Utrecht Institute of Linguistics OTS
Utrecht University

The Role of Positional Cues in Non-Adjacent Dependency-Learning

An Artificial Grammar Study

Ileana GRAMA
3429024

Master's Thesis

Linguistics: The Study of the Language Faculty

SUPERVISORS: PROF. DR. F.N.K. WIJNEN, DR. A.O. KERKHOFF
SECOND READER: PROF. DR. P.H.A COOPMANS

AUGUST 2011

ACKNOWLEDGEMENTS

Before I give my thanks to the many people who have made this piece of writing possible, I would like to point out one thing. The paper you are about to read is, for me, something of a triumph over my murky and sinuous thought processes. As I look back on the pages of this thesis, it seems so uncharacteristically comprehensible and *not* labyrinthic that I can hardly recognize my own brainchild.

Which is why my first big ‘thank you’ goes to Frank Wijnen: 1. Because for a year and a half now he has been helping me through every paragraph, midwifing my every underdeveloped idea, and correcting my every ill-spun sentence. 2. For giving me confidence in my work, freedom to pursue my research interests, and patiently helping me clarify them, even when they seemed vague and uncomely. 3. Most importantly, because throughout this collaboration I have felt a change in the way I think, argue my points, and define my concepts, which I can only attribute to his supervision.

I’m equally grateful to Annemarie Kerkhoff, without whom none of these experiments could have been carried out in due time and under acceptable conditions. She is also the main reason why this thesis is at all coherent. Her patience with my writing has been hugely helpful, her advice and help with the hands-on research – crucial; this thesis rests on her own previous research, as my (hopefully) future research rests on her and Frank’s input. I hope to have both of them there to guide it.

Thanks to Jacques Mehler and Marina Nespor for discussions that have shaped the thoughts behind this thesis and behind my PhD proposal; my stay in Trieste has helped clarify a lot of my questions, and given me a new perspective on my object of study. Thanks to David Maximiliano Gomez, Silvia Benavides Varela, Amanda Saksida, Alan Langus, and Damian Blasi for fruitful discussions, and a great time on Italian soil; also thanks to Martin Everaert whose generosity and faith in my research made it all possible.

Thanks to Iris Mulders for her help with subject-recruit, to Theo Veenker for technical support, to Desiree Capel and everyone at baby-lab meetings for advice and input, especially Lorijn Zaadnoordijk who gave me a preview of how baby-testing might go (gulp!). For valuable input I also thank René Kager, Serghey Avrutin, Sophia Katrenko, and Peter Coopmans. Thanks to all my subjects, especially my dear Dutch litkids, who spent a boring 20min inside a stuffy booth coloring mandalas – to quote a favorite author of mine, ‘scientific progress goes *boink!*’.

I am indebted to Peter Coopmans, my second reader, not just for availing himself of this role so wonderfully and on such short notice, but also for agreeing to supervise potential further research in this direction.

I owe a special debt of gratitude to Eddy Ruys – for hiring a theoretical linguistics renegade as his TA, for letting me do what I have dreamt of doing since kindergarten (teach), and for a recommendation letter which still makes me blush violently.

I dedicate this thesis to my family and friends, in Romania, the Netherlands, and all over the place – they know who they are and they know themselves to be awesome.

Table of Contents

Abstract: Previous research has shown that both adults and infants are capable of detecting and implicitly learning dependencies between non-adjacent elements in streams of spoken input. The learning mechanism underlying this ability, arguably a form of statistical learning, may be essential to infants' acquisition of morpho-syntactic dependencies in their native language. In this study we tested the limitations of this learning mechanism, using an artificial grammar that replicates some of the properties of morpho-syntactic dependencies in natural languages. Endress & Mehler (2009) proposed that non-adjacent dependencies are just as successfully tracked when they are instantiated at string edges, as when they occur string-internally (i.e. the dependency between a and b was shown to be learned just as easily in strings of the type XaYbZ as in strings of the type aXYZb). We predicted a replication of their results with our own materials, designed to emulate dependencies in natural languages. Subjects exposed to an artificial grammar with the structure YaXb (with a, b, monosyllabic, dependent elements, and Y, X, bisyllabic 'fillers') remained sensitive to the dependency despite the fact that the left edge of the dependency did not coincide with the left edge of the string, confirming our expectations. On the other hand, subjects exposed to a minimally different grammar, with the structure aYXb were no longer able to track the dependencies, suggesting that the length and nature of the intervening material plays an important part in the ability to keep track of non-adjacent dependencies.

1. Introduction

Recent research in the field of psycholinguistics has focused on demonstrating that human beings possess the capacity to perform abstract statistical computations on strings of spoken input and to extract certain regularities from them. Distributional cues at various levels of computation could play a crucial role in the acquisition of a child's native language: co-occurrence probabilities between syllables can help segment words from continuous speech (Saffran et al.,1996; Aslin et al., 1998) and the frequent co-occurrence of items could bootstrap categorization of the words occurring in between them (Mintz, 2003). In this paper we direct our attention to a specific instance where sensitivity to distributional properties of the input can support acquisition: the detection of dependencies between non-adjacent elements (Gomez, 2002, Peña et al, 2002, Newport & Aslin, 2004, Onnis et al., 2005, among others), whereby the presence of one token predicts the occurrence of a second, distinct token over a space of intervening material. We begin by a description of non-adjacent dependencies (NADs) at various linguistic levels, and discuss some of their properties.

1.1 NADs in Natural Languages

NADs in natural languages are groups of elements that co-occur frequently, either for being part of the same lexical root, the same paradigm, or instantiating some other formal relationship. Dependencies between non-adjacent elements are instantiated at several levels of representation.

At a morpho-phonological level, for instance, Semitic languages offer the example of consonantal roots. Arabic verbs are represented as triconsonantal verb roots, such as *ktb* for ‘write’, or *dhrj* for ‘roll’ (cf. McCarthy, 1999). These consonantal frames are then ‘filled’ with vowels, which themselves form dedicated patterns specific for the tense and voice of the verb. For instance, *katab* is the active, perfective form of ‘write’, *kutib* is its passive counterpart, and *aktub* the active form of the imperfective; accordingly, *dahraj* is the active perfective of ‘roll’, *duhrij* its passive version. Note that inserting vowels in these consonantal ‘grids’ does not alter the core meaning of the verb. Learners of Arabic, therefore, must keep track of the consonantal dependencies across vowels to learn the lexical entries of the verbs, and of the vocalic patterns accross consonants to learn the grammatical form of different tense/aspect paradigms.

The ability to keep track of NADs can also serve to detect morphological rules, like parasynthesis (Endress & Bonatti, 2007): parasynthesis triggers recategorization of a lexical root by simultaneously attaching both a prefix and a suffix to it: in Italian, the adjective *rosso* (red) becomes the adjective *arossire* (to redden), with neither **rossire* nor **arosso* being attested forms in the language. At a morpho-syntactic level, dependencies are instantiated between functional words or affixes, and serve to mark a formal relationship. Agreement, for instance, is often the dependency between inflectional morphemes and can mark the relationship between subject and verb (1a), subject and predicate nominal (1b), or determiner and noun (1c):

(1) a. **Nous** part**ons** demain. (French)
We leave.**1st pl**. tomorrow
b. Fete**le** sunt student**e** la medicina. (Romanian)
Girls-the.**fem.pl** are students.**fem.pl** of medicine
c. l**a** bambin**a** / **il** bambin**o** (Italian)
the.**fem** child.**fem** / the.**masc** child.**masc**

In languages like English or Dutch, auxiliary verbs are accompanied by specific suffixes on the verb stem:

(2) a. Vandaag **heb** ik de dokter **ge**beld. (Dutch)
Today **have** I he doctor called.**PART**

b. The train **is** leav**ing** in an hour.

In this paper we focus on the latter type of dependencies: we investigate the mechanism that allows infants to acquire morpho-syntactic patterns over non-adjacent functional elements as illustrated in (1) and (2). In the following section we present evidence that infants acquire morpho-syntactic dependencies at an early age.

1.2 Early Acquisition of NADs

An important problem in regard of the acquisition of morpho-syntactic dependencies is the question of functional elements and their absence from early spontaneous utterances (Guasti, 2002). Up to the age of about 3 years, children's speech is described as 'telegraphic', lacking many of the functional elements that are compulsory in adult grammatical speech. Two scenarios may account for this absence of functors from early speech: either infants do not attend to functional elements in the early stages of their development, due to the latter's reduced referential content and prosodic salience, or they do learn them quite early on, but are unable to produce them due to age-specific *production* limitations (cf. Gerken et al., 1990).

A number of studies propose that infants do, in fact, become sensitive to functional elements as early as 10-11 months, as a class distinct from that of lexical elements: Shi et al. (1999) showed that newborn infants were able to discriminate between lists of lexical and functional words in English; Shi et al. (2006) showed that 11-month-olds were aided by a real (but not nonce) high-frequency determiner in segmenting pseudo-nouns, and suggested that 8-month-olds had also acquired the determiner but with an underspecified phonological representation. Shafer et al. (1998) showed in an ERP study that 11-month-olds, but not 10-month-olds could discriminate between real and nonsense functors in a string of natural speech. Hochmann et al. (2010) suggested that 17-month-old infants are sensitive to the relative frequency of functors compared to the smaller frequency of lexical elements and identify the latter as a class of items that are more likely to be assigned referents in the visual world.

According to this line of research, infants exploit the properties of functional items to acquire them early on, as a set distinct from lexical words. If infants can track functors in speech input from such an early age, the next question is whether they are also sensitive to dependencies between them. Santelmann & Jusczyk (1998) showed that infants of 18 months were able to discriminate grammatical English sentences of the type: *At the bakery, everybody* ***is*** *bak**ing** bread*, from the minimally different but ungrammatical: **At the bakery, everybody* ***can*** *bak**ing** bread*. This sensitivity was not observed with younger infants of 15 months, and was also shown to be dependent on the length of the lexical material that intervened between the dependent morphemes: children could no longer detect the correct dependencies when more than 3 syllables intervened.

These results are supported by Gomez & Maye (2005), who tested infants' ability to detect and learn NADs in an artificial grammar: infants of 12, 15 and 18 months were exposed to an artificial language composed of strings of the form aXb, where 'a' and 'b' were monosyllabic nonce words who predicted each other with 100% probability (e.g. *pel* in string-initial position always co-occurred with *rud* in string-final position), and X belonged to a set of 12, 18 or 24 different bisyllabic elements (e.g. *kicey*, *puser*, *vamey, etc.*) which were combined exhaustively with the 2 a_b pairs. The aXb strings thus obtained were concatenated with 750ms pauses between them. Infants were familiarized with this artificial language for about 3 minutes, and were subsequently tested, using the head-turn preference procedure, on their discrimination between familiar aXb strings, and unfamiliar strings of the type aXb', where the b' element was not the one predicted by the a element. While 18- and 15-month-olds showed significant discrimination of trained and untrained dependencies, 12-month-olds did not, suggesting that computationally, the ability to track non-adjacent co-occurrence patterns could emerge around the age of 15 months.

Tincoff et al. (2000) also investigated a developmental aspect of NAD-learning: employing the same methodology as Santelmann & Jusczyk (1998), they tested sensitivity to different forms of the aspectual paradigm *be_ing* with 18-month-old infants. While these could distinguish between *was_ing* and **could_ing* (as well as between *is_ing* and **can_ing*, according to the 1998 study), they did not seem to have acquired dependencies like *are_ing* and *were_ing*. The authors tried to relate these findings to the frequency of the dependency in a corpus of child-directed speech: although *is_ing* was the most frequent occurrence, *was_ing* was in fact a lot less frequent than the present plural form *are_ing*. Thus, simple frequency of co-occurrence does not offer a full explanation for the learning pattern of NADs discovered by Tincoff et al.

Nevertheless, other distributional properties of morpho-syntactic dependencies may be relevant to the order of their acquisition. Heugten & Johnson (2010) studied the acquisition by Dutch infants of dependencies within the DP domain: on the one hand, the dependency between the determiner *het* and the diminutive suffix *–je* (e.g.*het hondje* 'the little dog'), and on the other the dependency between the determiner *de* and the plural suffix *–en* (e.g. *de paarden* 'the horses'). Dutch 24-month-olds, but not 17-month-olds, were sensitive to the diminutive dependency (instantiated over CVC nonce words, e.g. *het **kag**je*), while 24-month-olds had not yet acquired sensitivity to the plural dependency (e.g. *de **kag**en*). A corpus study showed that the two dependencies differed on a number of measures, including absolute frequency in the input, forward and backward transitional probabilities between the dependent elements, and number of syllables intervening between the dependent elements. All these measures predicted that the diminutive dependency would be acquired easier than the plural one, a fact confirmed by the infant study.

In conclusion, infants around the age of 18 months keep track of the functional elements in their native language and the formal dependencies between them. This sensitivity is modulated by various distributional properties of these dependencies (suggesting that a form of distributional learning is at stake), although other factors may come into play: functional elements, as we have seen, may be particularly salient due to their prosodic properties (Selkirk, 1996), or their contrast in length with lexical elements, making dependencies between them potentially easier to track; working memory limitations may constrain NAD-learning to only those instances where the dependent elements are sufficiently 'close'[1] to each other (cf. Santelmann & Jusczyk, 1998).

1.3 Issues in the Acquisition of NADs

As we have seen so far, dependencies between non-adjacent elements abound in natural languages at every level of representation. Furthermore, some of these dependencies are acquired and recognized by infants quite early on. A relevant question, therefore, is what mechanism might subserve the acquisition of NADs.

In what follows I present evidence from the literature that there is a distributional learning mechanism that allows learners to keep track of co-occurrence patterns between non-adjacent elements (otherwise known as higher-order dependencies). The following section presents relevant literature investigating the nature of the computations that allow NAD-learning, as well as their limitations, and the additional cues necessary to support these computations.

This study focuses on one particular cue, the salience of string edges, and investigates its relevance to NAD-learning. Part 3 discusses proposals on the salience of edges with respect to different types of computations, and part 4 presents an adult study conducted with the purpose of investigating the relevance of edge-salience to NAD-learning.

2. NADs in the literature

[1] Although Santelmann & Jusczyk define this 'closeness' in terms of syllabic units that separate the dependent elements, our discussion in section 4.3.2 will enumerate several ways in which we can construe the inability to keep track of dependencies across lengthier intervening material

2.1 NADs and Statistical Learning

The capacity to detect dependencies between non-adjacent elements has been studied in the literature as an instance of statistical learning. One proposal is that humans can keep track of transitional probabilities (TPs[2]) not only between adjacent elements (Saffran et al. 1996, Aslin et al. 1998, etc.), but also between non-adjacent ones (Newport & Aslin 2004, Endress & Bonatti 2007, etc.). That is, in a string of the type ABCD, listeners will not only be sensitive to TPs between A and B, or C and D, but also to the probability that C follows A, or that D follows B. In this scenario, NADs are detected by keeping track of the degree to which the first element in a dependency predicts the occurrence of the last one.

Newport & Aslin (2004) investigated adults' sensitivity to transitional probabilities between non-adjacent elements in a continuous string.. They claimed that if statistical learning is capable of keeping track of non-adjacent as well as adjacent regularities, then there must be a constraint on the type of non-adjacent TPs it can compute; otherwise, the ability to keep track of the co-occurrence statistics of any two items in a string would yield an explosion of possible computations, proving to be a costly learning mechanism. In order to avoid this computational explosion, one could assume that there are certain constraints limiting the operations of this statistical learning mechanism, making it apply selectively. Such constraints on a learning mechanism would most likely correlate with constraints on the type of patterns that are available in natural languages, since these patterns must be accessible to implicit learning. In a series of behavioral studies, the authors attempted to identify some of these constraints, employing continuous strings of syllables to test subjects' ability to extract various types of non-adjacent regularities: dependencies between syllables, consonants, or vowels.

In Newport & Aslin, (2004), subjects were exposed to 21 minute-recordings of synthesized continuous strings, composed of 5 syllabic 'frames' (pairs of syllables which co-occurred with a TP=1 across another intervening syllable) and 4 'fillers' (syllables that could occur in between any of the five frames). The TPs between adjacent syllables were, thus, .20-.25; if statistical computations could be performed between non-adjacent elements as well, the 5 frames should be statistically salient, in comparison to the low TP values of adjacent syllables. However, when faced with the task of discriminating between 'words' (aiXbi string from the exposure phase, where TP(bi|ai)=1) and 'part-words' (novel aiXbj strings, where TP(bj|ai)=0), subjects did not exceed chance performance.

[2] TPs express the probability, given the occurrence of a certain element X, that a certain element Y will follow in the input – in other words, the degree to which the occurrence of X predicts the occurrence of Y; the (forward) TP between X and Y in an XY sequence is calculated as the frequency of the XY pair divided by the frequency of X.

Contrastingly, subjects performed significantly above chance when the task was to track consonantal or vocalic frames composed of three elements (above 75% correct responses); they discriminated between words that respected the frames, and part-words that had also occurred in the exposure string, but that crossed a word-boundary (either the last two syllables of one word and the first of another, or the last syllable of one word and first two syllables of the following), even though statistic computations at the syllabic level would have provided no cues as to the valid words in the string.

The authors contended that the same type of statistical computation which succeeds in detecting regularities between phonemes fails when these regularities are instantiated between syllabic items. Several possible explanations were presented: on the one hand, non-adjacent patterns may be easier to detect when the items that instantiate them are interspersed with items that differ in some marked way. In other words, consonantal patterns may be easier to track across vowels and vice-versa precisely because the relevant items stand out due to their phonetic properties; at the same time, non-adjacent syllabic patterns were instantiated across syllables of the same kind. Alternatively, vowels and consonants may constitute several levels of representation (‘tiers’), as proposed by autosegmental phonology (cf. Goldsmith, 1999). It may be, therefore, that subjects were actually performing probabilistic computations on adjacent units at a given level of representation.

An alternative explanation is that statistical computations do not operate uniformly across linguistic levels (segmental, syllabic, etc.), but rather that certain computational constraints induce a preference for segmental rather than syllabic regularities. In other words, subjects may be most sensitive to regularities involving phonemic units, and may primarily compute statistical information at a segmental level; syllable-level statistics, then, may be constructed from the latter, and may represent a more complex type of computation.

Such an explanation relies on the assumption that there was no other difference in complexity between the study investigating syllabic regularities and the two investigating segmental regularities. A potential confounding factor, however, may have been the number of items that formed a dependency. Words in the syllabic condition were formed of a NAD accross two syllables; words in the segmental conditions were defined by more complex dependencies, between three items. That is, to segment a trisyllabic sequence as a word, participants could rely on the high TP value of one pair in the syllable condition, and that of two pairs in the phoneme conditions (the high conditional probability between the first and second item, as well as between the second and the third item in a frame). This potential increase in statistical cues may have facilitated the acquisition of segmental dependencies.

Nonetheless, the results of Newport & Aslin (2004) do suggest that the mechanism in charge of detecting NADs is constrained by specific factors such as salience; with no other cues in the input, subjects were only able to keep track of certain types of dependencies, showing that there are indeed constraints on non-adjacent statistical computations. However, the contention that these constraints correlate with limitations on the types of dependencies possible in natural languages may not hold entirely. Although segmental dependencies appear more accessible, dependencies between syllabic elements are often present in natural languages, as we have pointed out in the introduction; furthermore, we have seen that infants as young as 18 months are sensitive to syllabic/morphemic dependencies, modulo their distributional properties (Santelmann & Jusczyk, 1998, Tincoff et al., 2000, Heugten & Johnson, 2010). In what follows we shall see that NADs can be computed over syllabic units as well, provided there are certain cues marking the relevant items over which computations must be performed.

2.2 Variability

Gomez (2002) showed that even dependencies between syllabic units can be reliably detected given the necessary statistical cues. Her contention was that dependencies in natural languages are instantiated between frequent, closed-class elements, surrounded by open-class items with a much higher degree of variability; the frequent occurrence of functors makes them stand out against the relative infrequency and variability of lexical items and enables learners to keep track of the higher-order dependencies between them. In other words, the contrast between variability and fixed structure renders the latter more easily detectable.

Soderstrom et al. (2009) suggested that heterogeneity of variance is a key component of statistical-based language acquisition: for instance, they argued that the representation of linear input strings as sequences of peaks and valleys of variability (in this case TPs) is what drives segmentation of units; on the other hand, if a certain type of statistical computation yields no heterogeneity (for instance, if there are no great differences between the transitional probabilities in a string of elements), this might suggest that this particular type of computation is unproductive. Extending this argument, it is possible that heterogeneity of frequency[3] (marking functional elements as peaks and lexical ones as valleys) can alert the learner as to the relevant units over which non-adjacent computations are to be performed. Such a hypothesis also solves the problem raised by Newport and Aslin about combinatorial explosion: listeners would not have to keep track of dependencies between any two units at

[3] Note that, in natural languages, and for the purpose of this discussion, the notions of frequency (of dependencies) and variability (of the intervening material) are interchangeable: the variability of the intervening (lexical) elements correlates with their relatively low frequency (as elements of a larger set are expected to be less frequent), whereas the low variability of the dependent (functional) elements (as dictated by the smaller size of the functional category compared to the lexical one) correlates with their relatively high frequency

every possible level, but only of dependencies of elements marked by the salience of their frequency (or, as we will see later, other types of salience factors).

To test this, Gomez (2002) created an artificial language composed of aXb strings where the number of X-elements varied among different conditions: for three different a_b dependencies, there were 2 Xs in one condition, 6, 12, or 24 Xs in each of the others. One factor that distinguished between dependencies and X elements was the length: dependent elements were monosyllabic, while intervening Xs where bisyllabic; the aXb strings were separated by 750ms pauses. A comparison of performance (discriminating familiar aiXbi strings from novel aiXbj ones) between the four conditions for adults showed a significant difference between the largest set size (24X) and the other conditions; although subjects showed discrimination of trained and untrained strings in other conditions as well, the high variability in the 24X condition significantly facilitated NAD-learning. Similarly, 18-month-old infants were tested on the same type of language, in 3 conditions: set size 3, 12 and 24, and performed significantly above chance only in the third condition, suggesting that high variability may be a pre-requisite to NAD-learning for infants. Although the study demonstrates the relevance of frequency as a cue to NAD-detection, it is unlikely that this was the only cue to facilitate the extraction of dependencies: beside the difference in syllable number that separated the dependent from the intervening elements, the segmentation cue that Gomez introduced (750 ms pauses between aXb strings, leaving the dependent elements in string-initial and string-final positions, respectively) may have rendered the task considerably easier. As we will see in section 3, edges of strings seem to be privileged compared to mid-string positions in terms of salience. Additionally, the stimuli for this experiment were not synthesized, but recorded by a real native speaker; we may assume that certain prosodic or intonational cues may have also played a role in NAD-detection.

The studies cited so far seem to point to the conclusion that non-adjacent TP computations may not be the only factor involved in NAD-learning: higher-order dependencies are difficult to keep track of, and learners seem to require additional cues (such as stability against a background of variability) to identify the relevant units over which statistical computations can be made. It is possible that non-statistical cues as well play a crucial role in NAD-learning: prosody, pauses, or word-length may help segment the input in ways that makes dependent elements more salient, thus facilitating sensitivity to conditional probabilities between them (we return to this hypothesis in section 3).

Furthermore, it may be that NADs are reliably extracted through a different computational mechanism than the one argued for so far. In Gomez (2002), what distinguishes NADs is not only the non-adjacent TP value of 1, but also the high frequency of co-occurrence of the relevant elements. It may be that relative frequency of co-occurrence (simply counting the number of times that two elements co-occur), instead of conditional probability (comparing the number of times only one of the elements occurs, to the number of times they co-occur), is what determines sensitivity to NADs. Frequently co-occurring elements need not have high conditional probabilities: within the *be_ing* verbal paradigm, for instance, the verb *be* may frequently co-occur with the progressive *–ing* suffix (e.g. *he* ***is*** *sing**ing***), while also frequently occurring in other syntactic contexts, as a copular (e.g. *he* ***is*** *a writer*) or lexical (e.g. *there* ***is*** *a God*) verb. The relative frequency of the *be_ing* configuration would be high, but the TP of the pair may be lower, as influenced by the polysemous nature of the verb. Thus, the distinction between co-occurrence frequency and transitional probability is not only important from a theoretical point of view (identifying the exact type of statistical computation that supports NAD-learning), but also from an empirical point of view (making specific predictions about which dependencies should be easier to detect and learn).

2.3 Prosodic Cues and Different Types of Learning

Peña et al. (2002) investigated the relevance of non-statistical cues to NAD-extraction. In a series of experiments, employing synthesized strings of monosyllabic items, they showed that (even very subtle, almost imperceptible) segmentation cues may be crucial to learning NADs as rule-like regularities (if a then b, after an intervening X). Subjects were first exposed to a continuous concatenation of aXb strings (Experiment A) where a, X and b were pooled from three distinct sets each containing three distinct syllables (so that adjacents TPs were 0.3 within aXb strings and 0.5 between strings, whereas the TP between a and b was always 1.0). In both conditions, in the test phase, they were able to distinguish aiXbi ('words') from bjaiX ('part-words') strings, with test items taken from the training phase. This showed that participants were able to segment strings based on non-adjacent TPs even with no other cue available. However, when the task was to generalize[4] the dependency to a new context, that is to discriminate aiX'bi ('rule-words', with X elements which had occurred in the familiarization, but never in between ai and bi) from part-words (as defined before), subjects did not succeed when the input string was continuous (Experiment B), but only when subtle 25ms pauses were inserted between the aiXbi segments (Experiment C).

Based on these findings, the authors proposed the existence of two distinct learning mechanisms: a statistical one, which facilitates segmentation of continuous strings on the

[4] Our operational definition of generalization is the ability to detect a pattern (e.g. dependency) when it is instantiated in a context not heard before; this ensures that subjects do not simply recall strings from the familiarization phase, or remember familiar chunks, but that they are aware that the rule they have learned is productive, applicable to novel contexts. Consequently, Peña et al. (2002) define *rule*-learning (as opposed to, e.g., statistical learning) precisely as the type of learning that produces inferences which can be generalized to novel contexts.

basis of non-adjacent TPs, and is productive even in the absence of the subtle segmentation cues; and a rule-learning mechanism, conditioned by the existence of segmentation cues, whereby dependencies are acquired as structural rules (if a then b, after an intervening X) that can be generalized to novel contexts. In addition, increased exposure to unsegmented strings resulted in a significant preference for part-words rather than rule-words (Experiment D), suggesting that increasing exposure does not enhance performance on rule-extraction, but rather increases sensitivity to the 'statistical contingencies contained in the stream' (Peña et al., 606)[5].

Several objections may be raised as to the methodology employed by Peña et al. (2002). For instance, the design of the test phase in this study is different from previously cited ones (Gomez, 2002 and Newport & Aslin, 2004), which directly investigate knowledge of valid vs. ungrammatical dependencies – Peña et al. test subjects on their ability to segment the input with the help of non-adjacent regularities. The nature of the task in Experiment C, for example, makes it unclear whether subjects really acquired NADs from the input – that is, whether they show knowledge that a certain token of type a corresponds to only one specific token of type b. An alternative explanation would be that subjects simply learned that an element from the set of a tokens had to occur string-initially, while an element from the set of b tokens had to occur string-finally: this would exclude the necessity of positing a direct correspondence between a and b tokens (a token-dependency), and reduce the task to positing a correspondence between a certain set (e.g. the set of a tokens) and a certain position in a string (e.g. the first). This positional encoding of sets would have been enough to prompt preference of rule-words over part-words[6], and may have been conditioned by the existence of subtle segmentation cues (which marked the beginning and ending of strings, and therefore helped define structural positions within them).

Perruchet et al. (2004) also challenged the findings of Peña et al. (2002) on this account: if the relevant measure of NAD-learning is participants' ability to segment the input according to those NADs, then adding 25ms pauses may prove no more than the fact that

[5] Although this is not explicitly stated in the article, what the authors could presumably mean is that the adjacent TPs for the syllables composing part-words are higher than for the rule-words; therefore, with increased exposure, subjects may become more sensitive to these adjacent TPs. It is unclear why, in this case, sensitivity for non-adjacent TPs (which would prompt rejection of part-words in favor of rule-words) should be less powerful than sensitivity for adjacent TPs. The authors offer no detailed account of the exact statistical computations at hand, nor do they address the issue of the interaction between adjacent and non-adjacent probabilistic computations (i.e. what happens when the two prompt distinct responses).

[6] Whereas part-words were of the form baX or Xba, rule-words were of the type abb, or aab – thus, while the former violated positional encoding of classes at string edges, the latter also violated positional encoding, string-medially (i.e. where elements from the X set should have occurred). Nonetheless, we will see, in section 3.1, that the rule that assigns arbitrary sets of tokens to positions applies more effectively at string edges rather than string-medially (cf.Endress & Mehler, 2010)

discrete segmentation cues aid segmentation. In one experiment, they showed that participants could recall more correct trisyllabic sequences with than without 25ms gaps, even when there was absolutely no dependency between ‘a’ and ‘b’ elements in aXb strings. Furthermore, the authors ran 14 independent simulations (using the materials from Peña et al.’s study) with PARSER (cf. Perruchet & Vinter, 1998), a computational model that performs statistical computations only between adjacent elements and is, therefore, unable to track NADs. The scores obtained from these simulations closely resembled the performance of human subjects in Peña et al.’s Experiment A, suggesting that no sensitivity to non-adjacent patterns was necessary in order to successfully discriminate words from part-words in this artificial language: an associative learning mechanism (relying on the frequency of tri-syllabic chunks) was sufficient to segment the relevant ‘words’ from the continuous input string, without computing co-occurrence statistics between non-adjacent elements.

Onnis et al. (2005) also challenged the design of the Peña et al. (2002) experiments, by showing that the phonotactic properties of the artificial grammar may have provided additional cues; the syllables employed as NADs and those employed as Xs differed in that while the former began with plosive consonants, the latter began with liquids or fricatives. The authors showed that participants lost the ability to discriminate words from part-words when this phonotactic cues was removed, but were still able to make this distinction when only the phonotactic cue remained, and the statistical cue was eliminated (there were no NADs, TPs between non-adjacent elements were reduced to .33). In addition, subjects showed a general preference for strings that began with a plosive, even when these were statistically part-words in the familiarization phase.

Finally, Onnis et al. (2004) showed that subjects were, in fact, able to discriminate rule-words from part-words even in the absence of segmentation cues in the training, in an experimental setup similar to Peña et al. (2002): a crucial modification, apart from controlling for phonotactic cues, was the introduction of higher variability in the X elements: subjects performed significantly above chance in the discrimination task when the X syllable was drawn from a set of 24 elements. Onnis et al. (2004) thus conclude that the high variability of X is a sufficient cue to allow subjects not only to learn NADs, but also to recognize them in novel contexts.

2.4 Conclusions

In this chapter we examined some of the studies that investigate the cognitive mechanism responsible for the acquisition of NADs. As discussed, a prominent view in the literature is the construal of NAD-learning as a type of higher-order statistical learning: NADs may be acquired by a statistical learning mechanism that keeps track not only of adjacent transitional probabilities, but also of TPs between non-adjacent elements. A necessary implication of this model is the existence of certain constraints on this statistical-learning mechanism: additional cues are required to mark the relevant elements over which computations are to be performed. Otherwise the mechanism would be faced with a combinatorial explosion of possible pairs between which TPs should be calculated.

Certain studies such as Peña et al. (2002) claim that the necessary cues are in the organization of the signal, such as the existence of subtle pauses segmenting relevant units; this implies that in natural languages (where it is assumed that pauses are not consistent, dependency-marking cues), prosodic information may play a crucial role in NAD-learning (Jacques Mehler, p.c.). Other proposals (cf. Gomez, 2002 and Onnis et al., 2004) suggest that distributional cues, such as the stability of the NADs against a background of variability, are sufficient to support learning of dependencies in the absence of other cues.

Importantly, studies like Peña et al. (2002) impose a dichotomy between rule-learning and statistical learning mechanisms, suggesting that statistical computations cannot lead to the acquisition of NADs as generalizable rules, but that a distinct, algebraic rule-learning mechanism is responsible for this. This is crucial, as we expect morpho-syntactic dependencies in natural languages to be extracted as grammatical rules. Nonetheless, the belief that statistical learning cannot yield generalizable rules is not unanimously held in the literature: McClelland & Plaut (1999), as well as Seidenberg et al. (2002) question the assumption that ‘grammatical’ and ‘statistical’ learning are distinct processes: they cite studies that show neural networks being capable of reaching a number of generalizations, on the basis of statistical computations alone. Abstract rules have also been claimed to be inferred on the basis of analogies and similarities between frequent chunks encountered in the input (cf. Cleeremans et al., 1998, and references cited therein). The idea that generalizations cannot be derived from statistical/associationist computations is therefore subject to debate, and in our own study we choose not to take a definite position in this debate. The crucial matter that we take from the Peña et al. study is the proposal that generalizations about NADs cannot be obtained in the absence of segmentation cues; this proposal is contested in Onnis et al. (2004), which shows that under the right conditions of variability, NADs can extracted as generic rules even from continuous strings of spoken input.

In what follows we investigate the importance of another cue, namely positional information, which may be relevant to the detection of non-adjacent regularities. We examine the possibility that the position of dependent elements in a string (peripheral vs. string-internal) plays a role in the ease with which NADs are detected and learned. We begin by discussing claims to this effect that have been made in the literature so far, and present the results of our own experiments investigating the relevance of positional cues.

3. Edges

As we have seen in the previous section, the NAD-learning mechanism that we are investigating may rely on the existence of certain cues in the input: the study by Peña et al. (2002) seems to suggest that subtle pauses segmenting the input strings are necessary in order to acquire NADs as rules generalizable to novel contexts; other studies, like Onnis et al. (2004), propose variability of the intervening material as an alternative cue that can bootstrap NAD-learning. All of these studies, however, concur that NADs are only acquired in the presence of specific cues, be they distributional or segmentational in nature.

The 25ms pauses inserted between aXb strings in the experiments conducted by Peña et al. can be interpreted as marking the dependent a and b elements as edges of the string. This positional configuration itself may enhance the salience of the dependency (cf. Endress et al., 2005, Endress & Mehler, 2009 and Endress & Mehler, 2010). On the other hand, it is unclear to what extent such cues are present in natural languages: morpho-syntactic dependencies may not be consistently delimited by pauses, as initial and final elements of ready-segmented strings. In other words, dependencies in natural languages seem to be embedded in lengthier strings (e.g. *My mother* ***is*** *not feel****ing*** *well*), rather than appear as isolated AXB-type structures (e.g. ****Is*** *not walk****ing***). It is crucial, therefore, to investigate the importance of edge-salience to the acquisition of NADs. In this section we discuss a few important studies that deal with the role of edges in rule-extraction tasks, and the potential implications for our own object of study.

3.1 Salience and Perceptual Primitives

Endress et al. (2005) address the salience of edges in the context of constraints on the human capacity for algebraic[7] rule-computation. They propose that instead of being a general algebraic computational mechanism, the human capacity for rule-extraction is constrained by certain factors of 'perceptual salience' in the input: in other words, algebraic rules may be computed efficiently only in specific configurations, where the elements they refer to are somehow perceptually marked. Repetitions of tokens, for instance, may represent such a 'perceptual primitive' which facilitates rule-extraction: infants as young as 7 months, habituated with a sequence of trisyllabic strings containing a repetition (AAB), subsequently

[7] Algebraic rules are defined as rules that are computed over abstract symbols (variables that stand for all elements in a certain class, or having a certain feature), instead of tokens. Note that this is slightly different from Peña et al.'s definition of rules as patterns which can be generalized to novel contexts (but may also be patterns defined over tokens, like NADs): the former is a restricted version of the latter

preferred repetition (AAB) to non-repetition (ABA) strings even when though they were formed with items which had not been played during the habituation (Marcus et al., 1999); infants were, therefore, able to learn the repetitions as abstract rules that could be applied to novel stimuli[8]. Similarly, Endress et al. (2007) showed that adult subjects can generalize repetition-based rules over sequences of three piano-tones, but fail to abstract general rules from similar sequences which do not contain repetitions. It appears that abstract patterns are more efficiently detectable when repetitions are involved.

Another example of a 'perceptual primitive' is the salience of edges: rules that involve edges of strings are more easily acquired than rules that operate string-internally. To demonstrate this, Endress et al. (2005) tested subjects' ability to generalize repetition rules that applied either to the edge (ABCDEFF) or to the middle (ABCDDEF) of a string. Subjects in two experiments were exposed to 36 7-syllable strings containing a repetition either string-finally (e.g. /zOfesapitu*koko*/, for ABCDEFF) or string-internally (e.g. /zOfesa*pipi*tuko/, for ABCDDEF). As each sequence was presented, they were asked to memorize it, and subsequently discriminate between an identical sequence, and a minimally different one where the repetition was instantiated at a different position in the string. Participants performed equally well at this discrimination task, in both experiments. In the test phase they were given novel sequences and asked if these conformed to the patterns they had heard in the familiarization, in a two forced-choice task where they had to choose between grammatical and ungrammatical strings. In the first (string-final repetition) experiment these test strings were either grammatical ABCDEFF patterns, or ungrammatical ABCDEEF patterns; in the second, they were either grammatical ABCDDEF patterns, or ungrammatical ABCDEFF. Subjects performed above chance with the edge-final generalization, but failed to generalize the rule that applied string-internally.

Subsequent testing showed that this was not due to a general difficulty with encoding the positions of sequence-internal items, or repetitions: subjects were given pairs of strings of two types, and asked to judge whether the strings were identical or not: pairs of type 1 were made up of identical segments taken from the previous experiments (ABCDEFF vs. ABCDEFF, or ABCDDEF vs. ABCDDEF, respectively); pairs of type 2 were made up of two minimally different segments, both containing repetitions as well, but where the crucial difference was the *position* at which the repetition was instantiated (ABCDEFF vs.

---

[8] Recall that an algebraic rule is defined as any rule that operates over abstract symbols or classes as opposed to individual item, and which, once inferred from a limited amount of exposure, can be applied to novel contexts/stimuli. Throughout the body of research that is discussed in this section, the acquisition of these rules is maintained to be disjunct from any type of statistical/association-based learning, which deals with probabilistic computations over individual items. The possibility that statistical computations may be at the *basis* of abstract rule-inference, or involved in this process in any way, is not discussed. Perceptual primitives apply only to algebraic rule-learning, and are irrelevant to statistical learning, as we see in section 3.2

ABCDEEF, or ABCDDEF vs. ABCDEFF, respectively). The rate at which they correctly labeled identical (type 1) and non-identical (type 2) pairs was above 90%, irrespective of where the repetitions were located in the string, showing that there was no perceptual impediment in detecting the positions of the repetitions.

The authors concluded that string edges are marked positions for rules that rely on positional encoding. The task in each experiment required not only detecting repetitions but also encoding the precise positions where these repetitions were instantiated; thus, the algebraic rules obtained needed to specify both the existence of a repetition and its position in the string. Because it was shown above that the difficulty in detecting the position of string-internal repetitions was not perceptual in nature, it follows that the problem arose in the actual computation of the algebraic rules. The conclusions of this study are that rules that compute regularities over specific positions in a string are not learned by a general abstract learning mechanism, but by a more restricted, 'piecemeal operator'. This specialized mechanism is sensitive to the edges of individual strings in the input, but is unable to compute rules over string-internal positions.

It is important to note, however, that although participants in these experiments failed to generalize string-internal repetition rules, whereas they succeeded with string-peripheral ones, this does not mean that there is no mechanism strong enough to compute sequence-internal rules, merely that such a mechanism failed to produce the relevant generalizations in *these* specific experimental conditions. While there is evidence to support the existence of 'perceptual primitives', as mechanisms that facilitate the detection of a certain type of rule, we do not fully support the conclusion that they represent restrictions on the type of rules that can be computed by human learners. Furthermore, this study investigated the salience of repetition rules in string-final position – it is quite possible that this 'edge-salience' applies specifically to repetition-detection, or that the right edge of a string is in fact more salient than the left one. It is important, therefore, to investigate the extent to which string-edges are salient for other types of algebraic computations as well.

Endress & Mehler (2010) investigate the acquisition of phonotactic constraints string-internally and string-peripherally. In their first experiment they employ strings of the form C1VCCVC2 (e.g. /*kalRis*/), where C1 and C2 are each selected from a distinct set of three consonants (C1 from {k, t, f} and C2 from {s, ʃ, p}), and the VCCV clusters are taken from a set of 6 strings({/alRi/, /aRli/, /alni/, /anli/, /aRni/, /anRi/}); the phonotactic constraints, therefore, consisted of recognizing that the first and last consonants were each constrained to an arbitrary set of possible options. Participants were exposed to 36 strings, obtained from 6 out of 9 C1-C2 pairs, combined exhaustively with all the 6 VCCV clusters. Grammatical test strings were obtained by combining the remaining 3 C1-C2 pairs with 6 VCCV clusters which were identical to the familiarization, except that the vowels were switched (e.g. /ilRa/, /iRla/, etc.). Ungrammatical test strings were identical with the grammatical ones, only the C1 and the C2 were inverted. Subjects were able to discriminate the grammatical strings, showing that they had learned the phonotactic constraint and could apply it to novel strings.

In an analogous second experiment with strings of the type CVC1C2VC, subjects did not perform above chance in the discrimination task; however, when the arbitrary sets for C1 and C2 were replaced with natural classes (C1 was always a stop {k, t, p}, and C2 always a fricative {f, s, ʃ}), subjects acquired the phonotactic constraints even in string-medial positions. As before, the difficulty encountered by participants in the second experiment was not due to processing perceptual information in string-medial positions, but in extracting the general rules: additional testing showed that participants could encode consonants both at string edges and string-internally. Subjects were given test pairs consisting of grammatical strings from the test phases of one of the first two experiments, either paired with themselves or with their ungrammatical foils; they discriminated the identical from the non-identical pairs with above 95% accuracy.

Phonotactic regularities, therefore, are also more easily detected when they are at string edges; the authors conclude that 'positional generalizations' (algebraic rules that target certain specific positions in a string) are constrained by a perceptual primitive that makes encoding items at edges easier and/or more precise than in middle positions. On the other hand, the bias towards natural phonological classes seemed strong enough to allow phonotactic learning even in string-medial positions: preference for natural classes may be a different kind of perceptual primitive, which may outweigh the bias towards edge positions.

The studies reviewed so far are consistent with the hypothesis that certain features of the stimuli over which rules are computed facilitate the computation of the latter: generalizations that involve positional information are best computed at string-peripheral positions; generalizations that make use of natural classes are computed more efficiently than those that refer to arbitrary classes (in string-internal positions). The authors infer that this is indicative of the existence, instead of a generic mechanism for rule-detection, of specialized mechanisms that detect rules only under specific requirements, for instance, only string-peripherally.

These findings are highly relevant to the topic of our own investigation: as we have mentioned in the previous section, Peña et al. (2002) as well as Onnis et al. (2004) showed that NADs can be learned as rules, generalizable to unfamiliar contexts. As an instance of rule-learning, NAD-extraction is, therefore, predicted to be influenced by perceptual primitives such as edges salience. In the section that follows we address this prediction and the complications that arise.

3.2 NAD-Learning and the Salience of Edges

Endress & Bonatti (2007), addressed the relevance of edge-salience to the extraction of NADs by pointing out the cognitive differences between two processes: learning dependencies between classes and learning dependencies between tokens. The authors conducted a series of experiments replicating, to a large extent, the design in Peña et al. (2002). Employing the same familiarization material as in Peña et al. (2002), they tested discrimination of three types of test items as a function of duration of exposure: class-words (aiX'bj, X' taken from the set of a or b, and thus different from the X in the exposure phase), rule-words (aiX'bi) and part-words (aibjX or Xajbi)[9]. The authors found that while rule-words were preferred over class-words even better after longer exposure periods, the reverse pattern applied for the discrimination of class-words from part-words: if class-words were preferred after 2 minute exposure phases, this preference disappeared when the exposure was increased to 30 minutes, and a preference for part-words emerged after 60 minutes of exposure to the same materials.

The contrast between rule-words and class-words was taken be indicative of learning specific token-dependencies (NADs), whereas the preference for class-words over part-words was assumed to show that subjects have acquired the 'structural relations' between one class of items (ai) in string-initial position corresponds to another class (bj) in string-final position. The authors proposed that the distinct patterns of acquisition of class-words and rule-words stem from the fact that while NADs (rule-words) are acquired based on a purely associationist mechanism, class-rules are acquired by a mechanism dedicated to the computation of structural rules; the latter is assumed to lose its strength with greater exposure time, due to the increase in the memory representations of actual items in the input.

Endress & Mehler (2009) picked up on this distinction: in their study they compared the influence of positional cues on the acquisition of class-words and NADs, by observing the acquisition of both class and token dependencies instantiated string-peripherally (AiXYZCi), and string-internally (XAiYCiZ). Within the same experimental setup as before, class-words (AiXYZCj) were discriminated from part-words when they were instantiated at string edges, but performance was at chance-level when they occurred in the middle of the strings (XAiYCjZ). On the other hand, token dependencies were preferred over class ones when the relevant elements occurred string-medially (XAiYCiZ vs. XAiYCjZ), suggesting once more that the acquisition of NADs, unlike that of class rules, is not affected by positional cues.

[9] For instance, whereas the familiarization contained words like **be***fo***ga**, **ta***li***du**, or **pu***ra***ki**, part-words could be strings like *fo***gata** (Xba) or **kibe***fo* (baX), and class-words could be strings like **putaga** (aab), or **beduki**. (abb).

The results of these studies were offered as evidence in favor of a distinction between class-rule learning as an instance of algebraic computations, and NAD-learning as an instance of statistical learning. However, two important observations must be made. Firstly, we have pointed out in our discussion of Peña et al. (2002) that class-words may be learned on the basis of a mechanism that detects classes of items which can occur in specific positions within a string, therefore based on positional encoding; the same considerations apply in this case as well. Furthermore, in our discussion of Endress et al. (2005) we pointed out that edges are salient for positional encoding specifically. It is unsurprising, therefore, that class-rules were better detected at string edges.

Nonetheless, the main findings of Endress & Bonatti (2007) together with Endress & Mehler (2009) contradict our assumptions so far: studies discussed in the previous section suggested that rule-learning is subject to the influence of perceptual primitives – having shown that NADs can be learned as generalizable rules, we expected NAD-learning to be influenced by positional cues. At the same time, the experiments conducted by Endress & Mehler (2009) suggested that dependencies between tokens (NADs) could be tracked equally well both in string-peripheral and string-internal positions.

Several possibilities arise: on the one hand, perceptual primitives such as edge-salience have been shown to apply to rules that involve positional encoding; if NAD-rules are tracked independent of positional encoding, they may also not fall under the incidence of perceptual primitives. On the other hand, the studies cited above propose that the token-dependencies they investigate are not generalizable as (algebraic) rules; the crucial difference between these NADs, and NAD-rules as presented in Onnis et al. (2004), is that the latter are detectable as a results of the high variability of the intervening material.

The question that remains to be answered, therefore, is whether NAD-rules are indeed sensitive or not to the salience of edge positions. In this study we seek an answer to this question by conducting two similar studies: in the first, adult subjects will be tested on their sensitivity to a_b dependencies in strings with the structure YaXb; in the second, the same methodology will be employed, with the minimal difference that the structure of the strings will be aYXb. By comparing results, we aim to compare the efficiency of NAD-learning in contexts where NADs are not instantiated string-peripherally (Experiment 1), to the efficiency of detecting NADs at string edges (Experiment 2). If NAD rule-learning entails positional encoding and is sensitive to perceptual primitives, we predict that subjects in Experiment 1 will perform significantly worse than those exposed to the materials in Experiment 2. On the other hand, if NAD-learning does not rely on positional encoding, we expect participants in both experiments to perform equally well.

Our study is based on the methodology employed in Gomez (2002) and Onnis et al. (2004): the classes of Y and X elements are each made up of 18 different bisyllabic elements, stradding three monosyllabic a_b dependencies. Our design is intended to emulate the properties of morpho-syntactic dependencies in natural language, with the functors being shorter in length, more frequent and less variable than lexical items. In so doing, we hope to come closer to determining whether NAD rule-learning, as a mechanism potentially employed in language acquisition, is sensitive to positional salience.

4. The Salience of Edges in NAD-learning – An Artificial Grammar Study

This investigation takes as baselines two experiments conducted in the same laboratory, whose design and methods we adopted in our own study: the first (henceforth Experiment 01) replicated Gomez's (2002) findings in the high-variability condition of her Experiment 1, with Dutch adult participants; the stimuli were taken from the original study and adapted minimally in order to match the Dutch sound-pattern; three a_b dependencies were paired exhaustively with 18 X bisyllabic elements. Each string occurred 6 times in a familiarization phase that lasted approximately 15 minutes, during which subjects were given a mandala to color. Results showed a significant preference for trained over untrained strings in the test phase.

The second baseline for our study (Experiment 02) replicated the exact setup of Experiment 01, with the minor modification that the X elements in the test phase were novel bisyllabic nonce words that had not been heard in the familiarization at all. Subjects showed that they could generalize their knowledge of the NADs to novel contexts by endorsing grammatical over ungrammatical dependencies significantly more often.

In Experiment 1 we reproduced the design of Gomez (2002) and the first of our baseline studies, with the minor modification of introducing an additional bisyllabic Y-element at the left edge of every aXb string (yielding YaXb). The purpose was to verify whether subjects were still able to learn the a_b dependencies, despite the fact that one of the dependent elements was no longer marked as a string edge. The rationale behind employing the Gomez (2002) methodology was twofold: firstly, the study showed successful learning of NADs, and was furthermore successfully implemented in our own lab, for Dutch subjects; secondly, we considered it superior to other methodologies in testing precisely the mechanism of learning that could be employed for NAD-detection in natural languages.

In Experiment 2, we aimed to create a more adequate baseline for comparison: a minimally different grammar, that would match the previous in complexity, but crucially differ in that it instantiated the dependent elements at string edges (aYXb). A comparison between the two would reveal the effects of positional cues on NAD-learning: if subjects performed significantly better in the second experiment, this would suggest that the salience of string edges plays a role in the detection and acquisition of NADs.

4.1 Experiment 1

Subjects

34 participants (2 male), aged between 19 and 30, were recruited via email; subjects were required to have no hearing impairments or attention deficits, and to be native speakers of Dutch. The experiment lasted 20 minutes, and participants received a small (5 euro) reimbursement for their effort.

Materials

The familiarization phase consisted of continuous YaXb strings, concatenated with 250ms pauses between nonce words and 750ms pauses between individual strings. This grammar differed minimally from Gomez (2002) in that it introduced an additional bisyllabic Y element at the left periphery of the string.

The edges of the a_b dependencies were, therefore, no longer marked by the pauses separating the strings. Instead, subjects had to attend to the co-occurrence patterns between a string-internal and a string-final element, and to make subtle discriminations between grammatical and ungrammatical a_b pairs. If NADs are detected by a statistical learning mechanism which is only sensitive to frequencies and transitional probabilities, then the non-alignment of one of the dependent elements with the string edge should not affect subjects' ability to learn the dependencies.

Subjects were assigned to one of two different languages, L1 and L2; three a_b dependencies between monosyllabic elements were created for both L1 (a1_b1, a2_b2, a3_b3) and L2 (a1_b2, a2_c3, a3_c1), so that the dependencies grammatical 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. This was achieved by employing ordered lists of Y and X elements and combining each Yi with six Xs, from Xi to Xi+5 (the lists looped back on themselves, so that Y16, for instance, was combined with X16, X17, X18, X1, X2, and X3). Thus, the probability of co-occurrence (or the non-adjacent transitional probability, NATP) for a given YX pair was 0.16, and the frequency of occurrence for all Y and all X elements independently was equal (each appeared 6 times).

The pairs thus obtained (18x6=108) were combined exhaustively with the three dependencies in each language for a total of 324 strings per language (cf. Kerkhoff et al., 2011). Consequently, all Ya pairs had a forward transitional probability (fTP, the probability of a following Y, calculated as the number of occurrences of the pair Ya divided by the number of occurrences of Y) of 0.33, but a backward transitional probability (bTP, number of occurrences of Ya divided by the number of occurrences of a) of 0.055.

The test strings were created by selecting two novel YX 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 from combining each YX pair with all six dependencies in L1 and L2, resulting in 12 different strings; these strings were presented in a random order and participants were asked to judge whether they were consistent with what they had heard or not.

Stimuli

The 3 monosyllabic a and b nonce words, along with 24 bisyllabic Xs were taken from Kerkhoff et al. (in press); 24 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 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. *klépin tép poémer lút*), and with special emphasis placed on the monosyllabic dependent elements. The individual elements were subsequently spliced from the recording with 10ms of pause before, and 15ms after; only one token was used for each stimulus. 18 elements were chosen out of the 24 recorded for each of the X and Y sets, based on the quality of the recording (cf. Appendix 1).

The stimuli were then concatenated into strings as indicated above, with an additional 225ms pause in between each word, so that a total pause of 250ms was obtained between each nonce word in a string. 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’. 17 participants were assigned to each of two artificial languages, L1 and L2, respectively. The ‘sentences’ were presented over speakers in a sound-isolated 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.

In the test phase, subjects were presented with individual test strings and asked if they conformed to what they had heard in the training phase: they responded by pressing one of two buttons marked ‘Yes’ or ‘No’. They heard 12 strings of which only 6 were consistent with the subtle word-order patterns in the training, and 6 were not. For each string, they had to determine whether they thought it observed the patterns in the language or not. Subjects were told 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 participant, we calculated percentages of endorsement for Trained (‘yes’-answers to test strings which belonged to the training language) and Untrained (endorsement for test strings from the other language than that presented in the training phase). The mean percentage of endorsements for grammatical strings was 69.12% (SD=21.76), and the endorsement for ungrammatical strings was 60.29% (SD=17.89); t-tests showed that endorsement rates were significantly above 50%, both for grammatical (t (33)=5.122, p<.001) and ungrammatical (t(33)=3.355, p=.002) strings. Only one participant explicitly identified the pattern in the training strings, and none of the subjects scored above 83.3% correct.

We ran an RM ANOVA with Grammaticality (trained, untrained) as a within-subject factor, and Language (L1, L2) as a between-subject factor, and obtained a significant effect for Grammaticality: F(1, 32)=4.356, p=.045, $\eta 2$=.120, no significant Language effect (p=.859), and no interaction (p=.646).

The mean percentage of correct answers was 54.41%; a simple t-test revealed that this score was significantly above chance: t(33)=2.112, p=.042.

In order to identify the effect of edge salience on NAD-learning, we compared this experiment with the baseline Experiment 01 (25 participants). Results of this experiment indicated that subjects were reliably able to discriminate between grammatical (76.67% endorsement, SD=19.83) and ungrammatical (54.67% endorsement, SD=25.24) strings (F(1)=8.753, p=.007).

A RM ANOVA with Grammaticality (trained, untrained) as within-subject, and Experiment (Exp 01, Exp 1) as between-subject factor yielded a significant effect of

Grammaticality (*F*(1, 57)=14.790, *p*<.001), but no significant Grammaticality X Experiment interaction (*p*=.106).

The task in Experiment 01 was that of recognizing familiar strings, which had occurred 6 times in the familiarization; Experiment 02 (19 participants) tested subject's ability to recognize a_b dependencies in strings they had not heard previously (the endorsement for grammatical strings was 82.46% (SD=24.51), and for ungrammatical 52.63% (SD=40.54); a RM ANOVA with Grammaticality as within-subject factor showed a significant difference, *F*(1)=7.342, *p*=.014). Recall that our own test items were strings that had not occurred, as such, in the exposure. We therefore compared our results to those of this second study: an RM ANOVA with Grammaticality (trained, untrained) as within- and Experiment (Exp 02, Exp 1) as between-subjects factors yielded a significant Grammaticality X Experiment interaction was obtained: *F*(1, 75)=4.494, *p*=.039 (cf. Table 1). Subjects in Experiment 02 showed a significantly better discrimination of trained and untrained strings than subjects in Experiment 1, suggesting that the task in the former was easier than in the latter.

%Endorsement	Experiment 01	Experiment 02	Experiment 1
Grammatical	76.67% (SD 19.83)	82.46% (SD 24.51)	69.12% (SD 21.76)
Ungrammatical	54.67% (SD 25.24)	52.63% (SD 40.54)	60.29% (SD 17.29)

Table 1: Endorsement percentages for trained and untrained items, for Experiments 01, 02 and 1

Discussion

Subjects in this experiment showed a significant preference for grammatical versus ungrammatical dependencies, despite the fact that the left edge of the dependency no longer coincided with the left edge of the string. This experiment, therefore, not only replicates the results of Gomez (2002), but also demonstrates their reliability, by showing that the success of NAD-learning in the previous study did not rely primarily on the existence and nature of the segmentation cues.

Our results also replicate those of Endress & Mehler (2009), where subjects were also shown to discriminate ai_bi from ai_bj dependencies successfully even when the dependencies were embedded in longer strings. A crucial difference between the latter and our own study, however, rests in the variability of the material surrounding the dependencies. In this experiment, each non-dependent (X or Y) element occurred 6 times out of 324 (1.85%), while each dependency (a_b) occurred 108 times out of 324 (33.33%). In Endress & Mehler’s experiments, on the other hand, each of the dependent and non-dependent elements occurred equally frequently (8 times during a 3.45min exposure). Thus, in our study, not only were the dependencies salient due to the high (1.0) TP between elements a and b, but also due to their relatively high frequency, and the correlated variability of the surrounding material (the X or Y elements). As pointed out in Gomez (2002), this factor may be of great importance to NAD-learning, and may support learning of NADs as rules generalizable to novel contexts.

The aim of our investigation, therefore, was to verify whether NAD-learning as rule-learning was also effective in the absence of positional cues like edge-salience: that is, whether NADs were still detectable as generalizable rules in YaXb strings, and to what extent they were harder to learn than in grammars where they were marked as string edges.

One possible objection to the design of this experiment is the fact that only one of the dependent elements, namely the first, lost positional salience, whereas the right edge of the string was still aligned with the last item of the dependency. Adding a third bisyllabic Z to the right edge would have greatly increased not only the complexity of the language, but also the length of the training phase, thus proving an impractical solution. Given that one of the dependent elements was string-internal, we do not believe that subjects could have made the subtle distinction between grammatical and ungrammatical dependencies if they had only been attending to the edges of the strings. Nevertheless, a good suggestion for a future study would be to compare the results of the present experiment, with one where neither dependent element is aligned with the string edges.

In order to address our main research question about the relative effect of positional cues on NAD-learning, we compared the results of our experiment, with those of a previous study where participants were able to discriminate between aiXbi and aiXbj strings, the former of which they had heard in the familiarization phase. No significant difference in performance was found between Experiment 01 and Experiment 1. We also compared results with another experiment in which participants in the test phase were able to discriminate between aiX’bi and aiX’bj dependencies instantiated over novel X-elements. Performance in Experiment 02 was significantly better than in Experiment 1, as the difference between endorsements for grammatical and ungrammatical strings was higher in the baseline experiment (02) than in our own.

In order to interpret these results, we must understand which of the two baseline studies were comparable to our own: that is, whether the task in our own experiment prompted learning NADs as generalizable rules (similar to Experiment 02), or merely recognizing strings from the familiarization phase (similar to Experiment 01). If NADs were learned as rules, then the significant difference in performance from Experiment 02 would suggest that the insertion of the Y-element at the left periphery rendered the task of tracking NADs more difficult. On the other hand, if subjects in our experiment discriminated grammatical NADs on the basis of familiarity with chunks presented in the exposure, then comparison with Experiment 01 suggests that the insertion of Y at the left periphery did not have a significant effect, and thus that NAD-learning was uninfluenced by positional salience.

The test phase in Experiment 1 was designed to elicit grammaticality judgments on novel strings made up of non-novel elements: entire YaXb sequences were novel because the Y_X pairings had not occurred as such in the familiarization, and this was meant to ensure that sensitivity to NADs in the test phase showed the ability to generalize. Nevertheless, the individual elements Y, X, a and b were chosen from the same sets in the exposure as in the test phase, and although the two Y_X combinations in the test strings were new, the final aXb sequences were familiar, as each X element was straddled by each individual dependency in the training phase. It is possible, therefore, that subjects focused more on the second part of the strings, particularly the aXb sequences which they had heard in the exposure, and were able to recall aXb chunks from YaXb strings.

In addition to this uncertainty with regard to the nature of the task in our experiment, previous studies differ in another aspect as well: complexity. Our materials consisted of two sets of 18 bisyllabic filler items (instead of one set), and strings made up of four elements instead of three. Whereas, previously, subjects were exposed to 6 iterations of 54 different strings, our exposure consisted of 324 structurally different strings which were longer and more complex than the previous. Thus, the increase in complexity of Experiment 1, instead of the effect of edge-salience may have rendered performance in Experiment 02 significantly better than in this experiment.

In order to compare the results of this first investigation with a second experiment matched in all respects other than positional cues, we constructed a minimally different grammar aYXb; by employing the exact same materials, we hope to verify whether the string-peripheral or string-internal position of the first element of the dependency has an effect on NAD-learning. If subjects rely on edge salience as a cue, their performance in the second experiment should be significantly better than in the first, whereas if edge salience is not a factor, performance should be the same in both experiments.

4.2 Experiment 2

Subjects

34 participants (7 male), aged between 19 and 30, were recruited via email, with the same requirement as before; of these, two were excluded from the analysis, due to their

familiarity with previous studies on NAD-learning. All subjects were rewarded 5euro for participation. The duration of the experiment was the same as before.

Materials & Procedure

For this experiment, the same stimuli were employed as in the previous one, with the minimal modification of the structure of the strings: instead of strings of the type YaXb (e.g. *klepin tep poemer lut*), participants heard strings with the structure aYXb (e.g. *tep klepin poemer lut*). Strings were concatenated in the same way as before, with both the training and the test phase being identical in design. The procedure was the same, and none of the participants tested in this experiment had participated in the previous version. As before, an equal number of participants (17) was assigned to each language, respectively; however, both of the participants excluded were assigned to L1, leaving a total of 15 subjects for L1, and 17 for L2.

Results

As before, endorsement percentages were calculated for Trained and Untrained test strings. Subjects responded 'yes' to grammatical strings 69.79% of the time (SD=24.47), and to ungrammatical strings 56.25% of the time (SD=32.44). An RM ANOVA with Grammaticality (trained, untrained) as within-subject factor, and Language (L1, L2) as between-subjects factor yielded no significant effect of Grammaticality (p=.149), and no interaction (p=.547).

The mean percentage of correct responses was 57.29% (SD=24.75): a one-sample t-test revealed that this was not significantly different from chance (p=.106).

Although the percentages of endorsement for grammatical and ungrammatical strings were similar to those obtained in the previous experiment, their difference was far from significant – this may have been due to a greater amount of variance.

A potential cause for this variance is the fact that 7 of the 32 participants in this experiment (22.3%) recognized the existence of a dependency. All 7 obtained 100% correct scores, and most could reproduce all three dependencies; one other participant was aware of the existence of a dependency, but believed that it consisted of a vocalic rather than a syllabic pattern.

We considered these 7 participants explicit learners of the dependency. In order to evaluate performance on implicit learning alone, we excluded their data and recalculated the mean endorsement percentages: implicit learners supported Trained strings 61.33% of the time (SD=20.81), and Untrained strings, 70.67% of the time (SD=18.18). A paired-sample t-test yielded no significant difference between endorsement percentages (*p*=.085). Note, furthermore, that even if these results could be considered marginally significant, they would indicate a preference for ungrammatical, rather than grammatical, strings. A one-sample t-test showed that the mean percentage correct answers (46%, SD=13.02) was not significantly different from chance (*p*=.143). It seems, therefore, that the effect that made overall endorsement percentages similar to those in the previous experiment was carried by the explicit learners. Implicit learners showed no sensitivity to the NADs.

%Endorsement	Experiment 2 (all participants)	Experiment 2 (implicit learners)
Grammatical	69.79% (SD 24.47)	61.33% (SD 20.81)
Ungrammatical	56.25% (SD 32.44)	70.67% (SD 18.18)

Table 2: Discrimination of grammatical and ungrammatical strings in Experiment 2 for all subjects, compared to discrimination scores for implicit learners alone

Discussion

The purpose of this second experiment was to serve as a baseline for the previous one: both artificial languages were constructed using the same vocabulary, the same number of items and were intended to be matched in terms of complexity. By comparing learning performances on a YaXb grammar to those on an aYXb grammar, we aimed to draw certain conclusions about the relevance of positional salience to NAD-learning.

Unfortunately the results of the second experiment were not conclusive: our analysis points to the possibility that participants were not able to implicitly acquire the dependencies. At the same time participants in the previous experiment showed a significant preference for grammatical strings, suggesting that they had acquired some implicit sensitivity to the dependencies.

Our hypothesis, based on positional salience of string edges, predicted that performance in the second experiment would be just as good, if not significantly better than the one in the first. Given the apparent reversal of this pattern, we must assume that a different factor than edge salience also differentiated the design of the two grammars.

One possibility relates to the fact that the relative distance between the dependent elements in this final experiment was greater than in all previous studies: NADs in experiment 2 straddled two bisyllabic pseudo-words, instead of one. In a study by Santelmann & Juzczyk (1998), 18-month old infants displayed sensitivity to morpho-syntactic dependencies in their native language (discriminating them from ungrammatical dependencies), but only when the dependent elements were separated by 3 syllables or less. The authors concluded that a limitation on working memory capacities prevented infants from detecting dependencies across larger spaces of intervening material.

Although adults may differ greatly from infants in terms of their working memory, it may be that NADs are only detectable under certain conditions of proximity. It has been argued that listeners attending to longer strings of sounds break these down into smaller 'chunks', which are represented separately (cf. Perruchet & Pacton, 2006, and works cited therein). If NADs are instantiated over greater distances, and thus unable to be represented in the same memory chunk, patterns between dependent elements may be lost to the learner. It may be, therefore, that in experiment 2, listeners were unable to represent the ‘a’ and ‘b’ elements in the same memory chunks, and thus remained insensitive to the co-occurrence statistics between them.

A second potential explanation for chance performance on experiment 2 has to do with the issue of generalization. We have already pointed out our concern that in experiment 1 subjects were not showing their capacity to generalize knowledge of NADs to novel contexts: by attending preferentially to the final segments of individual strings, subjects may have simply recognized familiar aXb sequences from the training phase. On the other hand, subjects in the second experiment could not simply rely on familiarity with strings from the exposure: the material spanned by the a_b dependencies consisted of novel YX combinations, so in order to make the correct discriminations subjects had to attend selectively to the first and last elements of the strings. Given the comparative difficulty of generalizing learned patterns to novel contexts, the difference in performance between the two experiments may have resulted from precisely this difference, between recognizing familiar strings and generalizing a NAD-rule to novel contexts.

Although it is beyond the scope of this paper to investigate the cognitive difference between learning generalizable patterns vs. familiar sequences, it is worthy to note that no novel stimuli *per se* were employed in any of the two experiments. Although the sequences were new, all the individual stimuli were familiar from the training phase. In addition to this, the non-dependent, bisyllabic items in this grammar exhibited consistent behavior, by occurring only in certain positions, and allowing generalizations as to the classes of pseudo-words that could occur in a certain structural position (i.e. all the elements in the class Y occurred as second items in a string, and all elements in the X class occurred in third position). The importance of lexical-like elements being classifiable was pointed out in Hoehle et al. (2006), who showed that children fared better at identifying morpho-syntactic dependencies when the material contained in between them was identifiable as pertaining to the class of DPs. In both our experiments, therefore, the generalization of dependencies may have been facilitated by the fact that the bisyllabic fillers reliably formed distinctive classes.

4.3 General Discussion

Natural languages make use of various types of cues to mark regularities; infants presumably learn about the specific properties of lexical or functional elements in their native language by integrating these cues, by combining the output of various learning mechanisms operating at various levels of computation (cf. Soderstrom et al., 2009). Morpho-syntactic dependencies, for instance, are marked by several properties: distributional features such as high conditional probability or co-occurrence frequency, the specific phonetic and prosodic properties that mark functional elements in general, or the contrast in word-length between the functors that form dependencies and the lexical items that surround them. Investigating the importance of all these types of cues to learning can give a clearer picture of the process of acquisition: for instance, whether children are more attuned to statistical properties of stimuli, or whether prosody or other acoustic factors have a primary role in guiding the computations that can be performed over spoken input.

In this study we focused on the role of a specific type of cue: the relative salience of string edges in extracting generalizable rules. We asked what role edge-salience would play in the detection and learning of dependencies between non-adjacent elements. We began by looking at previous studies on NAD-learning, which identified a statistical learning mechanism capable of keeping track of patterns between non-adjacent syllables, and whose workings were modulated by the existence of certain constraints. At the same time, we discussed studies that looked into the relevance of positional cues to various types of learning: algebraic rule-learning was shown to be more effective when the rules referred to edge positions, rather than to string-internal ones. On the other hand, statistical learning was claimed to be unaffected by factors of positional salience. The predictions of these claims for our study would be that learning NADs as generalizable rules should entail a significant difference between -detecting a_b dependencies in a language of the type YaXb and in a language like aYXb: if NAD-learning involves positional encoding, subjects should find it easier to detect and acquire the dependencies in the latter rather than the former artificial grammar. Unfortunately, while experiment 1 showed that subjects were sensitive to the NADs in strings like YaXb, in the second experiment, subjects failed to extract the regularities in aYXb strings.

The conclusions that we may draw from this study are straightforward: firstly, we have shown that adult subjects can acquire NADs irrespective of the relative position of the dependent elements with respect to string edges, even with an artificial grammar more attuned to the properties of natural languages (i.e. where the dependent elements behave like natural language functors with respect to their contrast in frequency, low variability and word-length with adjacent, 'lexical' nonce words). Secondly, we have seen that the length or complexity of the material spanned by NADs may be a crucial factor in their acquisition.

4.3.1 Prosody and Edges

One of factors in the design of our experiments is worth discussing: the artificial language consisted of stimuli recorded by a native speaker, conforming to the Dutch stress-pattern and with special emphasis on the monosyllabic elements that formed the dependencies. Many of the subjects were particularly struck by the intonation of the strings, and some even became aware that the monosyllabic pseudo-words were important, due to the emphasis with which they were pronounced.

The prosodic cues in the strings that were used, therefore, may have been partly responsible for the acquisition of the dependencies: stress on the monosyllabic elements may have rendered them more salient. In natural languages, monosyllabic functional elements do not have fixed prosodic properties: Selkirk (1996) proposes that stress-assignment rules refer to lexical categories and their projections alone, concluding that a functor in a language like English, for example, can be assigned a variety of prosodic representations, depending on the context where it appears.

Peters and Stromqvist (1996) claim that this variation in prosodic status, contrasted with the stable prosodic pattern of lexical elements, may prove to be a useful tool in identifying functional elements in the input ('The Spotlight Hypothesis'). If the prosodic properties of functors are indeed a crucial factor in their acquisition, and thus in the acquisition of dependencies between them, then it is possible that the type of learning tested in our experiments is qualitatively different from the type of learning employed in detecting morpho-syntactic dependencies in natural languages.

An additional effect of the prosodic pattern of our materials is the possibility that it may have created 'edges' of its own. The work of Endress and colleagues examines the role of pause-marked string-edges in different types of computations; however, boundaries in spoken input are not marked by pauses alone. In our artificial grammar, for instance, the prosodic pattern SW S SW S (S – strong, i.e. stressed; W – weak, i.e. unstressed) may have marked a boundary between two stressed syllables (that is between the 'a' element, and the first syllable of an 'X' element), thus leaving the two dependent nonce words ('a' and 'b') as final edges of consecutive short strings.

To our knowledge, the work on edge-salience has not been extended to cover the effect of prosodically-marked boundaries on algebraic and statistical computations. Nonetheless, Endress & Mehler (2009) mention that in natural languages, string-edges may be marked in a variety of ways: prosodic 'break-points', affixation (inflectional morphemes may be marked as appearing on edges of lexical units or phrases), or any other device that segments continuous input into individual strings. It is to be expected, therefore, that boundaries marked through any linguistic mechanism can enhance the salience of the elements adjacent to these boundaries.

4.3.2 Distance and Complexity in NAD-learning

Another potential objection to our methodology was the length and complexity of the 'Y' and 'X' elements: in Experiment 2 we saw that NADs could not be detected over two bisyllabic intervening elements , even when the variability of this intervening material was increased to 108 distinct concatenations. Although this allowed us to conclude that there is a processing limitation to the detection of NADs, several other explanations are possible for the results of Experiment 2 (compared to, for instance, the above-chance performance in the baseline Experiments 01 and 02, and also in Experiment 1).

Firstly, the variability of the YX concatenations itself may have been too great: although Gomez (2002) demonstrates that there is a lower limit to the amount of variability that allowed NADs to be acquired, there is no reason to exclude the possibility of a higher limit as well. Subjects in Gomez's study heard each training string 6 times, over the entire course of the exposure phase, while in our exposure phase each string was heard only once. Gomez's high variability condition consisted of 24 intervening X items, whereas in our exposure a number of 36 different bisyllabic stimuli (2x18) were employed to create 108 different combinations. The information load, therefore, may have been simply too high for participants in our second experiment to successfully track the stable a_b dependencies. Hoehle et al. (2006) claim that a relevant factor in the detection of NADs in natural languages is what they call the 'analyzability' of the intervening material; their study suggests that NAD-detection relies on the infant having some knowledge of the nature of the element(s) spanned by the dependency – at the very least being able to categorize them. Gomez's materials may have allowed infants (and potentially adults as well) to learn the intervening X-elements as a category;, in our own design, although the bisyllabic elements consistently occurred in fixed positions (allowing them to be assigned to distinctive classes based on positional information), the number of such 'fillers' employed was greater than in Gomez's study: whereas Gomez used *one* class with 24 X items (maximum), we employed *two* classes, X and Y, with 18 items *each*. This higher number of nonce words (and the greater complexity introduced by the existence of two structural positions) may have rendered the intervening material harder to classify.

To verify this claim, one could restrict the number of possible YX combinations to 24, the same number of intervening strings as in Gomez's study, and test subjects' performance in a design more similar to Gomez (2002), where each training string was heard 6 times. Should performance improve, this would substantiate the claim that the nature and structure of the intervening material is relevant to the process of NAD-learning.

An alternative account for the failure to detect non-adjacent patterns across two bisyllabic fillers is the length of the span itself. Santelmann & Jusczyk (1998) suggest that infants cannot detect syntactic dependencies across more than 3 syllables; the authors attribute this failure to the developing working memory of 18-month infants. Our adult subjects, however, were also unable to distinguish non-adjacent patterns across a span of 4 syllables, suggesting that perhaps a working-memory limitation is not a sufficient explanation.

In the final part of their study, Santelmann and Jusczyk also point out that number of syllables may not be the correct measure for evaluating the distance between the dependent elements in terms of computational load. In the case of our own study, three possibilities come to mind: (i) NADs can be computed over a limited number of syllables, (ii) NADs can be computed over a limited number of separate morphological units, or (iii) NADs can only be computed within a limited time window. These three hypotheses correspond to three different ways of measuring the length of the intervening material.

A future project aims to evaluate each of these proposals in turn, and to give a clearer picture of the processing limitations that blocked learning of NADs in aYXb, but not in aXb. Experiment 2 will serve as a baseline for a number of studies which will modify each of the factors in (i), (ii) and (iii), while keeping the others constant, in order to identify the effect of each factor separately: for instance, to identify the effect of syllable number we will create a language with fewer intervening syllables (2), but with the same number of intervening units (2), and the same duration as Experiment 2 (a language of the type ayxb, where 'y' and 'x' are monosyllabic, and additional pauses are inserted between items in a string to match the mean duration of YX strings in Experiment 2). Understanding the nature of the limitations on the NAD-learning mechanism can offer fresh insight into the nature of this mechanism as a tool for language acquisition.

4.4 Conclusions

This project set out to investigate the role of positional cues in learning dependencies between non-adjacent elements. We have discussed the mechanism responsible for the detection and learning of NADs, and we have investigated the hypothesis that positional cues are relevant to the computation of NADs. Our results have confirmed previous findings that NADs can be computed string-peripherally as well as string-internally, and have further revealed the existence of processing limitations on NAD-learning. These results can serve as basis for further investigations into the precise nature of the limitations on NAD-learning, and for extending the debate as to whether dependency-learning is independent of positional information.

References

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 321 – 324.

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

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

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

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

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

Gerken, L., Landau, B. & Remez, R.E. (1990). Function Morphemes in Young Children's Speech Perception and Production. *Developmental Psychology*, 26 (2), 204-216

Goldsmith, J. A. (1999). An overview of Autosegmental phonology. In Goldsmith, J. A. (ed), *Phonological Theory: The Essential Readings*, Blackwell Publishers Ltd.,137-161.

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

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

Guasti, M. T. (2002). *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press.

Hochmann, J-R., Endress, A.D. & Mehler, J. (2010). Word frequency as a cue for identifying function words in infancy. *Cognition,* 115, 444-457.

Hohle, B., Schmitz, M., Santelmann, L.M. & Weissenborn, J. (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 & Johnson, E.K. (2010). Linking infants'distributional learning abilities to natural language acquisition. *Journal of Memory and Language*, 63, 197-209.

Kerkhoff, A., Bree, E. de, Klerk, M. de & Wijnen, F. (2011). Implicit learning and dyslexia: non-adjacent dependency-learning in infants at familial risk for dyslexia. In Danis, N. et al. (eds.). *BUCLD 35: Proceedings of the 35th Annual Boston University Conference on Language Development*, 35, 358-370.

Marcus, G.F., Vijayan, S., Bandi Rao, S. & Vishton, P.M. (1999). Rule-learning by seven-month-old infants. *Science* 283, 77-80.

McCarthy, J.J. (1999). A prosodic theory of nonconcatenative morphology. In Goldsmith, J. A. (ed), *Phonological Theory: The Essential Readings*, Blackwell Publishers Ltd.,162-184.

McClelland, J.L. & Plaut, D.C. (1999). Does generalization in infant learning implicate abstract algebra-like rules? *Trends in Cognitive Sciences*, 3 (5), 166-168.

Mintz, T. H. (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, 90, 91-117.

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

Onnis, L., Monaghan, P., Chater, N. & Richmond, K. (2005). Phonology impacts segmentation and generalization in speech processing. *Journal of Memory & Language*, 53, 225-237.

Onnis, L., Monaghan, P., Christiansen, M. H.& Chater, N. (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*. Erlbaum: Mahwah, NJ, 1047-1052.

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

Perfors, A., Tenenbaum, J.B., Griffiths, T.L. & Xu, F. (in press). A tutorial introduction to Bayesian models of cognitive development. *Cognition*.

Perruchet, P., Tyler, M.D., Galland, N. & Peereman, R. (2004). Learning nonadjacent dependencies: no need for algebraic-like computations. *Journal of Experimental Psychology: General*, 133 (4), 573-583.

Perruchet, P & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *Trends in Cognitive Sciences*, 10 (5), 233-238

Peters, A.M. & Stromqvist, S. (1996). The role of prosody in the acquisition of grammatical morphemes. In Morgan, J.L. & Demuth, K. (eds.), *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition*, Lawrence Erlbaum Associates, Inc., 215-232

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, 274, 1926 – 1928.

Saffran, J.R. & Wilson, D.P. (2003). From syllable to syntax: multilevel statistical learning by 12-month-old infants. *Infancy*, 4 (2), 273-284.

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

Seidenberg, M.S., MacDonald, M.C., Saffran, J.R. (2002). Does grammar start where statistics stop? *Science* 298, 553-554.

Selkirk, E. (1996). The prosodic structure of function words. (1996). In Morgan, J.L. & Demuth, K. (eds.), *Signal to Syntax: Bootstrapping From Speech to Grammar in Early Acquisition,* Lawrence Erlbaum Associates, Inc., 187-213

Shafer, V., Shucard, D., Shucard, J. & Gerken, L. (1998). An electrophysiological study of infants' sensitivity to the sound patterns of English speech. *Journal of Speech, Language, and Hearing Research*, 41, 874–886.

Shi, R., Cutler, A., Werker, J. & Cruickshank, M. (2006). Frequency and form as determinants of functor sensitivity in English-acquiring infants. *Journal of the Acoustical Society of America*, 119 (6), EL61-EL67.

Soderstrom, M., Conwell, E., Feldman, N & Morgan, J. (2009). The learner as statistician: three principles of computational success in language acquisition. *Developmental Science*, 12, 409-411.

Tincoff, R., Santelmann, L. & Jusczyk, P. (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 Language Development*, Sommerville, MA: Cascadilla Press, 726-737.

Appendix 1

Stimuli

No.	Y	a	X	b
1	Klepin	TEP	Poemer	LUT
2	Lotup	SOT	Kengel	JIK
3	Kiertan	RAK	Domo	TOEF
4	Fapoeg		Loga	
5	Griefup		Gopem	
6	Malon		Naspu	
7	Veirig		Vami	
8	Dufo		Snigger	
9	Tarsin		Rogges	
10	Seibor		Densim	

11	Nijfoe		Fidang	
12	Baduk		Rajee	
13	Floenie		Nilbo	
14	Tipla		Plizet	
15	Stepoer		Banip	
16	Blieker		Movig	
17	Muiblo		Sulep	
18	Kijbog		Wiffel	

(YX pairing were formed by combining each Y with the X on the same line with it, and with each of the 5 following Xs)

Dependencies

L1	L2
TEP_LUT SOT_JIK RAK_TOEF	TEP_JIK SOT_TOEF RAK_LUT

YX pairings in the test phase: Klepin _ Densim

Lotup _ Nilbo