Newport & Aslin, 2004; Endress & Bonatti, 2007; Endress & Mehler, 2009). Apart from variability (Gómez, 2002), other cues have been shown to contribute to the detection of dependencies in input: Peña et al. (2002) showed that participants exposed to a continuous string of concatenated aXb syllables such as 'rakibe' (with 'a' predicting 'b' with 100%

probability) did not prefer novel aX'b strings, or 'rule-words' (with correct dependencies, but novel Xs), over baX strings, or 'part-words', that they had heard in the familiarization (although they did prefer familiar aXb strings over the latter). However, when subtle 25ms pauses were inserted at the edges of the aXb chunks, participants were sensitive to the a_b dependencies even with novel Xs. The authors concluded that learning dependencies as generalizable *rules*[2] requires subtle segmentation cues to mark the aXb phrases (but see Onnis et al., 2004 for counterevidence). Endress & Mehler (2009) extended these results by suggesting that the necessity of segmentation cues was in fact related to the positional salience of the dependent elements: when aXb strings are isolated even by subtle pauses, the dependent items occupy string-initial and, respectively, string-final positions. Previous research (Endress et al., 2005; Endress et al., 2007; Endress & Mehler, 2010) showed that string-peripheral positions are highly salient to rule-learning, as it is easier for learners to encode the first and last items in a string than elements in string-medial position. Statistical leaning, as opposed to rule-learning, it was suggested, is not sensitive to positional salience. Endress & Mehler (2009) tested learning a_b dependencies both in string-peripheral (aXYZb) and string-medial (XaYbZ) positions: in each configuration, they were interested in whether participants would prefer 'rule-words' (a_iX'Y'Z'b_i or X'a_iY'b_iZ') over 'part-words' (YZ$b_i$$a_j$X or Y$b_iZXa_j$), and 'class-words' ($a_i$X'Y'Z'$b_j$ or X'a_iY'b_jZ') over 'part-words'.[3] While rule-words were preferred independent of configuration, class-words were only learned when the a and b 'classes' were instantiated at the edges of strings (subtly segmented by 25ms pauses). Endress & Mehler proposed a distinction between NAD rule-learning (learning the classes of a and b tokens, or 'class-words') and NAD statistical learning (learning the dependency between specific a and b tokens, or 'rule-words'), and showed that while statistical learning succeeded both in string-peripheral and string-medial configurations, rule-learning was constrained to string-peripheral positions.

The line of research put forth by Endress, Mehler and colleagues is directed at bringing evidence in favor of the existence of two separate computational mechanisms, one statistical, the other based on generalizations, which can be dissociated by their reliance on salience cues in the input. However, this does not directly afford clear predictions as to the cues necessary to acquire morpho-syntactic dependencies in a natural-language learning situation. To master morpho-syntactic patterns in their native language, infant L1 learners must capture the correct one-to-one mapping of grammatical morphemes (which, following the definitions in Endress & Mehler, 2009, should be an

---

[2] Peña et al. (2002), as well as subsequent work by Endress and colleagues, postulate the existence of two separate mechanisms: statistical learning, which deals with co-occurrence patterns between specific entities, and rule learning, with computes abstract patterns that are generalizable to novel contexts. Under this assumption NADs can be either statistical regularities or generalizable rules.

[3] 'Rule-words' respected the dependency between specific a_i and b_i items, 'class-words' respected the a and b categories and their position (a_i occurred in an a position, b_i in a b position), but violated the correspondence between specific a_i and b_j items; both were novel as they had not been heard as such during familiarization; 'part-words' violated both the position of and the dependencies between specific a_i and b_j items, but were a familiar sequence, as they represented the end of a 'word' and the beginning of another 'word' from familiarization.

instance of statistical learning), but at the same time integrate that mapping into a more abstract ‘rule’ (e.g. for agreement). They must be aware of the classes of morphemes that correlate (e.g. auxiliary and verb suffix), but also of the specific mappings between elements of those classes. Adopting Endress & Mehler’s terms, part of obtaining ultimate mastery of a dependency is being able to prefer ‘rule-words’ over ‘class-words’ – it would therefore be interesting to see how this preference would interact with factors of positional salience.

At the same time, the distributional properties of the input can play a crucial role in how NADs are learned. Onnis et al. (2005) showed that, contrary to Peña et al. (2002), dependencies could be learned as generalizations (recognized in novel aX’b strings) even in the absence of segmentation cues, provided that the variability of the intervening Xs was large enough in the familiarization phase (cf. Gómez, 2002). In natural languages, this variability is afforded freely by the universal contrast between open-class vs. closed-class items: items that instantiate morpho-syntactic dependencies are closed-class, or functional elements, and are therefore more frequent and invariable; words that intervene are open-class, or lexical elements (nouns, verbs, adverbs, etc.) and as such are less frequent because they are highly interchangeable (variable). It would be interesting, therefore, to see how positional salience interacts with this variability: does variability override the effect of positional configuration, or will the edges still hold an advantage for NAD-learning?

Furthermore, what counts as positional salience for a natural language learner? Morpho-syntactic dependencies ((1a,c), (2)) do not necessarily occur at the edges of sentences or even phrases, and infants as young as 17 months can track agreement patterns even when the dependent morphemes are in different phonological phrases (van Heugten & Shi, 2010). A more likely constraint of positional configuration for NAD-learning, is the distance between the dependent elements. Eighteen-month-olds could track the *is_ing* dependency (2) across three syllables (a bisyllabic adverb and a monosyllabic verb stem) – however, when the distance between the morphemes became larger than 3 syllables infants lost their sensitivity to the dependency (Santelmann & Jusczyk, 1998). The authors attributed this to the limited working memory capacity of their young subjects. The question, however, remains whether this limitation is truly due to reduced processing abilities of 18-month-olds, or whether a more general constraint limits the window in which NADs can be not only recognized, but also learned.[4]

In this study we investigate the role of positional cues to NAD-learning, in an artificial grammar learning paradigm with adult subjects. We adopt the original design from Gómez (2002) and Gómez & Maye (2005), adapted for Dutch subjects in Kerkhoff et al. (2013), to address the questions sketched above: What is the role of edge salience in learning dependencies between specific tokens? What is the role of distance in tracking remote dependencies? How can artificial grammar learning

[4] Here, we recognize the distinction between identifying, or ‘learning’ a novel dependency from repeated exposure on the one hand (the knowledge tested in artificial grammar learning), and recognizing an already known dependency (or violation thereof), on the other hand (as tested in, e.g., Santelmann & Jusczyk, 1998).

studies inform our expectations about the cues found in natural language and their role in learning morpho-syntactic patterns?

Experiments 1 (Part-Edge), 2 (Edge, High Variability) and 3 (Edge, Low Variability)

We tested grammaticality judgement of correct $a_i\_b_i$ vs. incorrect $a_i\_b_j$ dependencies in two positional configurations: XaYb (Part-Edge), where the initial dependent item was not in a string-peripheral position, but the dependency spanned only one intervening element Y, and aXYb (Edge), where the dependent items were both at the edges, but the dependency spanned two bisyllabic words X and Y[5]. Experiment 1 tested NAD-learning in a Part-Edge configuration, Experiment 2 tested NAD-learning in an Edge configuration with High Variability (the number of unique combinations between a, X, Y and b was matched with Experiment 1, but, because in Experiment 2 there were two words between a_b instead of one, the variability *within* the a_b dependency was higher), and Experiment 3 tested NAD-learning in a Low Variability Edge configuration (there were fewer unique combination between a, X, Y and b, but the variability between a_b was matched with that in Experiment 1, i.e. there were as many XY pairings as there were Y elements).

If edge salience is crucial to NAD-learning, participants will fare poorly in the first experiment, where one of the dependent elements is string-medial, but will be successful in the second, where the dependency is string-peripheral; if, on the other hand, edge salience is not crucial, participants should perform above chance in the first experiment as well. On the other hand, if distance is a crucial factor in NAD-learning, participants in the Edge experiment are predicted to perform poorly due to the increased distance spanned by the dependency (2+2=4 syllables, which in the Santelmann & Jusczyk study was shown to be too large a span for 18-month-olds), and participants in the Part-Edge experiment should perform better. In keeping with research on sensitivity to natural language dependencies (Santelmann & Jusczyk, 1998; van Heugten & Shi, 2010) we predicted that distance, but not edge salience, would be crucial to the detection of the non-adjacent patterns.

By adopting a design similar to Gómez (2002) we aimed to better imitate the distributional properties of natural languages (cf. Onnis et al., 2005), and make our results more amenable to predictions regarding acquisition. By employing adult participants we aimed to eliminate the confound of younger infants' supposedly limited processing capacities, and directly test the hypothesis that the distance limitation on NAD-tracking is a general constraint that applies to all age groups.

[5] Because the X and Y elements were bisyllabic, we did not test the dependencies as completely embedded (e.g. XaYbZ) as the strings would have been too long and complex.

Methods

Participants

Participants in all three experiments were recruited via email. All subjects that reported hearing impairments or attention deficits, or who were not native speakers of Dutch were excluded. The experiment lasted 20 minutes, and participants received a small (5 euro) reimbursement for their effort. For Experiment 1, 34 participants (2 male) aged between 19 and 30 years (M = 22.15) were recruited. Thirty-four participants were recruited for Experiment 2 (7 male), aged between 20 and 30 years (M = 22.97) – two were excluded due to familiarity with research on NAD-learning (University students in linguistics). For Experiment 3, 31 participants (3 male), aged between 20 and 30 years (M = 21.71) were recruited.

Materials

We tested the importance of positional information in 3 experiments: in Experiment 1 (Part-Edge), subjects were tested on their sensitivity to partly-embedded a_b dependencies, in four-word, XaYb, strings. Subjects were assigned to one of two different languages, LANG1 and LANG2; three a_b dependencies between monosyllabic elements were created for both LANG1 (a1_b1, a2_b2, a3_b3) and LANG2 (a1_b2, a2_c3, a3_c1, see Appendix 1), such that the grammatical dependencies in one language were ungrammatical in the other. The X and Y elements were selected from two separate sets of 18 elements each. Pairs of X and Y were formed so that each X was paired with 6 of the 18 Y elements, and each Y was paired with 6 Xs (making the transitional probability between X and Y 0.16). The pairs thus obtained (18x6=108) were combined exhaustively with the three dependencies in each language for a total of 324 strings per language, which were semi-randomized for each participant (never allowing the same dependency to occur more than three times in a row).

The test strings were created by selecting two novel XY pairs from the set of combinations that had not been used in the familiarization phase. This was meant to ensure that subjects did not merely recall entire strings from the training phase. Test strings were formed by combining each XY pair with all six dependencies in LANG1 and LANG2, resulting in 12 different strings; these strings were randomized for each participant, and for each string participants were asked to judge whether it was consistent with what they had heard or not. We measured participants' endorsement ('yes' answers) for strings containing dependencies that were consistent with their language of exposure (Consistent, i.e. "hits"), and compared this to endorsement for dependencies that were inconsistent with their language of exposure (Inconsistent, i.e. "false alarms"). Participants learned the dependencies if they endorsed consistent test items significantly more often than inconsistent ones.

Experiment 2 (Edge, high variability), was identical in all respects to Experiment 1, except that the structure of the strings, instead of being XaYb, was aXYb; this ensured that the a_b dependency was string-peripheral, but, at the same time, that the distance spanned was greater (2 words, i.e. four syllables). The same XY combinations were used, so that the stimuli were completely identical except for the order of words in a string.

Experiment 3 (Edge, low variability), was identical in all respects to Experiment 2, except that instead of 108XY pairs, we used only 18 XY pairs, and repeated the 54 aXYb combinations thus obtained six times (for the 324 familiarization tokens). This was done to control for the fact that, in Experiment 1, although there were 108 XY pairs, the variability *between* the dependent a_b elements was only 18 (as there were only 18 different Ys). Hence, Experiment 3 was matched with Experiment 1 for intra-dependency (between a_b) variability (while Experiment 2 was matched with Experiment 1 for overall variability). In this way, we also wanted to eliminate the potentially confounding factor that performance in Experiment 2 might actually be negatively affected by the large variability / number of XY pairs (108) occurring between a_b.[6]

Stimuli

The 3 monosyllabic a and b nonce words, along with 18 bisyllabic Xs were taken from Kerkhoff et al. (2013); 18 new Y elements were created and submitted to the approval of 11 different native speakers, to check for their suitability as Dutch-sounding nonce words, and to control for their similarity to real Dutch words. The stimuli (Appendix 1) were recorded using a Grundig Fine-Arts High-Definition DAT Recorder (DAT-9009), at a wave frequency of 48 kHz. A female voice read out YaXb strings with a lively intonation and initial stress on every word (e.g. klepin tep poemer lut, cf. Appendix 1 for IPA transcription), with special emphasis placed on the monosyllabic dependent elements. The individual elements were subsequently spliced from the recording and concatenated into strings as indicated above (for all three experiments), with 250ms pauses separating every two words in a string (i.e. Y from a, a from X, and X from b, as in Gómez, 2002). A total of 324 strings were obtained per language, and 12 strings were concatenated for the test phase.

Procedure

Subjects were told they would listen to short ‘sentences’ in an ‘alien language’, which had certain regularities pertaining to ‘word-order’. Participants were randomly assigned to one of two artificial languages, LANG1 and LANG2, respectively. The ‘sentences’ were presented over speakers, in a sound-attenuated booth; participants were seated at a table and attended to the task of coloring a mandala while listening to the stimuli. They could follow the progress of the training phase on a computer screen which showed a countdown of the strings from 324 to 1. Familiarization lasted 19 minutes. In the test phase, subjects were presented with 12 individual test strings and asked if they ‘belonged to the language they had just learned’ in the training phase. Subjects responded by pressing

[6] Although Gómez (2002) showed that intra-dependency variability affects learning positively (the more variability between a_b, the better the dependency is learned), there is still the possibility of a superior limit on variability, that is, of a threshold above which variability becomes more confusing than helpful.

one of two buttons marked ‘Yes’ or ‘No’. Before the start of the test phase, subjects were told that half of the sentences would conform to their training language. Hence, six strings were consistent with the word-order patterns in the training, and 6 were not. Subjects were informed that the patterns were not consciously detectable, but that they would have to rely on their intuition to answer the questions. Both the order of training stimuli and of the test strings were randomized for each participant, irrespective of language.

Results

For each Experiment, we compared subject’s endorsement (‘yes’ answers) of Consistent and Inconsistent strings. Percentages of endorsement for consistent and inconsistent test items for all three experiments are presented in Table 1. For each experiment, we ran a Repeated Measures ANOVA with Consistency as a within-subject, and Language (LANG1, LANG2) as a between-subjects factor. For Experiment 1 (Part-Edge), there was a significant effect of Consistency, $F(1, 32) = 4.356$, $p = .045$, $\eta^2 = .120$, indicating that endorsements for consistent strings were higher than endorsements for inconsistent strings. There was no significant effect of Language ($p = .859$), and no interaction between Consistency and Language ($p = .646$). For Experiment 2 (Edge, high variability), we obtained no significant effects for either Consistency ($p = .149$), Language ($p = .930$), or the interaction between them ($p = .547$). Similarly, for Experiment 3 (Edge, low variability), there was no effect of Consistency ($p = .643$), or Language ($p = .373$), and no interaction ($p = .526$).[7]

		Experiment 1 (Part-Edge)	**Experiment 2 (Edge, high var.)**	**Experiment 3 (Edge, low var.)**
	String structure	**XaYb**	**aXYb**	**aXYb**
All learners	**Number**	34	32	31
	Consistent vs. Inconsistent	69.12% (SD=21.76) 60.29% (SD=17.89)	69.79% (SD=24.47) 56.25% (SD=32.44)	65.05% (SD=17.93) 62.07% (SD=26.52)
Explicit learners	**Number**	1	7	3
	Consistent vs. Inconsistent	83.33% 33.33%	100% (SD=0) 4.76% (SD=12.59)	77.78% (SD=9.62) 16.67% (SD=28.86)
Implicit learners	**Number**	33	25	28
	Consistent vs. Inconsistent	68.69% (SD=21.95) 61.11% (SD=17.51)	61.33% (SD=20.81) 70.67% (SD=18.18)	63.69%(SD=18.18) 67.26% (SD=21.51)

Table 1: Percentage Endorsements for Consistent vs. Inconsistent strings

[7] A Repeated Measures ANOVA with Consistency as within- and Experiment as between-subjects factor revealed a significant effect of Consistency, $F(1) = 4.910$, $p = .029$ (indicating higher endorsement rates for consistent sentences), but no Interaction between Consistency and Experiment ($p = .51$), suggesting that the differences in performance between experiments were too small to reach significance.

After the test, participants were informally asked for feedback on the experiment they had just completed. They were asked what they had noticed about the language, and if they had any particular strategy in evaluating the test items. Although many participants were aware of the structure of the strings (that there were four words in a string, two mono- and two bisyllabic, with a specific order), few were explicitly aware of the dependencies. In Experiment 1, only one participant realized that there was a one-to-one correspondence between the monosyllabic elements. In contrast, 7 of the 32 participants in Experiment 2 were explicitly aware of dependencies, and 3 of the 31 participants in Experiment 3 discovered the regularity. In Table 1 we also present the endorsement rates for the explicit and implicit learners separately, showing that while explicit learners show a marked preference for consistent test items, implicit learners in Experiments 2 and 3 do not display a preference for consistent items at all.

Discussion

We tested subject's sensitivity to dependencies between non-adjacent elements in complex four-word strings in an artificial grammar learning paradigm. Participants exposed to strings of the type XaYb (Experiment 1, Part-Edge) showed a significant preference for Consistent $a_i\_b_i$ over Inconsistent $*a_i\_b_j$ dependencies (in strings with familiar words, but novel word-combinations[8]), suggesting that discriminating 'rule-words' from 'class-words', and thus sensitivity to one-to-one mappings between dependent elements, was possible even when the dependencies were not (entirely) string-peripheral. Participants' capacity to learn non-adjacent patterns embedded in complex strings is even more impressive if we consider the fact that they were distracted by a secondary task during familiarization, and seemed to be largely unaware of the patterns they were learning.

Conversely, irrespective of the variability of intervening material, participants exposed to string-peripheral, but more distant dependencies (aXYb, Experiment 2 and 3) were not able to reliably distinguish between Consistent and Inconsistent dependencies. This pattern of results suggests that increasing the distance between dependent elements renders NAD-learning more demanding; our study, identifying a distance constraint on adult NAD-learning in an artificial grammar learning paradigm, corroborates the findings of Santelmann & Jusczyk (1998), identifying a distance constraint on 18month-olds' sensitivity to morpho-syntactic dependencies, and prompts the hypothesis that distance is a general computational constraint on the detection of NADs. Future research is needed to shed light on the precise nature of this constraint: does NAD-learning degrade linearly as a function of distance, or is there a cut-off point (e.g. Santelmann & Jusczyk's three syllables) beyond which

[8] In particular, what was novel was the combination between X and Y; note, however, that the _aYb fragments of every XaYb test item had been had been heard 6 times in the familiarization phase.

learning becomes impossible?[9] Furthermore, how is this distance best quantified: number of syllables (as proposed by Santelmann & Jusczyk), number of words, or temporal duration?

Höhle et al. (2006) propose that distance might actually be quantifiable in terms of 'analyzability' of the intervening material: German infants were able to track the aspectual dependency *hat_ge-* (analogous to the English *have_en*) across determiner+noun phrases (categorized as nominal phrases due to the determiner *das/den*), but were unable to track the same dependency across adverbs of the same length (adverbs in German are not marked by any specific morpheme, unlike English adverbs which frequently end in the suffix *-ly*). The authors concluded that the child's ability to recognize/analyze the intervening material in a dependency may affect their ability to track the respective dependency; they suggested that, in Santelmann & Jusczyk (1998), infants' failure to track the dependency in the conditions with 4 or 5 intervening syllables was not due to linear distance as such, but to the fact that the adverbs used were less recognizable/familiar to the infants (few of them had the *–ly* marker). In our own study, although the X and Y classes were (structurally) consistent (X words appeared in X positions, and Y words in Y positions), and each X or Y item occurred six times in the familiarization phase, there was no overt marker for either the X or the Y class, which may have rendered X and Y elements harder to categorize and therefore less 'analyzable'. One way to test Höhle et al.'s hypothesis would be to introduce categorical markers for the X/Y classes in our materials for Experiments 2 and 3, and see if this improves performance on NAD-learning.

Although participants' discrimination performance did not reach significance in Experiments 2 and 3, it is interesting to note that numerically, the difference between endorsement rates for Consistent vs. Inconsistent test items is actually largest in Experiment 2 (see Table 1). This numerical pattern is likely carried by the seven explicit learners in Experiment 2 who displayed a ceiling effect (accepted all Consistent strings, rejected almost all Inconsistent ones, and were even able to recall at least some of the specific dependencies). On the other hand, implicit learners in this experiment even had higher endorsement rates for inconsistent items. Although the higher number of explicit learners in Experiment 2, and their markedly different performance compared to implicit learners, might be purely coincidental, it is worth considering a more interesting possibility: that the specific structural properties of the language may prompt or inhibit a specific type of learning. For instance, edge-positions might be salient for explicit recall, but not for implicit computations; or, contrastingly, distance constraints may operate on implicit learning alone, but leave explicit learning unaffected. Because NAD-learning studies so far do not seem to report subjects' explicit knowledge, or directly investigate the interaction between various salience factors and the explicit/implicit learning divide, it is difficult to make specific claims or predictions at this stage. We hope that this study serves as a

[9] Note that the possibility of a limited window of computation would be consistent with hypotheses of NAD-learning as abstraction of non-adjacent patterns from fixed n-grams, or chunks.

starting point for more extensive investigation into possible computational differences between implicit and explicit sensitivity to non-adjacent patterns.

Several final observations are required in order to obtain a balanced and correct perspective of the scope of our findings.

Firstly, sensitivity to part-edge dependencies in Experiment 1 does not completely overrule the hypothesis of an edge-salience effect. Learners were shown to perform above chance despite the lack of positional salience, but there is no reason to assume that this salience would not have had a facilitating effect – in Experiments 2 and 3, the importance of distance may have simply overridden this effect. Our study simply argues against the notion that this salience effect is absolutely crucial to detecting and learning non-adjacent patterns.

Secondly, our conclusions on the importance of distance to NAD-learning can be challenged on two accounts. On the one hand, note that the stimuli were spliced from recordings of full XaYb strings; thus, in Experiments 2 and 3, when the words were recombined into aXYb, switching the first two nonce words may have disrupted the overall prosodic pattern of the strings, rendering stimuli in Experiments 2 and 3 slightly less prosodically natural.[10] Furthermore, although learning was only observed in the first experiment and not the others, we did not find a statistically significant difference in performance between the 3 experiments, possibly due to the small effect sizes obtained. However, the fact that learning was significant only in Experiment 1, where the distance spanned by dependencies was smaller, is in line with the findings of Santelmann & Jusczyk (1998), and prompts further inquiry into the possibility of a distance restriction on NAD-learning.

The current study identifies a number of factors in NAD-learning that have so far received little attention, such as the importance of distance or the implicit/explicit learning divide. Why is there a distance limitation on NAD-learning? Future research could examine the effect of temporal duration, number of words or number of syllables, by manipulating pause length or the nature of the intervening material in an aXYb configuration (e.g. would an intervening tetrasyllabic word, instead of two bisyllabic XY words, facilitate NAD-learning?). Also, individual differences in working memory capacity could correlate with performance on NAD-learning at a distance, if Santelmann & Jusczyk (1998) were correct in their claim that working memory limitations are at stake when sensitivity to non-adjacent patterns breaks down. On the other hand, categorical marking of the X and Y classes would facilitate NAD learning if the relevant factor were not linear distance, but processing load, or 'analyzability' of the intervening material. Furthermore, the importance of 'implicit' learning could be looked into by instructing participants to explicitly look for rules, and then observing how this

[10] Recent findings from our lab suggest that acoustic and prosodic factors may play a substantial role in NAD-learning; subjects presented with aXb strings with an unnatural prosodic contour (rising contour on the final element) ignored the 750ms pauses between aXb strings, and segmented baX strings instead (the X elements did have a falling pitch contour on their final syllable). Learners, therefore, may be highly sensitive to prosodic cues, and performance on dependency-learning could be affected by unnatural prosody.

strategy influences learning of dependencies in different configurations like XaYb or aXYb. Finally, examining the effects of distance and categorical marking on artificial grammar NAD-learning with infants should confirm the relevance of all these cues to young L1 learners, or, on the contrary, point out crucial differences between infant and adult learning patterns.

References

Endress, A.D., B.J. Scholl & J. Mehler (2005). The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General*, 134(3), 406-419.

Endress, A.D. & L.L. Bonatti (2007). Rapid learning of syllable classes from a perceptually continuous stream. *Cognition*, 105(2), 247-299.

Endress, A.D., G. Dehaene-Lambertz & J. Mehler (2007). Perceptual constraints and the learnability of simple grammars. *Cognition*,105(3), 577-614.

Endress, A.D.& J. Mehler (2009). Primitive computations in speech processing. *The Quarterly Journal of Experimental Psychology*, 62(11), 2187-2209.

Endress, A.D. & J. Mehler (2010). Perceptual constraints in phonotactic learning. *Journal of Experimental Psychology: Human Perception and Performance*, 36(1), 235-250.

Gómez, R.L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431-436.

Gómez, R.L. & J. Maye (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, 7, 183–206.

Höhle, B., M. Schmitz, L. M. Santelmann & J. Weissenborn (2006). The recognition of discontinuous verbal dependencies by German 19-month-olds: Evidence for lexical and structural influences on children's early processing capacities. *Language Learning and Development*, 2(4), 277-300.

Heugten, M. van & E.K. Johnson (2010). Linking infants'distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, 63, 197-209.

Heugten, M. van & R. Shi (2010). Infant's sensitivity to non-adjacent dependencies across phonological phrase boundaries. *Journal of the Acoustical Society of America*, 128, EL223-EL228.

Kerkhoff, A.O., E. de Bree, M. de Klerk & F.N.K. Wijnen (2013). Non-adjacent dependency learning in infants at familial risk of dyslexia. *Journal of Child Language*, 40(1), 11-28.

Newport, E. L. & R.N. Aslin (2004). Learning at a distance 1. Statistical learning of nonadjacent dependencies. *Cognitive Psychology*, 48, 127-162.

Onnis, L., P. Monaghan, M.H. Christiansen & N. Chater (2004). Variability is the spice of learning and a crucial ingredient for detecting and generalizing in nonadjacent dependencies. *Proceedings of the 26th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum, 1047-1052.

Peña, M., L.L. Bonatti, M. Nespor & J. Mehler (2002). Signal-driven computations in speech processing. *Science*, 298, 604-607.

Santelmann, L. M. & P.W. Jusczyk (1998). Sensitivity to discontinuous dependencies in language learners: evidence for the limitations in processing space. *Cognition*, 69, 105-134.

Tincoff, R., L.M. Santelmann & P. Jusczyk (2000). Auxiliary verb learning and 18-month-old's acquisition of morphological relationships. In Howell, S.C., S. Fish & T. Keith-Lucas (eds.), *BUCLD 24:*

Proceedings of the 24th Annual Boston University Conference on Languag Development, Sommerville, MA: Cascadilla Press, 726-737.

Wilsenach, A.C. & F.N.K. Wijnen (2004). Perceptual sensitivity to morphosyntactic agreement in language learners. Evidence from Dutch children at risk for developing dyslexia. In A. Brugos & L. Micciulla (Eds.), *Proceedings of 28th BU Conference on Language Development*. Somerville, MA: Cascadilla Press, 645-656.

Appendix 1: list of stimuli (with IPA transcriptions)

No.	**X**		**a**		**Y**		**b**	
1	klepin	[klepɪn]	TEP	[tɛp]	poemer	[pumər]	LUT	[lʌt]
2	lotup	[lotʌp]	SOT	[sɔt]	kengel	[kɛŋəl]	JIK	[jɪk]
3	kiertan	[kirtɑn]	RAK	[rɑk]	domo	[domo]	TOEF	[tuf]
4	fapoeg	[fapux]			loga	[loxa]		
5	griefup	[xrifʌp]			gopem	[xopəm]		
6	malon	[malɔn]			naspu	[nɑspu]		
7	veirig	[vɛirɪx]			vami	[vɑmi]		
8	dufo	[dyfo]			snigger	[snɪxər]		
9	tarzin	[tɑrzɪn]			rogges	[rɔxəs]		
10	seibor	[sɛibɔr]			densim	[dɛnsɪm]		
11	nijfoe	[nɛifu]			fidang	[fidɑŋ]		
12	baduk	[badʌk]			rajee	[raje]		
13	floenie	[fluni]			nilbo	[nɪlbo]		
14	tipla	[tɪpla]			plizet	[plizɛt]		
15	stepoer	[stepur]			banip	[banɪp]		
16	blieker	[blikər]			movig	[movɪx]		
17	muiblo	[mœyblo]			sulep	[sylɛp]		
18	kijbog	[kɛibɔx]			wiffel	[wɪfəl]		

Dependencies

LANG1	**LANG2**
TEP_LUT	**TEP_JIK**
SOT_JIK	**SOT_TOEF**
RAK_TOEF	**RAK_LUT**

Note: YX pairing were formed by combining each Y with the X on the same line, and with each of the 5 following Xs. YX pairings in the test phase were **klepin _ densim** and **lotup _ nilbo.**