Universiteit Utrecht

Nonadjacent Dependency Learning Across Domains in Typically Developing Children and Poor Readers

MA Thesis (30 ECTS)

Britt Hakvoort (0418552)

18-08-2009

Supervised by dr. Elise de Bree and dr. Annemarie Kerkhoff

External reader: Prof. dr. Frank Wijnen

Table of Contents

MA Thesis

Nonadjacent Dependency Learning Across Domains in Typically Developing School-Aged Children and Poor Readers

Abstract

Frequency has been demonstrated to be of great influence in language learning (a.o. Gómez 2002). Language development might be influenced by a deficiency in the ability to derive rules from frequent input. In fact, such a deficiency could even account for certain problems people with dyslexia experience in language. The proposed study set out to provide more insight into children's sensitivity to frequent patterns in (artificial) language and in the auditory non-linguistic domain. Based on Gómez (2002), Dutch school-aged children were presented with two nonadjacent dependency learning tasks. Neither typically developing children nor poor readers appeared to be sensitive to nonadjacent dependencies of the linguistic or non-linguistic type. Because the findings are likely to be influenced by several methodological factors, no clear conclusions can be drawn about the participants' capacity to learn nonadjacent dependencies from frequent input.

1. Background

Most people take their ability to read and write for granted. After all, these are skills acquired at a fairly early age and, though a tedious learning process is involved, most children pass through it successfully. Only some 3 to 10 % of children do not acquire literacy skills to the fullest: though their intelligence level is comparable to that of same-age peers, they experience problems in reading and writing (Snowling, 2000). This difficulty with reading and spelling, discrepant with intelligence, has been labeled dyslexia. Dyslexia has been categorized a language-based disorder. Several studies have revealed it to be a familial syndrome; children with dyslexic parents run a significantly larger risk (32 to 66%) of developing the disorder compared to children who do not have a dyslexic parent (e.g. Scarborough, 1990). Interestingly, literacy skill appears to be related to the child's early language development. Precursor studies have indicated that the development of a child at-risk of dyslexia in terms of spoken language may "contain indicators of dyslexia in individual children" (De Bree, 2007: 7). If we indeed assume that reading is a skill partly reliant on a child's level of spoken language, it is likely that if a child experiences difficulties in its development of spoken language, it will also encounter problems in acquiring literacy skills at a later age (De Bree, 2007).

Dyslexia as a Phonological Deficit

Language acquisition as a whole can be said to be strongly reliant on phonological development. Phonological development can be defined as a process in which the child acquires adult speech sound systems and patterns, needed to produce language. The phonological acquisition process has long since been studied in order to determine what constitutes normal and delayed development of the phonological system (Bernhardt and Stemberger, 1998), and to investigate how a delay in phonological development affects the language acquisition process. It has been found that deficits in the phonological system, for example, may result in a communication disorder not only affecting speech (in the case of Specific Language Impairment), but also the development of literacy skills (Ingram, 1981). Since people with dyslexia experience problems with the acquisition of literacy skills, it has been hypothesized that a core cognitive deficit in phonology is the underlying source of all the difficulties they experience (e.g. Vellutino et. al 2004). Though several researchers have brought forward their own version of a phonological deficit hypothesis (e.g. Brady 1997, who named it the *phonological representations hypothesis,* or Elbro 1996, under the name of *distinctness hypothesis*), the main assumption remains that dyslexics have poor phonological representations which cause them to experience language difficulty. "[The] phonological representations [of dyslexics] have been variously referred to as 'holistic', 'weak', 'fragile', 'fuzzy', 'indistinct' or 'underspecified'. In other words, the way in which their brain codes phonology is less efficient than that of normally reading children" (De Bree, 2007: 9). Due to a problem in their phonological development, they are less aware of and able to apply phonological rules and knowledge. This view has been supported by several studies, among those a longitudinal study by Bradley and Bryant (1983), which, according to Snowling, is one of the most influential studies that examines "the relationship between children's early phonological skills and their later reading achievement" (2000: 76). It demonstrates a relationship between a child's phonological awareness at age 4, and its literacy skills in terms of reading and spelling at age 8.

The terms phonological "awareness" and "skill" are rather broad. It was questioned by many which specific phonological skills can be said to be causally related to reading and writing. Muter et al. (1998) suggest that segmentation and sensitivity to rhyme are crucial predictors of a child's level of phonological awareness and later literacy skill. They tested a group of 4 to 6 year old children on segmenting ability and rhyming ability. Segmenting skill was tested by conducting a phoneme identification task, in which the child was presented with a picture of one-syllable word (e.g. a cat). The experimenter would pronounce *ca-,* after which the child had to complete the word. Another test to assess segmentation skill was a phoneme deletion task, in which children were presented with a word (e.g. *bus*), and were then asked to say the word with the initial phoneme deleted (e.g. *bus* then becomes *us*). Rhyming ability was defined by the child's abilities to detect and produce rhyme. The rhyme detection task required the child to match (rhyming) pictures to a target word. The rhyme production tasks required a child to produce rhymes for two target words. Muter et al.'s study (1998) demonstrated that the child's performance on the segmentation tasks was eventually a better predictor of literacy skill than rhyming ability. Overall, it has been suggested that "the key to reading development lies in performance on metaphonological tasks – tasks that require children to reflect consciously on the structure of spoken words" – like the tasks in Muter et al.'s study (Snowling, 2000: 79). Acquiring skills such as reading and writing strongly rely on these reflective abilities. It is likely that because of this, dyslexic children are slow in shaping their alphabetic skills compared to their normally developing peers. Because of the supposed problems they have with forming phonological representations, constructing a representation of phonemes and mapping this representation to a grapheme is difficult. Researchers supporting for example the connectionist view on language assume that a dyslexic child has a foundation that is not sufficiently optimized to cater for a proper mapping between phonology and orthography. Mappings that associate a complete word with its pronunciation are available to the dyslexic child but these mappings do not define the fine-grained links that exist between sequences of graphemes and phonemes (Brown, 1997). In other words, the mappings "are adequate for the establishment of associations between whole words and their pronunciations but [do] not support the development of fine-grained links between grapheme and phoneme sequences" (Snowling, 2000:88). As a consequence, dyslexic children will encounter difficulties encoding words they are not familiar with, such as nonsense words (Snowling[1], 2000). Thus, dyslexic children often try to apply unconventional reading strategies: it is possible that they build sight vocabularies[2] which allow them to read words without developing the decoding skills their typically developing peers have. Words that occur more frequently might become easier to read for the dyslexic child because he can rely on the visual knowledge and directly link the visual input to the meaning of a word. Typically developing children may also use this manner of reading, yet also develop decoding skills allowing them to process novel words more easily (Coltheart, 1978). From this perspective, an inference can be made about the speed at which a sight-vocabulary may expand when dyslexic children encounter novel words, for example when presented with a nonsense-word reading task: it is likely that this process is much slower (Snowling, 2000).

Dyslexic children also experience problems in spelling, which is predictable considering their difficulty in phoneme-grapheme conversion. It may be so that their knowledge of morphological regularities and certain orthographic conventions improves over time, yet, the core problem, namely their reduced phonological awareness, causes their new knowledge not to be brought forward properly. In other words: "If dyslexic children have phonological difficulties, then this will affect their ability to spell words phonetically, which [is] the foundation for their orthographic skill" (Snowling, 2000: 97).

However, a deficit on a phonological level might not be solely related to problems in orthographic skill. It may also affect other linguistic processes that are, for example, involved in syntactic or semantic processing, in that way causing a dyslexic person to have trouble with grammar or interpretation of sentences when reading or listening to spoken language.

Grammatical Deficits in Dyslexia

Grammatical rules aid a reader in comprehending and interpreting a written or spoken text. It is known that people with dyslexia often experience problems with sentence comprehension, agreement structures and interpretation. It thus appears their morphosyntactic knowledge is affected, too. Some scholars attribute the problems dyslexic people experience in the grammatical domain to a deficit in the phonological system as well. Rispens and Been (2007) investigated awareness of agreement structure in dyslexic and typically developing Dutch children. Children were

[1] Snowling does not endorse the connectionist view.

[2] A vocabulary based on visual recognition of a word as a whole.

presented with sentences like **de leuke clown maak een grapje* versus *de leuke clown maakt een grapje* (**the funny clown make a joke* versus *the funny clown makes a joke*) (Rispens and Been 2007). They were asked to indicate whether the sentence presented was correct or incorrect. Rispens and Been found that children with dyslexia were significantly less aware of subject-verb agreement structure than typically developing children. Their performance also correlated with performance on a non-word repetition task, "underscoring the idea that morphosyntactic deficits may stem from a problem with phonological working memory" (2007:303).

With respect to the reading problems and the problems with the interpretation of sentences that people with dyslexia encounter, Shankweiler and Crain (1986) have proposed the *processing limitation hypothesis*, which states that all linguistic structures are mastered before the age at which children begin acquiring literacy skills. It acknowledges that phonological decoding skills are essential in acquiring literacy skills, but also that phonological decoding is not the sole necessary element for reading. Working memory should also work efficiently in order to connect phonological and morphological representations. Assuming that the language faculty is a modular system, Shankweiler and Crain state that

> *"[a] modular view of the language mechanism raises the possibility that any number of components of the system might be the source of reading difficulties. At the same time, the fact that these components are related in a hierarchical fashion creates the possibility that a complex of symptoms of reading disorder may derive from a* single *affected component" (1986:140, emphasis in original).*

Shankweiler and Crain identify the "flow of information between the phonological buffer and the higher-level parsers" (1986:152) to be crucial for reading. In a modular system, the phonological information is transferred bottom-up and proceeds from the phonological to the semantic and syntactic parsers. If the phonological information is impaired, this may lead to a bottleneck which affects the flow of information, disturbing other processes in the language faculty, like syntactic processing. In other words, "[because] the bottom up flow of information from the phonological buffer is impeded by the difficulties in accessing and processing phonological information […] all subsequent processes in the language system will be adversely affected [in dyslexics]" (Crain and Shankweiler, 1990:543). Additionally, the cost for processing orthographic information becomes much higher when the phonological system is impaired. This puts a strain on working memory; resulting in reduced sentence comprehension, a symptom often recognized in dyslexics (Crain and Shankweiler, 1986). Empirical evidence supporting their view comes for example from Gottardo et al. (1996). Using a hierarchical regression analysis with the predictors verbal working memory, syntactic processing and phonological sensitivity, they found that the performance on several reading tasks cannot be predicted by the syntactic processing factor once the factors working memory and phonological sensitivity were entered into the regression model. In other words, phonological sensitivity and working memory are better predictors of the outcome of reading tasks than syntactic processing. Gottardo et al. state that "[t]his finding is certainly consistent with the idea that the predictive power of syntactic processing is an epiphenomenon of more basic limitations in phonological processes" (1996:578), in accordance with the *processing limitation hypothesis* by Crain and Shankweiler (1986).

Though the *processing limitation hypothesis* can provide an explanation for many difficulties dyslexics encounter in a variety of modules of language, some phenomena can hardly be accounted for by this theory. For example, Stein et al. showed that people with dyslexia tend to attribute the agent-role to the grammatical subject in a passive construction (1984). Structures like *John is eager to please* pose no problem for the dyslexic reader, yet in *John is easy to please* dyslexics interpret *John* as the agent of *please.* The *processing limitation hypothesis* cannot provide a unified explanation for these results because "both structures have identical surface forms, and therefore also an identical load on the verbal [working memory] system" (Rispens, 2004:19). It is possible that grammatical structures are delayed or even impaired, while not related to phonological processing deficits. Together, both impaired or delayed modules may impede on the acquisition of literacy skill.

Sensitivity to Grammatical Patterns in Typically Developing Infants

In order to be able to determine whether the acquisition of grammatical structures is actually delayed in dyslexics, the acquisition of such structures by typically developing children has to be examined first.

A number of scholars have investigated the sensitivity of infants to grammatical patterns in natural language. Santelmann and Jusczyk (1998) investigated sensitivity to morphosyntactic dependencies in typically developing English infants at 15 and 18 months using a Head Turn Preference procedure. The infants were presented with grammatical and ungrammatical dependencies in natural language of the respective type *is X Y-ing* and *can X Y-ing*, where X was an adverb and Y a verb. Santelmann and Jusczyk (1998) found that the infants were more sensitive to the grammatical dependencies at 18 months compared to 15 months olds, since the 15-month olds did not demonstrate a listening preference to the passages where the *is X Y-ing* dependency was maintained. This implies that the sensitivity to the dependency at hand is developed somewhere between 15 and 18 months. However, the 18-month old infants were only sensitive to the dependency when there were no more than 3 intervening syllables between the two key elements. The authors suggested that this was due to a more limited processing window. As age increases, the processing capacity increases and so at a certain point, children are able to process nonadjacent elements with a larger number of intervening syllables. It is thus possible that processing capacity and acquisition of grammatical elements are related.

Höhle et al. (2006) examined sensitivity to grammatical patterns in natural language in 19-month typically developing German infants based on Santelmann and Jusczyk's study. The German infants appeared not to be sensitive to the morphosyntactic dependencies. Given the structural differences between English and German (German has a freer word order, which often results in more than 3 intervening syllables between the nonadjacent elements) Santelmann and Jusczyk's reasoning could not explain Höhle's results. Having a limited processing window would barely allow German children to learn the nonadjacent dependencies in their language. Höhle et al. (2006) therefore suggest that the nature of intervening lexical items might keep the participants in their study from recognizing the nonadjacent dependency. Adverbs in German do not carry a marker, like in English (*-ly*), which might increase the difficulty of classifying these elements. This may lead the infants to focus on adjacent elements more, because they need to categorize the X element first before they can shift their attention towards the nonadjacent elements.

Interestingly, Mintz (2003) conducted a corpus study in which he demonstrated that words surrounded by the same frequently occurring frame can be categorized as belonging to a certain grammatical category. A frame here can be defined as two words separated by one intervening word that always occur together, for example *you X it*, where *X* is a transitive verb. An experiment with 12-month-old American infants shows that they are sensitive to these frames (Mintz, 2006). The infants were trained on sentences containing nonsense verbs and nouns, e.g. *She wants to lonk it* vs. *I see the bist in the room.* In the test phase, they were presented with the sentences they were trained on, but also with ungrammatical items, e.g. *She wants to bist it* and *I see the lonk in the room.* The infants showed a preference towards the grammatical sentences, implying they had indeed categorized *lonk* as a verb and *bist* as a noun. These infants were raised in an English speaking environment, and whether Mintz' findings are also applicable cross-linguistically is not clear. Yet, "[a] fundamental property of a frequent frame is that it is a relatively local context defined by frequently co-occuring units" (Mintz, 2006:39), and in heavily inflected languages these units are likely to be (highly frequent) inflectional morphemes. These morphological frames can still provide the necessary category information (Mintz, 2006). Erkelens, Kerkhoff and De Bree (in prep) conducted a study investigating frequent morpheme frames, based on the assumption that "infants search input for the smallest frequently co-occurring non-adjacent units" (Erkelens, 2009: 114). Using a Head Turn Preference Procedure, 16-month old Dutch infants were presented with sentences containing nonsense words, which were for example affixed with the Dutch diminuative marker *–je* or a plural verb marker *–en,* resulting in sentences like *ik zie een* plif-*je* vs. *wij* sook-*en niet hoor*. In the test phase of the experiment, the infants were presented with *ik zie een* sook-*je* and *wij* plif-*fen niet hoor.* Infants were expected to have categorized *plif* as a noun, and *sook* as a verb based on the morphological frames in the training phase. Indeed, children listened longer to the consistent than to the inconsistent items, suggesting that by 16-months Dutch infants can use frequent morpheme frames in the categorization of nonsense words.

Though many questions remain unanswered, the studies by Santelmann and Jusczyk (1998), Mintz (2003, 2006), Erkelens, Kerkhoff and De Bree (in prep) and Höhle et. al (2006) all appear to reach the same conclusion in some respect. All studies either implicitly or explicitly "suggest that […] distributional information is a viable basis for young children's early word categorization" (Mintz 2003:107). Children may categorize words based on input patterns, and the categorization of words may in turn help to unveil the grammatical structure. Frequency and distributional information thus appear to be an important cue in language learning as a whole. Perhaps then, a deficit in the processing of distributional and frequency information can account for the problems that people with dyslexia experience in the grammatical and phonological domain.

Frequency and Statistical or Sequential Learning

One of the available perspectives attempting to explain the problems dyslexics have was brought forward by Wijnen (2006). He hypothesized that a deficit in implicit sequence learning, which might underlie the categorization of phonological and grammatical elements, could partly cause dyslexia. As Wijnen states, pattern recognition is not only essential in acquiring phonemic categories. It is also important at the lexical, morphological and syntactic level. After all, it has been demonstrated that dyslexic children also experience problems in those areas (Wijnen, 2006). Frequency plays a crucial role in this type of (statistical) learning, as it proves to be a tool for infants to categorize linguistic input (see Gerken, 2002; Mintz 2003,2006). Taking Wijnen's hypothesis as a starting point, both the grammatical difficulties as well as the phonological developmental problems that dyslexics have can possibly be accounted for. If an infant is unable to extract statistical regularities from a frequently occurring pattern, it is likely that this may cause problems in categorizing linguistic elements in the domains of grammar and phonology, influencing the infant's early language development.

A study that took Wijnen's (2006) theory as a starting point was conducted by Wilsenach (2006). Based on Santelmann and Jusczyk (1998), she studied morphosyntactic dependency learning in Dutch 19-month old infants and infants at-risk of dyslexia. She demonstrated that 19-month old typically developing infants were sensitive to grammatical patterns in sentences such as *de zon heeft helder geschenen* vs. **de zon kan helder geschenen* (*the sun has shone brightly* vs. **the sun can shone brightly* where *can* is a modal verb in Dutch). Using a head-turn-preference procedure the children showed a preference for the grammatical sentence. However, the infants at-risk of dyslexia did not display the same pattern. They showed no preference for either the grammatical or ungrammatical sentence. These findings could indicate they were not able to track the morphosyntactic dependency between *heeft* and *ge-,* implying that the at-risk children lagged behind "in their ability to recognize a frequently occurring pattern in the input" (Wilsenach, 2006:78) and thus that their implicit sequence learning might be impaired. Possibly, this might be attributed to a reduced processing capacity in infants at-risk of dyslexia. However, Wilsenach indicated that the typically developing infants possibly only preferred the grammatical passages because *heeft* occurs more frequently than *can* in the input. Yet, a corpus study (van der Ven, 2004) showed that *kan* has a higher frequency than *heeft*, which would lead to the expectation that infants would have to prefer *kan X ge-* over *heeft X ge-* (X in this case being an adverbial expression, such as *helder*). The typically developing infants however did not. However, van der Ven's study also indicated that *heeft* and *ge-* occurred more often than *can* and *ge-* which may explain Wilsenach's findings.

Nonadjacent Dependency Learning

Important studies investigating statistical learning in the grammatical domain were conducted by (a.o.) Gómez (2002) and Gómez and Maye (2005). These studies served as a basis for studies with Dutch typically developing infants and adults, and infants at-risk of dyslexia and dyslexic adults (De Bree and Kerkhoff, in prep, see below). The studies by Gómez (2002) and Gómez and Maye (2005), with typically developing American infants, addressed nonadjacent dependency learning based on frequently reoccurring grammatical patterns in artificial language. The dependencies they used resemble the morphosyntactic dependencies in Santelmann and Juszcyk's study (1998), where *is* and *X–ing* always occurred together, separated by an adverb – a frequently occurring pattern in English. Though Santelmann and Juszcyk's 1998 experiment was conducted using natural language, Gómez (2002) and Gómez and Maye (2005) used artificial language in their studies. An advantage of using an artificial language is that infants are not familiar with the language. If they use statistical information to extract regularities in their native language, they should be able to learn the workings of the artificial language as well. Gómez and Maye (2005) rightfully point out that there are differences between natural and artificial language which cannot be overlooked. However, if similar trends are found in the studies by Santelmann and Jusczyk (1998) and Gómez (2002) and Gómez and Maye

(2005), then this "suggest[s] we are in fact tapping into related sensitivities" (Gómez and Maye, 2005: 187). Results from both Gómez (2002) and Gómez and Maye (2005) seem to support this statement.

Gómez (2002) investigated infants' sensitivity to statistical cues using an artificial language and found that 18-month old infants were able to discriminate untrained stimuli from trained stimuli. These stimuli consisted of 3 words, forming a string, where the first and last word always appeared together and the middle element was always different, forming an aXb and cXd dependency (e.g. *pel-wadim-jic*, *pel-kicey-jic*). When infants were trained on an aXb and cXd dependency in the familiarization-phase of the experiment and were presented with an aXb (grammatical) or aXd (ungrammatical) string in the test phase of the experiment, the infants noted the difference between the stimuli (Gómez, 2002). Moreover, Gómez and Maye (2005) demonstrated that the effects found differ depending on the age of the participants. They tested nonadjacent dependency learning (henceforth NADL) on infants of 12, 15 and 18 months of age. 12-month old infants did not appear to track the dependencies[3], whereas 15- and 18-month olds did. Yet, the ability to differentiate between trained and untrained strings seems to decrease when the set size of the intervening element X decreases. Gómez (2002) also tested NADL with adults, in order to investigate whether adults were also (or, still) sensitive to the dependencies. They were found to perform around chance level for set size X =2 and X=12, yet obtained an accuracy of almost 90% when X was 24. High variability of X thus appears to lead to better perception of nonadjacent dependencies, possibly because when conditional probabilities between adjacent elements are low (so, from A to X, X to B), participants switch their focus to the nonadjacent elements in the string (from A to B directly) (Gómez and Maye, 2005). When X is a 24-element set, the conditional probabilities of adjacent elements would be lower than when X is a set of 2 elements. Hence, it seems likely that the more frequent an X element occurs (thus, when the set size is small), it becomes harder to categorize the third (b) element because the focus is on the first two elements (a and X) leading to an increasing difficulty in detecting the nonadjacent dependency.

De Bree and Kerkhoff (in prep) conducted Gómez's experiment (2002) with Dutch typically developing infants and infants with a familial risk of dyslexia, and with adults, both good and poor readers. They used non-words of the artificial language as designed by Gómez (2002), adapted to Dutch phonotactics to make them sound like Dutch. De Bree and Kerkhoff did not replicate Gómez's results in their infant study. The infants were 18 months old and were trained on two types of dependencies (aXb and cXd), with a set size of 24 for the X elements. The infants did not show a preference for trained or untrained trials in the test-phase, suggesting they did not learn the nonadjacent dependencies. No differences were found between typically developing infants and infants at-risk of dyslexia. Their adult study, however, yielded results. The adults were trained on 3 different dependencies, aXb, cXd and eXf, using an X-set size of 24. In the test-phase, the participants were asked to identify familiar strings by answering yes or no. They answered yes more often if they

[3] Lany and Gómez (2008) found, however, that 12-month infants were able to generalize experience with adjacent dependencies to novel nonadjacent dependencies. When familiarized with an artificial language which contained adjacent dependencies, infants showed enhanced learning of nonadjacent dependencies. Infants trained on an artificial language which lacked adjacent dependencies did not display this pattern. This suggests that infants may benefit from prior experience with statistical learning.

heard a trained trial than when they heard an untrained trial, indicating that adults, both the good and the poor readers, are sensitive to the type of grammatical relation presented to them (De Bree, Kerkhoff and Wijnen, 2008). These findings raise some questions. The fact that Dutch-speaking adults are able to learn nonadjacent dependencies but that infants cannot, suggests that nonadjacent dependencies become available somewhere in the period in between infancy and adulthood – but when? The findings with respect to children at-risk of dyslexia on morphosyntactic dependency learning by Wilsenach (2006) suggested that the infants at-risk of dyslexia in De Bree and Kerkhoff's study were likely to perform differently from the typically developing group, yet they did not. Poor reading adults did not even perform differently from a control group at all. Is there an age at which these two groups do perform differently?

The present study sets out to answer these questions by targeting nonadjacent dependency learning in typically developing school-aged children of five and eight years old, and a group of eight-year old poor readers, using the Dutch stimuli developed by De Bree and Kerkhoff (in prep).

Nonadjacent Dependency Learning: Domain Specific or Domain General?

Some scholars have brought forward the idea that a statistical learning mechanism subserving language might not be language specific, but a general learning mechanism that might also aid us in developing our auditory and visual systems (e.g. Wijnen, 2006). Saffran et al. intended to study "whether learning in different domains can be at least partly subserved by the same knowledge-acquisition processes" (1999:28). Next to administering a word-segmentation task[4], Saffran et al. also presented their participants with a non-linguistic segmentation task. The stimuli here consisted of tones without any phonetic content that had the same statistical distributions as the stimuli in the speech task (Saffran et al., 1999). Both tasks were designed to have a full grammar[5]. Saffran et al. concluded that it is more likely that one segmentation mechanism is responsible for the segmentation of the tone-stream and the artificial language, rather than two domain-specific mechanisms. In a 2002 study, Saffran however hypothesizes that this type of statistical learning might be constrained. "[L]earners perform only a subset of the logically possible computations" (Saffran 2002: 173) so as to avoid irrelevant information in the input that possibly may cause a processing problem. In her study, Saffran aimed to study whether "predictive dependencies affect the learnability of sequential structure" and whether this possible constraint on learning is domain-general or domain specific (2002: 174). She found that predictive dependencies in the grammar designed for her experiment are important in acquiring syntax in language, but also in pattern recognition in auditory non-linguistic and visual domains. The predictive dependencies in her grammar seem to have aided the statistical learning process.

Evans et al. (2009) conducted a study similar to Saffran's, with school-aged language impaired (SLI) children between age 6;5 and 14;4. They hypothesized that if SLI could be characterized by a deficit in domain-general implicit learning capacity, they would find that children with SLI would perform

[4] This task was not targeting grammatical development but did investigate into statistical learning.

[5] Note the difference with Gómez 2002 and De Bree and Kerkhoff (in prep): their studies highlight a specific element present in a grammar: nonadjacent dependencies, whereas Saffran et al.'s experiment (1999) consisted of a complete grammar.

poorly on linguistic and non-linguistic stimuli compared to typically developing controls. When presented with a continuous speech stream, SLI children did not appear to be as sensitive to statistical information as typically developing children. After being exposed to the language for 21 minutes, meanwhile drawing a picture on a computer, they did not appear to have learned the existing word boundaries (that were defined by transitional probabilities), contrary to the typically developing children. SLI children needed double the exposure time to learn the word boundaries. A second experiment, using tones analogous to the speech stream, revealed that typically developing children were "able to group sequences of auditory 'events' in the same manner regardless of whether the input is linguistic […] or non-linguistic" (Evans et al., 2009:330). SLI children were not, suggesting that their mechanism which processes statistical information might not be fully functional. These findings lead to the question whether Dutch school-aged children with poor literacy skills might also experience these problems, since "the language problems of children with SLI and developmental dyslexia show a strong resemblance: the same type of problems are observed in the two clinically distinct groups" (Rispens and Been, 2007:295). The studies by Saffran et al. (1999), Saffran (2002), and Evans et al. (2009) moreover raise the question whether nonadjacent dependency learning might be specific to language, or not. The present study sets out to answer these questions by conducting a non-linguistic nonadjacent dependency experiment with Dutch typically developing five- and eight-year old children, and eight-year old poor readers.

2. Research Questions

The present study intends to investigate statistical learning in the context of nonadjacent dependencies, by testing Dutch eight-year old children on a task similar to that of Gómez (2002). Eight-year old children are in primary school and have learned how to read and write, hence possible problems in this area can already be identified and differences between good and poor readers can be assessed. Additionally, a group of typically developing five-year olds, who are still in their pre-literacy stage of development, will be tested. The performance of both these age-groups may be different and may thus provide more insight into the learning of (novel) nonadjacent dependencies.

Next to investigating whether eight-year old poor readers perhaps experience problems in learning the nonadjacent dependencies, as observed in the SLI participants in the Evans et al. (2009) study, our study also aims to establish whether the learning is domain specific or not in Dutch children (typically developing and poor readers). This will be done by also administering an auditory non-linguistic NADL task. The study will therefore attempt to answer the following research questions.

1. *Are school-aged children sensitive to nonadjacent dependencies?*
2. *Are there any differences in performance between poor readers and typically developing children?*
3. *Is there a difference in performance on the task between eight-year olds and five-year olds?*

4. *Are there differences in the children's performance in the linguistic and non-linguistic domain?*

The study at hand will attempt to answer these research questions by presenting poor readers and typically developing participants with a nonadjacent dependency learning task similar to the task as designed by Gómez (2002) and De Bree and Kerkhoff (in prep) and a similar task using non-linguistic auditory stimuli. Participants will be familiarized with an artificial language or, in the case of the non-linguistic NADL, tone sequences for 10 minutes, while focusing their attention on coloring a picture. In the test phase of the experiment, they are offered 6 grammatical items, that occurred in the training phase, and 6 ungrammatical, unfamiliar items. They are asked to listen to these items and to indicate on a score sheet whether they are familiar with the item at hand by circling "yes" or "no". If the participants learned the nonadjacent dependencies, they are expected to answer "yes" to all trained items, as these ought to be familiar to them, and "no" to all untrained trials.

The results of the study by Evans et al. (2009) suggest that the results of the present study can be expected to be different for controls and poor readers, but are expected to be roughly the same across domains based on Saffran et al., who found near-identical results in both the linguistic and non-linguistic domain in their 1999 study.

3. A Nonadjacent Dependency Learning Study

The present study consists of 4 nonadjacent dependency learning experiments: a linguistic task and a non-linguistic task administered to eight-year old participants, and a linguistic and non-linguistic task with five-year old participants. In total, 88 children took part in this study.

To examine whether age played a role in learning nonadjacent dependencies, a total of 31 five-year old children participated in the study. They were divided into two groups: 16 children took part in the linguistic version of the experiment (experiment 1), and 15 in the non-linguistic version (experiment 2). Children in these two groups were also asked to complete a series of background measures, assessing (phonological) short term memory and letter naming skill, in order to be able to identify children that might experience reading problems in the future.

The third and fourth experiments were conducted with 33 children without literacy difficulties and 24 eight-year old children at-risk of dyslexia[6] who were, on the basis of a series of background measures, classified as poor readers[7]. These background measures evaluated the children's reading abilities, (phonological) memory and spelling skills (for a description of the background measures, see Appendix II). If children in the control group performed poorly on these tasks, they were transferred to the poor readers group.

---

[6] Children at-risk of dyslexia are children with at least 1 dyslexic parent. Children who have a dyslexic parent run a larger risk (32 to 66% when both parents are dyslexic) of becoming dyslexic themselves (e.g. Scarborough, 1990)

[7] Most of them had not received an official diagnosis for dyslexia (yet), but their performance on the background measures was indicative of reading problems.

The typically developing children and the poor readers took part in both studies to be able to assess within-subject differences across tasks. The first experiment assessed children's ability to learn a linguistic type of nonadjacent dependency, whereas the second tested this ability based on non-linguistic strings of tones. To avoid any transfer of learning effects, the second experiment was conducted approximately 3 months after the first.

4. Experiment 1: NADL with Five-year Olds

Experiment 1 investigated whether five-year old typically developing children were able to learn nonadjacent dependencies in an artificial language.

Participants

Sixteen control participants between 5;5 and 6;8 (mean 5;4, SD 0;5) took part in this study. All participants were recruited through a primary school in Ooij, the Netherlands. Their primary school verified none of the participants suffered from mental- or health conditions that could have compromised their performance.

Apparatus

Stimuli were presented using an UiL-OTS laptop (Fujitsu Siemens LifeBook S7110), running Linux and the experimental program FEP (Veenker, 1998). Since participants were tested individually, stimuli were presented using Plantronics headphones.

Stimuli

In the first experiment, the children listened to one set of artificial language stimuli of the NADL experiment as designed by De Bree and Kerkhoff (in prep), e.g. *tep-wadim-jik, tep-poemer-jik*. For a complete list of stimuli, see Appendix I. There were 2 familiarization languages to ensure a balanced design, L1 and L2. In both languages, the set of X-elements comprised 18 items. The number of different dependencies was set to 2 considering the age of the participants.[8] In total, there were 36 different strings in each language, $aX_{1-18}b$, and $cX_{1-18}d$ in L1, and $aX_{1-18}d$ and $cX_{1-18}b$ in L2. Strings were presented randomly during the training phase, allowing for several repetitions of the same type of dependency. Each unique string was presented 6 times, totaling 216 strings. After the listening period, lasting approximately 10 minutes using an interstimulus interval of 750 ms, children were presented with 12 test stimuli; strings from L1 (6 strings) and L2 (6 strings). No dependency type was

[8] In De Bree and Kerkhoff's infant study, the set size for X was 24 with 2 dependencies, and for the adult study it was 18 with 3 dependencies. Considering the age of the children taking part in the current study, 3 dependencies are possibly too difficult. However, completely adopting the infant paradigm might result in the task becoming too easy. Hence, the participants were presented with a mixture of the adult and infant experiment.

repeated more than two times in a row. The participants were asked to answer "yes" when they thought they had heard the string before, and "no" if they had not. Because they had been familiarized with all adjacent pairs (aX…, cX…,) they were expected to note the crucial differences between the nonadjacent elements in the test stimuli. For an illustration of the stimuli used in the test-phase, see Table 1 (De Bree, Kerkhoff and Wijnen, 2008).

L1		L2	
aXb	**cXd**	**aXd**	**cXb**
tep-1-lut	sot-1-jik	tep-1-jik	sot-1-lut
tep-2-lut	sot-2-jik	tep-2-jik	sot-2-lut
tep-3-lut	sot-3-jik	tep-3-jik	sot-3-lut

Table 1 – Stimuli in the test-phase artificial language by De Bree and Kerkhoff. For children familiarized with L1, stimuli from L2 are ungrammatical and vice versa (De Bree, Kerkhoff and Wijnen, 2008).

Procedure

Children were tested individually. They were instructed to listen to strings of words. Using a Dutch example (*de mooie poes* vs. *poes de mooie, the pretty cat* vs. *cat the pretty*), the experimenter emphasized that the order of the words was important, as one particular word order in the artificial language denotes grammaticality. Pupils were presented with a coloring-picture which they could complete while listening to the stimuli. After the training phase, lasting approximately 10 minutes, the participants were told they would be presented with 12 strings of words in a particular order that they may or may not have heard during their familiarization. They were told that they would be asked whether they had heard the strings, with the words in that particular order, before. They were asked to answer with "yes" or "no". The experimenter then emphasized once more that the order of the words was important because if this would not be done, children would be more likely to interpret the question as "have you heard these words before". This might lead to the participants focusing on adjacent instead of nonadjacent elements, resulting in only "yes"-answers. After the instruction, the pupils were presented with the test-items, one by one. The experimenter entered their answers into the computer. For the full instructions, see Appendix III.

Each five-year old completed three background measures that may predict future literacy skills: a non-word repetition task (Gathercole et al., 1994) , a digit span forwards[9] (Wechsler, 2008) and a letter naming task (Wentink et al., 2008) to assess (phonological) memory and letter-sound recognition. Scores on these tasks may aid in identifying children that may experience reading problems in the future (which might influence their performance on the main task). For a complete description of the background measures, see Appendix II.

[9] Previously, five-year old children were tested on a digit span backwards, but this proved to be too difficult for them. Hence the choice for the forward version.

Results

This section presents the results on the background measures and of Experiment 1.

Background Measures

Background Measures - Overview	
Measure	**Mean Score**
NRT	21.4 (6.1)
Digit Span FWD	6.2 (1.2)
Letter naming	15.2 (10.5)

Table 2 – Mean scores and standard deviations (in brackets) of five-year old participants on background measures.

Only one child obtained weak scores on all three background measures, as her score was more than 1 standard deviation below the mean in all instances. Her score on these tasks might be indicative of future reading problems. Her scores on the main task were, however, not left out of the eventual analysis as in- or exclusion of her data did not influence the results.

NADL

Before analyzing the data, the participants' yes and no answers were converted into percentages of endorsements (yes-answers), per grammatical and ungrammatical (i.e. the familiar and unfamiliar) test trials. This was done because it was expected that the participants would answer yes to all trained trials and no to all untrained trials if they had learned the nonadjacent dependencies, resulting in 100% and 0% endorsements respectively. By calculating the percentages of endorsements, the trained and untrained condition could be compared properly.

Table 3 displays mean percentage endorsements for the group of participants as a whole, and divided by gender.

NADL - Overview			
Group/gender	**n**	**Mean % endorsements on trained trials**	**Mean % endorsements on untrained trials**
Control 5 yo	16	76.0 (25.1)	72.9 (30,9)
female	6	66.6 (38.0)	55.5 (40.4)
male	10	81.6 (12.3)	83.3 (19.2)

Table 3 – NADL overview. Mean percentage endorsements in the trained and untrained condition for the 5 year-old participants as a group, and per gender.

A Repeated Measures ANOVA with percentage endorsements as a dependent variable, with the within-subjects variable "grammaticality" (two levels: trained vs. untrained trials) and gender (male, female) as a between-subjects factor indicated there was no main grammaticality effect ($p=0.47$). Participants answered "yes" as often in the trained as in the untrained condition. Though differences in mean percentage endorsement appeared large, no main effect for gender was found ($p=0.10$). There were no interaction effects for gender and grammaticality either ($p=0.33$). Boys and girls did not score significantly different from each other.

To assess whether this learning is domain-specific for language, a non-linguistic NADL experiment (henceforth NL-NADL) was conducted with a second group of five-year old children.

5. Experiment 2: NL-NADL with Five-year Olds

Experiment 2 investigated non-linguistic nonadjacent dependency learning in five-year old typically developing children.

Participants

15 participants between 5;7 and 6;8 (mean 5;7, SD 0;5) took part in this study. All participants were recruited through a primary school in the Netherlands. None of the participants suffered from mental- or health conditions that could have compromised their performance, which was confirmed by the school.

Apparatus

See Experiment 1.

Stimuli

Stimuli in this experiment resembled stimuli in the linguistic task as much as possible. Each stimulus was built from 4 musical tones, forming an aXb/cXd sequence. In total, there were 36 different strings in each familiarization language, $aX_{1-18}b$, and $cX_{1-18}d$ in L1, and $aX_{1-18}d$ and $cX_{1-18}b$ in L2. The tones were chosen such that the string would not be melodic, to construct a nonsensical sequence like in the artificial language used in Experiment 1. The 18 different X-elements consisted of 2 tones, which both had half the duration of the first and last tone. This way, the non-linguistic strings were analogous to the four syllables forming the linguistic strings. The X-elements were chosen such that none of them contained the tones that comprised the a, b, c and d elements. Both a and c were tones in the lower register of the piano, and b and d in the higher register. This way, differences between L1 and L2 were reduced.[10]

All tones were recorded onto an Athlon 64 computer using a Korg SP250 stagepiano, and then edited for duration with Audacity for Windows. The average duration of the stimuli was 2 seconds. Stimuli were presented using FEP (Veenker, 1998) with an interstimulus interval of 750 ms. The training period lasted approximately 10 minutes, equivalent to a total number of 216 strings - each unique

[10] If a and b both would have been tones in the higher register in L1, it would become easier to distinguish strings from L2 in the test phase of the experiment. In L1, aXb would have been highXhigh and cXd would have been lowXlow. In L2, aXd and cXb would respectively have become highXlow and lowXhigh, a difference more easily distinguishable. The actual stimuli used for this experiment maintained their lowXhigh structure and thus continuity was preserved, as in the linguistic version of the experiment.

string was presented 6 times. Strings were presented randomly during the training phase, allowing for several repetitions of the same type of dependency. Participants were either trained on L1 or L2. After the training phase, children were presented with 12 test stimuli; strings from L1 (6 strings) and L2 (6 strings). No dependency type was repeated more than two times in a row. The participants were asked to answer “yes” when they thought they had heard the string before, and “no” if they had not. Because they had been familiarized with all adjacent pairs (aX…, cX…,) they were expected to note the crucial differences between the nonadjacent elements in the test stimuli. For an illustration of the stimuli used in the test-phase, see Table 4 and 5. For an overview of the stimuli, see Appendix I. For children familiarized with L1, stimuli from L2 are ungrammatical and vice versa.

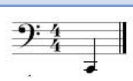
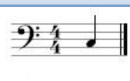
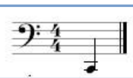
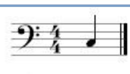
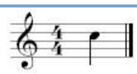
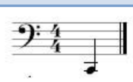
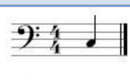
L1	
aXb	cXd
 -1- 	 -1- 
 -2- 	 -2- 
 -3- 	 -3- 

Table 4 – L1 stimuli in the test-phase of the non-linguistic NADL

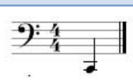
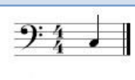
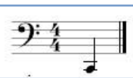
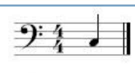
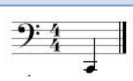
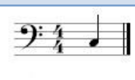
L2	
aXd	cXb
 -1- 	 -1- 
 -2- 	 -2- 
 -3- 	 -3- 

Table 5 – L2 stimuli in the test-phase of the non-linguistic NADL

Procedure

See Experiment 1. The instruction text was slightly altered for the NL-NADL: participants were told they were going to listen to sequences of musical tones, similar to language in that the order of the tones was important. This was exemplified using the *de mooie poes* example from Experiment 1 again. See Appendix III for the full instructions.

Results

The results of the NL-NADL can be found below, in addition to a comparison between the scores of the five-year olds that partook in the NADL task and the NL-NADL task.

NL-NADL

Table 6 represents the mean percentage endorsements on the NL-NADL.

NL-NADL - Overview			
Group/gender	**n**	**Mean % endorsements on trained trials**	**Mean % endorsements on untrained trials**
Control 5 yo	15	83.3 (17.8)	82.2 (19.3)
female	8	91.6 (12.6)	95.8 (7.7)
male	7	73.8 (18.9)	66.6 (16.6)

Table 6 – NL-NADL overview. Mean % endorsements in the trained and untrained condition for the 5 year-old participants as a group, and per gender.

A Repeated Measures ANOVA with dependent variable percentage endorsements , with the within-subjects variable “grammaticality” (two levels: trained vs. untrained trials) and gender (male, female) as a between-subjects factor indicated there was no main grammaticality effect ($p=0.65$). No interaction effect for grammaticality and gender was found ($p=0.10$). However, there was a main effect for gender ($F(1,13)=12.4$, $p=0.004$), indicating that males and females scored differently on the task. In this case, this specifically means that girls answered “yes” significantly more often than boys (see Table 6).

NADL and NL-NADL Compared

To investigate whether children experienced more difficulty with NADL or NL-NADL, percentages of endorsements in the trained and untrained conditions were used to calculate difference scores for both experiments. An independent t-test (two-tailed) with the difference scores on NADL and NL-NADL as a dependent variable and the group of five-year olds as an independent variable indicated that the mean scores on the tasks (NADL: mean 3.1 (24.5), NL-NADL: mean 1.1 (13.3)) did not differ significantly ($p=0.78$). This suggests that both tasks were equally difficult for the five-year olds.

The outcomes of Experiment 1 and 2 suggest that Dutch five-year old children are not able to learn nonadjacent dependencies in either the linguistic or non-linguistic domain. To assess whether their age may have influenced this, both the NADL and NL-NADL experiments were repeated with a group of typically developing eight-year old children. In addition, a group of poor readers took part in the study to assess whether there are differences in nonadjacent dependency learning between typically developing children and poor readers.

6. Experiment 3: NADL with Eight-year Olds

The third experiment studies nonadjacent dependency learning with an artificial language in typically developing eight-year olds and poor readers.

Participants

Thirty-three control participants between 8;1 and 9;11 (mean 8;4) and 24 poor readers between 7;7 and 9;1 (mean 8;2, SD 0;5) took part in this experiment. The participants were recruited through several primary schools in the Netherlands, where the teachers identified them as poor readers, and through the participant-database of a Utrecht University study entitled "Early language development in specific language impairment and dyslexia: A prospective and comparative study", conducted at the Utrecht Institute of Linguistics OTS. All children's native language was Dutch. Participation of bilingual children was minimized.[11] Teachers and parents confirmed none of the other participants suffered from mental conditions or hearing problems that might have compromised their performance on the main task[12].

Apparatus

See Experiment 1. Stimuli were presented to a group of pupils using a portable speaker system or, in the case of most[13] poor readers, a pair of headphones. Participants received coloring pictures and pencils, and a scoresheet on which they could note down their answers.

Stimuli

See Experiment 1.

Procedure

Similar to Experiment 1. However, in the case of the typically developing eight-year old children, the experiment took place in a classroom setting, using a portable speaker system. Instead of entering answers into the computer, children were given an answering sheet. If they thought they had heard a string with a particular word order before, they were instructed to circle "yes" on the answering sheet, and "no" if they thought they had not. After the instruction, the pupils were presented with the test-items, one by one. When all participants filled out their answer, the experimenter proceeded to play the next sound fragment, until all 12 questions were answered. For the score-sheet and the exact instructions, see Appendix III.

---

[11] The 2 bilingual children in the control group that partook in the study did not perform differently from the Dutch natives and were therefore included in the study. Their data being in- or excluded did not influence the results.

[12] One male participant in the poor reader-group suffered from the Von-Recklinghausen syndrome (neurofibromatosis, which may influence language development) but was not excluded from the study because in- or exclusion of his data did not influence the results.

[13] Seven poor readers, who all attended the same primary school, were tested simultaneously, following the procedure employed for control participants.

The poor readers were mostly tested individually. The procedure used was almost identical to the procedure for the typically developing children, except that the experimenter entered their answers to the questions directly into the computer instead of using a score sheet.

Either before or after the main task, each child individually completed the Een-Minuut-Test (Brus and Voeten, 1972), De Klepel (van den Bos et al., 1994) a digit span backwards (Wechsler, 2003), AVI-reading task (Visser et al., 1994) and a non-word repetition task (Gathercole et al., 1994) which were used to assess the child's reading abilities, working memory (a measure of intelligence) and phonological short term memory. Additionally, a dictation test (van den Bosch et al., 1993) was administered to assess the child's spelling skills. Scores on these tasks were used to identify poor readers in the control group. For a full description of the background measures, see Appendix II.

Results

This section presents the results of the administered background measures, and those of Experiment 3. The differences between the five-year olds and the eight-year olds will also be presented.

Background Measures

Before analyzing the data of Experiment 3, the participants' performance on the background measures was evaluated using an independent t-test, with EMT, Klepel, NRT, AVI, Digit Span Backwards and Dictation as test variables and group (i.e. typically developing or poor reader) as a grouping variable.

Background Measures – Overview			
Measure	**Typically Developing**	**Poor Readers**	**Independent t-test**
*EMT	55.1 (16.0)	33.0 (17.9)	t= 4.8, df 1,55, p=0.00
*Klepel	46.5 (23.9)	23.6 (16.3)	t= 4.0, df 1,55, p=0.00
*NRT	30.1 (5.7)	26.5 (6.9)	t= 2.1, df 1,54, p=0.03
*AVI	7.0 (2.1)	4.1 (2.6)	t= 4.5, df 1,54, p=0.00
Digit Span BWD	4.5 (1.2)	4.6 (1.2)	t=-0.2, df 1,54, p=0.80
*Dictation	34.8 (3.9)	20.9 (11.3)	t= 6.2, df 1,51, p=0.00

Table 7 – The mean scores and standard deviations (in brackets) for all background measures per group, including t-test outcomes (* indicates significant differences).

These background measures were conducted in order to establish the children's literacy skills. The means obtained by the typically developing children differed significantly from the mean scores of the children at-risk of dyslexia on all background measures except the Digit Span Backwards (see Table 7). Based on their performance on the background measures, children in the control group could be identified as poor readers. Each typically developing child completed the background measures and was by default assigned a reading profile, based on their (standardized) scores on the EMT and Klepel, two Dutch norm-referenced word-reading tests. Their combined scores result in a ++, +-, -+ or - - reading profile. Two plusses indicate a good score on both EMT and Klepel, +- indicates a good score on EMT and a poor performance on Klepel, whereas -+ indicates a poor performance on EMT and a good score on Klepel. Two minuses indicate a poor score on both tasks.

Children in the control group who were assigned a +-, -+ or - - reading profile and who scored more than one standard deviation below the group mean on at least two other tasks out of the NRT, AVI, and Dictation[14], were placed in the poor readers-group. This resulted in a transfer of 6 children. The analysis of the NADL experiment was thus conducted with a group of 27 typically developing children and 30 poor readers.

NADL

NADL-Overview			
Group/gender	**n**	**Mean % endorsements on trained trials**	**Mean % endorsements on untrained trials**
Control	27	74.7 (19.2)	73.5 (16.2)
female	17	75.5 (20.5)	76.5 (16.5)
male	10	73.3 (17.9)	68.3 (16.6)
Poor Readers	30	66.7 (17.5)	68.3 (22.9)
female	13	69.2 (15.0)	69.2 (21.3)
male	17	64.7 (19.4)	67.6 (24.6)

Table 8 – NADL overview. Mean percentage endorsements in the trained and untrained condition per gender, group, and gender per group. Standard deviations are between brackets.

Both groups' performance on NADL was assessed using a Repeated Measures ANOVA. The dependent variable was the percentage of endorsements, with as the within-subject variable grammaticality, which had two levels: trained trials vs. untrained trials. This way it was possible to investigate whether children were able to distinguish grammatical from ungrammatical strings. The factors group (typically developing vs. poor readers) and gender (male, female) were the between-subject variables, to investigate whether typically developing children and at-risk children performed differently and to identify possible gender-based differences.

No main effect for grammaticality was found, indicating that the percentage of "yes" responses in the trained, grammatical condition did not differ significantly from the percentage of endorsements in the untrained, ungrammatical condition (p=0.93). No main effect was found for group (p=0.17) and for gender (p=0.32), suggesting that there were no differences in performance on the task between the typically developing children and the poor readers, and between males and females. There were no significant interaction effects for grammaticality x group (p=0.59), grammaticality x gender (p=0.81) and grammaticality x group x gender (p=0.49) either. This indicates that both groups and both genders did not score significantly different from each other on each grammatical condition, and that both genders in each group did not score significantly different from each other on each grammatical condition.

[14] Digit Span Backwards assesses working memory, not a type of literacy skill. Therefore, only the performance on the core literacy tasks was used to classify the participants. The children had to score poorly on EMT, Klepel and at least two other tasks, because the children in the poor readers group on average also scored poorly on EMT, Klepel and two other measures out of NRT, AVI and Dictation (excluding Digit Span Backwards).

To ensure that both familiarization languages were equally difficult to learn, the effect of the familiarization language on grammaticality judgment was evaluated using a Repeated Measures ANOVA with grammaticality as within subject variables and familiarization language as a between subjects factor. This analysis yielded no significant result ($p=0.56$), ensuring the training languages did not differ from each other to the extent that it would have influenced the results.

A Repeated Measures ANOVA with the percentages of endorsements on trained and untrained trials as within-subjects variable and group as a between-subjects variable was conducted to test for a difference between five-year olds and eight-year olds, yet yielded no significant outcome ($p=0.46$). This implies there is no difference in performance between the five-year old children, the typically developing eight-year old children and the poor readers.

The results of Experiment 3 suggest that neither typically developing children nor poor readers at age 8 are able to learn linguistic nonadjacent dependencies in an artificial language. To assess whether this learning is domain-specific for language, NL-NADL was conducted with the same groups of children, 3 months after the linguistic version to prevent learning effects.

7. Experiment 4: NL-NADL with Eight-year Olds

Experiment 4 investigates the domain generality of nonadjacent dependency learning by subjecting eight-year old typically developing children and poor readers to auditory, non-linguistic nonadjacent dependencies.

Participants

See Experiment 3. Some participants were not able to participate due to personal circumstances, and in some cases, a system failure prevented participants from continuing the task. Hence, 24 control participants between 8;1 and 9;5 (mean 8;4, SD 0;5) and 15 eight-year old poor readers between 7;7 and 9;1 (mean 8;3, SD 0;5) took part in Experiment 4.

Apparatus

The apparatus used was identical to that in Experiment 3.

Stimuli

See Experiment 2.

Procedure

See Experiment 3. The instruction text was slightly altered for the NL-NADL: participants were told they were going to listen to sequences of musical tones, similar to language in that the order of the tones was important. This was exemplified using the *de mooie poes* example from Experiment 1 again. See Appendix III for the full instructions.

Results

The following section presents the results of Experiment 4 and the results of a comparison between the results of Experiment 3 and 4, and the results of a comparison between the eight-year olds and five-year olds.

NL-NADL

Table 9 displays mean % endorsement scores on NL-NADL obtained by typically developing children and poor readers in both the trained and untrained condition.

NL-NADL – Overview			
Group/gender	**n**	**Mean % endorsements on trained trials**	**Mean % endorsements on untrained trials**
Control	24	73.6 (23.5)	68.8 (26.2)
female	17	74.5 (21.3)	66.6 (26.3)
male	7	71.4 (29.9)	73.8 (26.9)
Poor Readers	15	75.5 (17.6)	75.5 (21.7)
female	6	63.9 (19.5)	69.4 (19.5)
male	9	83.3 (11.8)	79.6 (23.2)

Table 9 – NL-NADL overview. Mean % endorsements in the trained and untrained condition per gender, group and gender per group. Standard deviation between brackets.

Both groups' performance on NL-NADL was assessed using a Repeated Measures ANOVA. The dependent variable was the percentage of endorsements. The within-subject variable was grammaticality, which consisted of the two levels trained trials and untrained trials. This way it was possible to investigate whether children were able to distinguish grammatical from ungrammatical strings. The factors group (typically developing vs. poor readers) and gender (male, female) were the between-subject variables, to investigate whether typically developing children and at-risk children performed differently and to identify possible gender-based differences.

No main effect for grammaticality was found, indicating that the percentage of "yes" responses in the trained condition did not differ significantly from the percentage of endorsements in the untrained condition ($p=0.87$). No main effect was found for group ($p=0.67$) or for gender ($p=0.16$), suggesting that there were no differences in performance on the task between the typically developing children and the poor readers or between males and females. There were no significant interaction effects for grammaticality x group ($p=0.73$), grammaticality x gender ($p=0.96$) and grammaticality x group x gender ($p=0.37$) either. This indicates that both groups and both genders did not score significantly different from each other on each grammatical condition, and that both genders in each group did not score significantly different from each other on each grammatical condition.

The differences in average percentage endorsements on trained and untrained trials between the eight-year olds and five-year olds were assessed using a Repeated Measures ANOVA with the percentages of endorsements on trained and untrained trials as within-subjects variable and group as a between-subjects variable. The differences in performance were found not to be significant (p=0.75). There was no main effect for group (p=0.29) and there were no interaction effects between grammaticality and group (p=0.79).

To ensure that both familiarization languages in NL-NADL were equally difficult to learn, the effect of the familiarization language on grammaticality judgment was evaluated using a Repeated Measures ANOVA with trained and untrained endorsement percentages as within subject variables and familiarization language as a between subjects factor. This analysis yielded no significant result (p=0.41), ensuring the training languages did not differ from each other to the extent that it would have influenced the results.

NADL and NL-NADL compared

Neither Experiment 3 nor Experiment 4 yielded significant results. Taking reported difficulties participants had with non-linguistic NADL experiments (e.g. Evans et al. 2009) into consideration however, it is possible that Experiment 3 was still "easier" for participants than Experiment 4, despite the lack of learning effects in both tasks. To investigate this, the percentages of endorsements for the untrained condition of both experiments was subtracted from the percentage endorsements of the trained trials, resulting in a difference-score on each task. The difference scores on NADL and NL-NADL were compared for each group using a paired-samples t-test.

Difference NADL and NL-NADL		
	Mean NADL	**Mean NL-NADL**
Control	0.7 (20.5)	4.8 (35.2)
Poor readers	8.9 (24.3)	0.0 (21.8)

Table 10 – L2 stimuli in the test-phase of the non-linguistic NADL

There was no significant difference in the difference scores between the two tasks for controls (p=0.62) and for poor readers (p=0.22). A Pearson correlation showed no relation between difference scores on NADL and NL-NADL for controls (r=-0.03, p=0.89) or poor readers (r=0.3, p=0.27). Both tasks appear to have been equally difficult.

8. General Discussion

Expectations and Outcomes

Experiment 1 and 2 were conducted to provide an answer to the question whether school-aged children are sensitive to nonadjacent dependencies across domains. NADL-studies with Dutch infants and adults (De Bree and Kerkhoff, in prep) demonstrated that infants were not able to track the discontinuous dependencies in the stimuli, but adults were. This suggests that as age increases, at some point the ability to process and identify nonadjacent elements should become available. Hence, next to the group of eight-year old participants, a group of typically developing five-year old children took part the study to assess whether there were differences between them and the eight-year old controls and poor readers.

The results on Experiment 1, the linguistic NADL, suggest that the five-year olds have not perceived the pattern in the stimuli. The outcome of Experiment 2, the NL-NADL, implies that the children were not sensitive to the dependencies between the tonal stimuli either. In both cases, the mean percentage endorsements in the trained and untrained condition did not differ significantly from one another. Comparing the difference scores between the two experiments did not yield any significant outcome either, suggesting that the five-year olds perform equally poorly on the linguistic and non-linguistic task.

Experiment 3 and 4 meant to provide answers to the questions of whether age plays a role in the sensitivity to nonadjacent dependencies, whether this sensitivity is domain-general or domain-specific and whether typically developing children perform differently from poor readers. Results of previous studies suggest that typically developing infants, but also school-aged children are able to learn certain linguistic patterns implicitly and that this type of learning is domain-general (Evans et al., 2009; Gómez, 2002; Saffran, 2002). Yet, language impaired children need more exposure time to recognize linguistic patterns and that non-linguistic input is more difficult (Evans et al. 2009). Based on similarities between dyslexia and SLI, these difficulties were expected to surface in the results of the poor readers that took part in the present study.

The results of Experiment 3, the linguistic version of NADL, suggest that eight-year old children, both typically developing and poor readers, are not sensitive to the nonadjacent patterns in the stimuli. The results of both groups of participants on Experiment 4, the NL-NADL, did not meet the expectations either, as the children appeared not to have learned the dependencies. In that respect, the (absent) learning of nonadjacent dependencies appears to be domain-general, though considering the outcome of these experiments the evidence for domain-generality is not conclusive enough. After all, the poor performance on both tasks could have been influenced by extraneous factors (see below). A comparison between the two tasks revealed that the participants did not perform better on NADL than on NL-NADL or vice versa, which was also the case for five-year olds, suggesting both tasks were equally difficult for both groups. A comparison of scores between the eight-year olds and five-year olds did not yield any significant results either. This leads to the assumption that five-year olds and eight-year olds were not capable of acquiring the nonadjacent dependencies, both groups scored equally poorly. It could be observed that percentages endorsements on NADL were much higher for the five-year old boys, but this difference was not significant. The percentages endorsements on NL-NADL were much higher for girls, and here the difference between genders was significant. These results are difficult to interpret, but may be related to the relatively small number of participants that participated in each experiment.

Interpretation of the Outcomes

The outcomes of the four experiments did not match the expectations based on the literature. There are several possible methodological factors that may have influenced the results of both NADL and NL-NADL: testing method, the set-size of X, familiarization time, the cognitive demands on the tasks and in the case of NL-NADL the stimuli themselves.

The outcome of the NL-NADL may have been affected by exposure time and the number of X-elements (among others, see below), but it is possible that the obtained results were also influenced by the type of stimuli used. As mentioned, the musical tones used in the NL-NADL resembled an actual melody as little as possible to maintain the resemblance to the artificial language stimuli used in the linguistic version of the experiment. However, according to musicologist Wiering (p.c.) the perception of a tone in a sequence of 4 will always be dependent on the tone that has preceded it, which might have increased the difficulty of the NL-NADL. A way to solve this problem would be to devise sequences of 5 tones, which increases the number of intervals to 4. However, if this is done then firstly this would jeopardize the analogy between NADL and NL-NADL, and secondly, listeners would immediately perceive a key in the melody. This is not desirable, because then the stimuli would sound like actual music. Wiering (p.c.) advises, if the sequence has to consist of 4 tones, to use different types of instruments to create these tones. Perhaps an NL-NADL with that type of stimuli will yield different outcomes. However, the sounds of different instruments could be perceived as different voices pronouncing one word after another, which is undesirable as it might influence the perception. Also, the elements in the string would be more easy to label because some instruments are easily recognizable, which may influence the results. Children with a musical background would then have an advantage. The option as proposed by Wiering (p.c.) was actually taken into consideration before devising the stimuli, but was refuted based on the aforementioned problems.

The testing method employed in NADL and NL-NADL for typically developing eight-year olds, five-year olds and poor readers was different, one being in a classroom format and the other two on an individual basis. It is possible that the observed difference in mean percentage endorsement scores would have been higher if the typically developing eight-year old children had been tested individually as well. The poor readers and five-year olds might then have scored lower than the typically developing eight-year olds, the first group because they have difficulty extracting regularities from input, and the five-year olds because their learning mechanism still needs to develop further before they can complete a task successfully while at the same time being presented with distractor tasks. What could have happened presently, though, is that the group of eight-year old participants got distracted more easily during the task due to more prominent background noise from the classroom. Background noise is usually blocked when a participant wears headphones, which might positively influence performance. Perhaps, a future study could shed more light on this matter by examining all participants individually. If this does not yield different results, then the fact that poor readers do not score differently from controls may have to do with their grammatical abilities generally being less affected than their reading and spelling abilities (e.g. Snowling, 2000). This would then surface in the results of the linguistic NADL.

Another possible explanation for the absence of learning effects is the set size of the X-elements. Nonadjacent structures are not easy to learn. Onnis et al. suggest that this is because "learners have to overcome the bias toward adjacent transitional probabilities [first]" (2004: 1051). Gómez and Maye (2005) indicated that a larger set size of X makes it easier to perceive nonadjacent dependencies. The higher the number of X's, the more transitional probabilities between adjacent elements decrease. This makes it easier to focus on nonadjacent elements. The present study had 18 X-elements, which may explain the relatively high yes-bias: transitional probabilities between adjacent elements were relatively high, which may have led the participants to focus on adjacent elements only. All adjacent pairs in the test-phase had indeed been heard before, so when the children were asked whether they had heard a string of words or tones in a particular order before, they were more likely to say yes. They were familiar with the adjacent words and had failed to focus on the nonadjacent words due to the high(er) transitional probabilities between the adjacent elements. An increase of the set size of X, for example to 24 or 30, may reduce the focus on adjacent pairs and yield a different outcome overall because it reduces the overall difficulty of the task (but, see below).

Familiarization time might also have influenced the results. Evans et al. (2009) demonstrated that language impaired children require more input in order to be able to extract regularities from a continuous speech stream than typically developing children. Language impaired children needed 42 rather than the 21 minutes of training typically developing children needed, before they were able to distinguish boundaries in an artificial language. Also, it seems that adults need a lot of input when it concerns auditory non-linguistic stimuli. Gebhart et al. (2009) trained adults on dependencies consisting of 4 types of computerized noise triplets that were presented in a continuous stream. In the test phase, they listened to a part triplet and a triplet, and were asked with which of the two they were most familiar. Adults were not able to do this after 40 or 80 minutes of training. The investigators found that the participants needed 100 minutes of exposure time before they could identify the nonadjacent dependencies to be the familiar items.[15] Though Evans et al.'s experiment (2009) and Gebhart et al.'s experiment (2009) were slightly different in nature from the experiments in the present study, their outcomes do indicate that the amount of exposure time influences learning. It is possible that children, both typically developing and poor readers, need more exposure time to learn the nonadjacent dependencies than the mere 10-12 minutes of familiarization in the present study. However, De Bree and Kerkhoff (in prep) demonstrated that a familiarization time of approximately 12 minutes is enough time for adults to learn a total number of 3 dependencies with a set size of X=18 in an artificial language. Perhaps the cognitive capacity of children is still too limited to be able to analyze frequency patterns in the stimuli over such a short timeframe while at the same time focusing their attention on a distractor task, even though they were presented with only 2 dependencies. The size of the set of X-elements may even further increase the difficulty of the task (see above), so perhaps 10-12 minutes of familiarization with 18 different X-elements was just not sufficient for them. Yet, this raises the question why the *infants* in the study by Gómez (2002) were able to learn 2 nonadjacent dependencies with 18 different X-elements with only about *3 minutes* of exposure – their cognitive capacity is supposedly lower than that of school-aged children and the degree of difficulty of the stimuli is the same. It thus appears strange that the Dutch school-aged children show no learning effects. Evans (2009) suggests that the differences between the testing methods employed for infants and children have an influence on the learning. Because infants are subject to a preferential looking paradigm, they have a smaller cognitive load when listening to the stimuli than the school-aged children. Children need to complete a distractor task, coloring a picture,

[15] Stoimenovski (2009) also conducted a non-linguistic NADL, not with auditory but with visual stimuli. Good and poor adult readers were exposed to the stimuli for 10 minutes and were able to distinguish trained from untrained items. This is in sharp contrast with the results of Gebhart et al. (2009), but it might be attributable to the nature of the dependencies. Visual information can perhaps be labeled more easily than non-linguistic auditory information, in that the stimuli can be described, which might have aided the learning process.

to prevent them from actively listening to the stimuli[16]. The infants were not burdened with completing other tasks, making it easier for them to focus on what they hear, learning hereby becoming more explicit. This might explain why the infants in the studies of Gómez (2002) and Gómez and Maye (2005) did react to ungrammatical strings after a short exposure time, and also why the children in the present study did not, even though their familiarization phase lasted longer. This explanation, however, leaves the results of the adult study by De Bree and Kerkhoff (in prep) unaccounted for. The adults, both the controls and the poor readers, were able to distinguish trained from untrained nonadjacent items after 12 minutes of training. They also had to complete a distractor task, but their results on the main task imply that this did not cause any problems with respect to their cognitive load. A possible explanation for the adults' performance may be that they actually are aware that they are completing another task (the coloring), to distract them from the stimuli. The knowledge that they should focus on something other than what they hear can make them reflect more consciously on the stimuli (and learn them explicitly) because they want to perform well. They expect to be questioned about something specific present within the auditory material. It is possible that they are afraid that they will not identify this when they concentrate on the distractor task too much. Perhaps the children in the present study would have been able to learn the nonadjacent dependencies as well if they would have focused more of their attention on the stimuli instead of on the coloring. However, this almost requires them not to complete a distractor task, or a very simple one. However, not completing a distractor task would take away from implicit learning, and implicitness is an important premise of statistical learning. It is likely that the adults in De Bree and Kerkhoff (in prep) did not learn implicitly. The question thus arises whether they actually learned statistically or whether they employed conscious learning strategies.

The idea that infants can learn nonadjacent dependencies because they have a lower cognitive load, attributable to the testing method, is not only not applicable to the adult study by De Bree and Kerkhoff (in prep): it does not hold for the infant study by De Bree and Kerkhoff (in prep) either. Despite the fact that the infants are able to fully focus their attention on the stimuli during the training phase, they do not demonstrate the slightest learning effects in the test phase. It thus appears that a different explanation has to be found, one that, perhaps, can provide a unified explanation for the results of the Dutch infant and school-aged children studies. After all, it is highly unlikely that some humans, e.g. the infants in Gómez (2002) and Gómez and Maye (2005) and the children in Evans et al. (2009), are able to learn implicitly, but that others are not.

A major difference, and perhaps the only real major difference, between the present study and De Bree and Kerkhoff (in prep) and Gómez (2002) and Gómez and Maye (2005) and Evans et al. (2009) which can be identified is that the participants did not share the same native language. The children in the present study and in De Bree and Kerkhoff (in prep) have Dutch as their mother tongue, whereas the participants' native language in the other studies was English. It is possible that there is, then, to some extent, some L1 transfer to the artificial languages. As word order in Dutch is freer compared to English, it might be harder for Dutch infants and children to categorize the X element in

[16] But, see Andrade (2009): doodling may have a positive effect on learning. Coloring is not entirely the same but there are similarities, so instead of having enhanced the children's cognitive load, it could have enhanced the learning, though this was not apparent.

an artificial language because they know from their native language that there are several options to choose from. In a sentence such as *is X aan het Y* , (English: *is X Y-ing*), where Y is an infinitive, the X-element could be a noun, but also an adverb. An illustration: *is de tafel aan het schoonmaken* (*is cleaning the table)* or *is grondig aan het schoonmaken (is thoroughly cleaning).* In the English construction *is X Y-ing,* X can only be an adverb. Putting a noun in that position would render the sentence ungrammatical: **is the table cleaning*. It is possible that because Dutch native speakers have more categorization options for X available, their processing load increases when they try to apply L1 knowledge to an artificial structure, hence making it more difficult to learn the nonadjacent dependency because they spend more time on categorizing X. This could be happening in both the Dutch infants and in the school-aged children.[17] Höhle et al. (2006) demonstrated that German infants were not able to discern nonadjacent elements in natural language because it was harder for them to categorize the X, whereas the native-English infants in Santelmann and Jusczyk (1998) were able to, also using natural language. They suggest that indeed the grammatical content of X plays a role.

The explanation in terms of L1 transfer would also explain Evans et. al's (2009) and Gebhart et al.'s (2009) results on their studies with auditory non-linguistic stimuli in children and adults, mainly why such a long exposure time was needed. Because the stimuli were non-linguistic in nature and did thus not carry intrinsic word order information, it was not possible to transfer L1 knowledge onto the tonal stimuli.[18] Categorization had to start "from scratch" - assuming that the learning mechanism in place is domain-general. This may call for a longer exposure time. L1 knowledge is built by many hours of exposure. Applying this knowledge to a novel language should theoretically not take up much time, but as soon as this knowledge cannot be relied upon anymore, more input is needed. This would also explain why the children who partook in the NL-NADL experiment in the present study were not able to learn the dependencies. Perhaps an increase of exposure time would yield different results.

Though the L1 transfer-view may provide explanations, some problems arise concerning the linguistic version of NADL. An artificial language structure of the type aXb is analogous to multiple structures in natural language, for example *the X-s, and is X-Ying.* The construction *the X-s* is the same in English and Dutch, hence the difference in L1 should not influence the categorization of X, because in both languages X ought to be categorized as a noun. No other options are available, which should lead the participants to be able to focus their attention on the nonadjacent elements faster, and thus to learn the dependencies (learning also still depends on the set size of X). So, possibly, L1 transfer only applies to verbal constructions of the type *is X-Ying.* However, it is of course impossible to determine which type of structure the participants assign to the dependent elements in the artificial language – though the results of Gómez (2002) and Gómez and Maye (2005) versus those of De Bree and Kerkhoff (in prep) suggest that it was one of the type *is X-Ying*, as that would explain why the infants in Gómez (2002) and Gómez and Maye (2005) were able to learn the dependencies and the infants in

[17] The fact that Dutch adults are able to learn the dependencies, as mentioned above, can be attributed to a more conscious learning process.

[18] Though the visual stimuli in Stoimenovski (2009) did not carry word-order information intrinsically either, the fact that the adults in the study were able to learn the dependencies may have been due to the fact that the stimuli were easier to define and label.

De Bree and Kerkhoff (in prep) were not. A comparative study investigating nonadjacent dependency learning with both Dutch and English children could perhaps shed more light on this matter. Using explicit instruction, two groups of children, one English and one Dutch, can be made aware that in one case a, b, c and d elements in the artificial language are verbal constructions, and another two groups can be made aware that these are part of a nominal construction. Based on the differences between Dutch and English, the English group can be expected to learn the nonadjacent dependencies in the verbal construction because they experience no difficulty in categorizing the X, whereas the Dutch children cannot. The groups are not expected to differ on the other task. The outcomes of such a study could contribute to the knowledge about nonadjacent dependency learning as such, and provide insight into why children in the present study were not able to learn them. It will also be interesting to include poor readers in the proposed study. If it was indeed the case that none of the children were able to learn the nonadjacent dependencies because of the ambiguous X-category, in that they could assign X to multiple categories, then disambiguation should reveal possible differences between poor readers and typically developing children, for example if poor readers need a longer exposure time to be able to extract information from the input.

9. Conclusion

This study attempted to evaluate nonadjacent dependency learning in Dutch typically developing children at age 8 and 5, and poor readers at age 8 across the non-linguistic and linguistic domains. Overall, the outcome of the study suggests that Dutch children at age 5 and 8 do not perceive nonadjacent dependencies in the linguistic or non-linguistic domain. It is not clear why the children in the current study were not able to learn the nonadjacent dependencies, whereas children in other studies were. Also, why no differences were found between the poor readers and the typically developing children remains unclear. All differences between the results of the present study and those of other studies may be attributed to the amount of exposure time to the training language, cognitive demands, the variability of the X-element and testing methods. In the case of NL-NADL the type of stimuli may also have played a role. However, the fact that infants and children in other studies are able to extract information from input implicitly using the same experimental procedures and settings, does raise questions about the workings of statistical learning. It is possible that not only transitional probabilities play a role in learning nonadjacent elements, but also the categorization of X and the participants' native language. Perhaps, a study looking into L1 knowledge transfer may provide more insight into this.

Acknowledgements

First and foremost I would like to say that in general, working on this thesis has been great fun. I have a bunch of people to thank for that. NADL-meisjes, in particular supervisors Annemarie and Elise, it has been great working with all of you. Thank you for your patience, our talks, and the theekransjes with or without brownies. Frank, Desiree, and all the other "Monday-morning-labmeeting"-regulars, thank you for all your valuable input. Girls from the "dungeon", thanks for making working in the dark much more pleasant, I had a blast. To all my other friends and loved ones, Tasja in particular: thank you for putting up with me, and for showing interest in what I was doing, even though you probably had no idea what "naddul" was.

Last, but not least, I would like to express my gratitude to all children and their parents, and the schools who partook in this study. Your contribution was invaluable.

References

Andrade, J. (2009). "What Does Doodling Do?" *Applied Cognitive Psychology.* Published online by Wiley InterScience.

Bernhardt, B.H., and Stemberger, J.P. (1998). *Handbook of Phonological Development.* New York: Academic Press.

Bos, K.P van den, Lutje Spelberg, H.C. Scheepstra A.J.M. and Vries J.R. de. (1994). *De Klepel, een Test Voor de Leesvaardigheid van Pseudowoorden. Verantwoording, Handleiding, Diagnostiek en Behandeling*. Nijmegen: Berkhout.

Bosch, L., van den, Gillijns, P., Krom, R., and Moelands, F. (1993). *Leerling Volgsysteem: Schaal Vorderingen in Spellingvaardigheid 1, Groep 3– 4. Handleiding.* Arnhem: CITO.

Brady, S.S. (1997). "Ability to Encode Phonological Representations: An Underlying Difficulty of Poor Readers". In B.A. Blachman (ed.) *Foundations of Reading Acquisition and Dyslexia: Implications for Early Intervention.* Mahwah: Lawrence Erlbaum, 21-48.

Bree, E. H. de (2007). *Dyslexia and Phonology: A Study of the Phonological Abilities of Dutch Children At-Risk of Dyslexia*. Academisch Proefschrift, Universiteit Utrecht / Landelijke Onderzoeksschool Taalwetenschap. No. 155.

Bree, E. H. de and Kerkhoff, A. O. (in prep). *Nonadjacent Dependency Learning in Children (At-risk of) Dyslexia and Adults with Dyslexia.* Unpublished manuscript. Utrecht, UiL-OTS.

Bree, E. H. de, Kerkhoff, A. O. and Wijnen, F. N.K. (2008). *Plif, Saen, Jik and Lut: Category Formation in Phonology and Grammar in Infants.* Psychologie, Pedagogiek en Taal-Seminar. 11 dec. 2008

Brown, G. D. A. (1997). "Connectionism, Phonology, Reading, and Regularity in Developmental Dyslexia". *Brain and Language*. 59: 207-235.

Brus, B. Th., and Voeten, M. J. M. (1972). *Een-Minuut Test. Vorm A en B.* Nijmegen: Berkhout.

Coltheart, M. (1978). "Lexical Access in Simple Reading Tasks." In G. Underwood (ed.), *Strategies of Information Processing,* London: Academic Press, 151-216.

Crain, S. and Shankweiler, D. (1990). "Explaining Failures in Spoken Language Comprehension by Children with Reading Disability." In: D.A. Balota, G.B. Flores d'Arcais, and K. Rayner (eds.) *Comprehension Processes in Reading.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Erkelens, M. A. (2009). *Learning to Categorize Verbs and Nouns.* Academisch Proefschrift, Universiteit van Amsterdam / Landelijke Onderzoeksschool Taalwetenschap. No. 211.

Erkelens, M.A., Kerkhoff, A.O. and Bree, E.H. de (in prep). *Dutch 16-month-olds Categorize Nonsense Words Based on Frequent Morpheme Frames.* Unpublished manuscript, Amsterdam, ACLC / Utrecht, UiL-OTS.

Evans, J. L., Saffran, J. R. and Robe-Torres, K. (2009) "Statistical Learning in Children with Specific Language Impairment". *Journal of Speech, Language, and Hearing Research*. 52: 321-335.

Gathercole, S.E., Willis, C.S., Baddeley, A.D., and Emslie, H. (1994). "The children's test of nonword repetition: A test of phonological working memory." *Memory,* 2: 103-127.

Gebhart, A.L., Newport, E.L. and Aslin, R.N. (2009). "Statistical Learning of Adjacent and Nonadjacent Dependencies among Nonlinguistic Sounds". *Psychonomic Bulletin and Review,* 16: 486-490.

Gerken, L. (2002). Early sensitivity to linguistic form. *Annual Review of Language Acquisition*. 2: 1-36.

Gómez, R. (2002) "Variability and Detection of Invariant Structure". *Psychological Science,* 13: 431-436.

Gómez, R. and Maye, J. (2005). "The Developmental Trajectory of Nonadjacent Dependency Learning". *Infancy*, 7: 183-206.

Gottardo, A., Stanovich, K. and Siegel, L. (1996). "The Relationships between Phonological Sensitivity, Syntactic Processing, and Verbal Working Memory in the Reading Performance of Third-Grade Children". *Journal of Experimental Child Psychology*, 63: 563-582.

Höhle, B., Shmitz, M., Santelmann, L.M., and 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: 277–300.

Ingram, D. (1981). *Procedures for the Phonological Analysis of Children's Language.* Baltimore: University Park Press.

Lany, J. and Gómez, R. (2008). "Twelve-Month-Old Infants Benefit from Prior Experience in Statistical Learning". *Psychological Science*, 19: 1247-1252.

Mintz, T.H. (2003). "Frequent Frames as a Cue for Grammatical Categories in Child Directed Speech". *Cognition,* 90: 91-117.

Mintz, T.H. (2006). "Finding the Verbs: Distributional Cues to Categories Available to Young Learners". In K. Hirsh-Pasek and R. M. Golinkoff (Eds.), *Action Meets Word: How Children Learn Verbs*. Oxford: Oxford University Press, 61-63..

Muter, V., Hulme, C., Snowling, M., and Taylor, S. (1998). "Segmentation, Not Rhyming, Predicts Early Progress in Learning to Read". *Journal of Experimental Child Psychology,* 71: 3-27.

Ramus, F., Rosen, S., Dakin, S.C., Day, B.L., Castellote, J.M., White, S. and Frith, U. (2003). "Theories of Developmental Dyslexia: Insights from a Multiple Case Study of Dyslexic Adults". *Brain*, 126: 841-865.

Rispens, J. E. (2004). Syntactic and Phonological Processing in Developmental Dyslexia. Doctoral Dissertation. Groningen: Grodil. No. 48.

Rispens, J. E. and Been, P.H. (2007). “Subject-verb agreement and phonological processing in developmental dyslexia and SLI: A closer look”. *International Journal of Language and Communication Disorders*, 42: 293-305.

Saffran, J.R. (2002). “Constraints on Statistical Language Learning”. *Journal of Memory and Language,* 47: 172-196.

Saffran, J.R., Johnson, E.K., Aslin, R.N. and Newport, E.L. (1999). “Statistical Learning of Tone Sequences by Human Infants and Adults”. *Cognition*, 70: 27-52.

Santelmann, L.M. and Jusczyk, P. W. (1998). “Sensitivity to Discontinuous Dependencies in Language Learners: Evidence for Limitations in Processing Space”. *Cognition*, 69: 105-134.

Scarborough, H.S. (1990). “Very Early Language Deficits in Dyslexic Children”. *Child Development,* 61: 1728-1743.

Shankweiler, D. and Crain, S. (1986). “Language Mechanisms and Reading Disorder: A Modular Approach”. *Cognition,* 24: 139-168.

Snowling, M.J (2000). *Dyslexia.* Oxford, Blackwell Publishers Ltd.

Stein, C., Cairns, H.S., and Zurif, E. (1984). “Sentence Comprehension Limitations Related to Syntactic Deficits in Reading Disabled Children”. *Applied Psycholinguistics,* 5: 305-321.

Stoimenovski, M. (2009). *Non-Adjacent Dependency Learning in Adults and Adults with Dyslexia: The Visual Domain*. Unpublished manuscript. Utrecht, UiL-OTS.

Veenker, T. (1998) *FEP.* Experimental script for UiL-OTS Labs.

Ven, S. van der (2004). *Developmental Dyslexia and Implicit Knowledge of Morphosyntactic Agreement: a Preferential Listening Study with 19-month-old Infants.* Unpublished manuscript. Utrecht, UiL-OTS.

Vellutino, F.R., Fletcher, J. M., Snowling, M.J. and Scanlon, D.M. (2004). “Specific Reading Disability (Dyslexia): What Have We Learned in the Past Four Decades?” *Journal of Child Psychology and Psychiatry.* 45: 2-40.

Visser, J., Laarhoven, A. van, and Beek, A. ter (1994). *AVI toetspakket: Handleiding.* KPC: ’s Hertogenbosch.

Wiering, F. (2009). Personal Communication, May 2009.

Wechsler, D. (2003). “Digit Span Backwards”. *Wechsler Intelligence Scale for Children –IV*. Pearson: San Antonio.

Wechsler, D. (2008). “Digit Span Forwards”. *Wechsler Adult Intelligence Scale-IV.* Pearson: San Antonio.

Appendix I

This appendix contains the stimuli of NADL and NL-NADL respectively.

NADL stimuli

The dependencies used in the linguistic NADL were the following:

a. Tep (tɛp)
b. Lut (lʏt)
c. Sot (sɔt)
d. Jik (jɪk)

resulting in tep-$X_{1\text{-}18}$-lut, tep-$X_{1\text{-}18}$-jik, sot-$X_{1\text{-}18}$-jik and sot-$X_{1\text{-}18}$-lut. The X-elements can be found in the table below.

X-Elements - Overview					
Number	**X**	**Number**	**X**	**Number**	**X**
1	Wadim (‘wadɪm)	7	Godem (‘χodəm)	13	Rogges (‘roχəs)
2	Kasi (‘kaisi)	8	Naspoe (‘nɑspu)	14	Deuzem (‘døzəm)
3	Poemer (’pumər)	9	Hiftam (‘hɪftɑm)	15	Fidang (‘fidɑŋ)
4	Kengel (‘kɛŋəl)	10	Ditsja (‘ditʃa)	16	Gepag (‘ʒepɑχ)
5	Domo (‘domo)	11	Vaki (‘vaki)	17	Seta (‘seta)
6	Loga (‘loχa)	12	Snigger (‘sniχər)	18	Noeba (‘nuba)

Table 11 – The phonetic transcription of all 18 X-elements used in the linguistic NADL

NL-NADL stimuli

The dependencies used in the non-linguistic NADL are displayed in the table below.

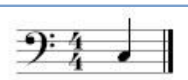
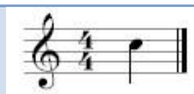
Dependencies - Overview		
Dependency	**Visual representation**	**Note name**
a		C_2
b		C_4
c		C_3
d		C_5

Table 12 - Dependencies for NL-NADL

The dependencies thus are C_2- $X_{1\text{-}18}$- C_4, C_2-$X_{1\text{-}18}$- C_5, C_3-$X_{1\text{-}18}$- C_5 and C_3-$X_{1\text{-}18}$- C_4. The X-elements can be found in the table below.

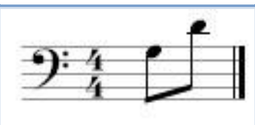
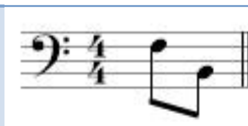
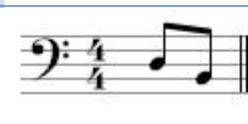
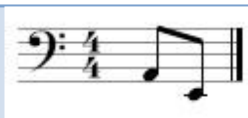
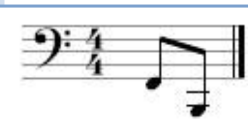
X-Elements - Overview					
X	**Visual representation**	**Note name**	**X**	**Visual representation**	**Note name**
1		D_4 F_4	10		D_5 B_5
2		E_4 A_5	11		A_5 E_4
3		G_4 D_5	12		F_4 B_4
4		D_3 G_3	13		D_4 B_4
5		E_3 A_4	14		A_4 E_3
6		F_3 D_4	15		F_3 B_3
7		D_2 G_2	16		D_3 B_3
8		E_2 A_3	17		A_3 E_2
9		G_2 D_3	18		F_2 B_2

Table 13 – X-Elements for NL-NADL

Appendix II

This appendix carries a description of all background measures used, both for the eight-year old and the five-year old participants. An (8) or (5) indicates to which age group the task was administered.

Eén Minuut Test (EMT) (8)

B.Th. Brus and M.J.M. Voeten (1972)

The EMT is a task which can establish a general level of technical reading skill. Technical reading indicates the pronouncing of a printed or written text. The EMT consists of a series papers with words on it, printed in rows. The words and papers become more difficult as the number of sounds and syllables per word increases. The score is determined by the number of correctly read words in a one-minute timeframe.

Klepel: reading nonsense words (8)

K.P. van den Bos, H.C. Lutje Spelberg, A.J.M. Scheepstra and J.R. de Vries (1994)

The Klepel, like the EMT, is meant to determine the level of technical reading skill. It consists of a paper with nonsense words on it, printed in rows. Because of the absence of word meaning, the Klepel requires a child to really be able to directly translate the word into sound. The nonsense words become more difficult. The score is determined by the number of words read correctly within a two-minute timeframe.

Schaal Vorderingen in Spellingvaardigheid (SVS) CITO (8)

L. van den Bosch, P. Gillijns, R. Krom, and F. Moelands (1993)

With the Schaal Vorderingen in Spellingvaardigheid, the children's spelling abilities are tested. This study made use of the version for 4^{th} grade children. The children are presented with a sentences, and writes down the word that is mentioned again after the sentence. For example: "there is a picture in this book"…"write down, book". The test consists of 38 items. The score is determined by the number of correctly spelled words.

AVI-niveau lezen (8)

J. Visser, A. van Laarhoven and A. ten Beek (1994)

The AVI-niveau reading task determines the child's reading skill when reading stories. One AVI-series consists of 9 stories, increasing in difficulty. Each story corresponds to the reading level, ranging from AVI 1 to 9. Accuracy and speed are important. Each AVI-level assesses the number of mistakes and the time it took the child to complete the story. The reading level is thus determined by the reading time and the number of mistakes.

Digit Span Backwards (8)

D. Wechsler (2003)

The Digit Span Backwards assesses working memory. The experimenter reads out a series of numbers, which the child has to repeat in reverse order, for example: 2-3, to which the correct answer is 3-2. The items become increasingly difficult. There are 12 items in total. Scores are determined through the number of correct items.

Nonwoord-Repetitie-Test (5, 8)

S.E. Gathercole, C.S. Willis, A.D. Badderly en M. Enslie (1994)

The Nonwoord Repetitie Test (Nonword Repetition Task) is meant to determine the level of the verbal short term memory. Evidence exists that memory is related to early literacy skill. Children with weak literacy skill often score poorly on verbal memory tests. The Nonwoord Repetitie Test consists of 48 nonsense words, presented to the child through headphones. The degree of difficulty ranges from 2 to 5 syllables per word. The child has to repeat the word. The score is determined through the number of correct items.

Digit Span Forwards (5)

D. Wechsler, 2008

The Digit Span Forwards assesses short-term memory. The child is presented with a series of numbers and is asked to repeat this. The items become more difficult as the test progresses. There are 16 items in total. Scores are determined through the number of correct items.

Letternaming (5)

H. Wentink, L. Verhoeven and M. van Druenen (2008)

The letternaming task requires children to name letters and combinations of letters that occur in the Dutch language. They have to make the sound that corresponds to the letter. Scores on this task may be indicative of future literacy skill. In total, there are 36 items. Scores are determined through the number of correct items.

Appendix III

Scoresheets/answer booklets used with eight-year-old children, for NADL and NL-NADL, and the corresponding instructions.

Instructions NADL

Voor jullie hebben jullie allemaal een boekje, daar mogen jullie je naam zetten.

Dadelijk krijgen jullie een luisterfragment te horen. Als je dadelijk luistert hoor je misschien wat gekke woordjes; dat komt omdat jullie naar zinnetjes uit een nieuwe taal gaan luisteren. Ieder zinnetje heeft 3 woordjes, en de woordjes staan in een bepaalde volgorde. Bijvoorbeeld: hum-dossim-mif. De volgorde van de woordjes is erg belangrijk, net als in onze taal. Als ik bijvoorbeeld zeg "Poes de mooie", dan weten jullie vast wel dat dat zinnetje niet helemaal klopt, want het moet natuurlijk zijn: "De mooie poes" Jullie gaan dus naar deze taal luisteren en ondertussen mogen jullie de kleurplaat in jullie boekje inkleuren. Als het luisterfragment afgelopen is, dan krijgen jullie 12 vragen. De vragen gaan we meteen na het luisteren beantwoorden. Is iedereen er klaar voor? Dan gaan we beginnen.

(start experiment)

We gaan nu de vragen beantwoorden. De vragen staan in jullie boekje. Sla je bladzijde maar om en kijk naar vraag 1. Iedere vraag gaat zo: heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin? Je hoeft alleen maar ja of nee te antwoorden, dat mag je omcirkelen in het boekje, en het gaat bij dit onderzoekje echt om wat jij gehoord hebt! Dus niet wat je buurman of buurvrouw gehoord heeft. Er zijn geen goede of foute antwoorden, dus werk gewoon lekker voor jezelf. Ik speel ieder zinnetje één keertje af. Let dus goed op. Als er geen vragen zijn dan gaan we beginnen. Denk erom, de volgorde van de woordjes is belangrijk!

[You all received a booklet, you may write your name on it. In a few minutes you are going to listen to something. If you listen carefully you will hear some funny words, that is because you are going to listen to sentences from a new language. Every sentence has 3 words, which are in a particular order, for example: hum-dossim-mif. The word order is very important, just as it is in our language. If I say: "cat the nice', then you probably all know that that sentence is not correct, because the right answer is "the nice cat". So, you are going to listen to this language and meanwhile you may color a picture. As soon as the listening period is over, you will get 12 questions. We will answer them straight after listening. Is everybody ready? Let's start.

(experiment starts)

We will now answer the questions. The questions are in your booklet. Turn the page and look at question 1. Every question goes as follows: did you hear this sentence in this order before, with exactly the same words in it? You only have to answer yes or no, which you can circle in your booklet. This experiment is all about what you heard! So what your neighbor answers is not important. There are no right or wrong answers, so just do your best. I play each sentence only once, so pay close attention! If there are no questions we start.

Remember, the order of the words is important!]

Onderzoek Dyslexie

Universiteit Utrecht

Naam:..

School:...

Groep:..

Geboortedatum:..

Testdatum:..

Vragenlijstje

1. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

2. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

3. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

4. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

5. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

6. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

7. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

8. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

9. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

   Ja nee

10. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

    Ja nee

11. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

    Ja nee

12. Heb je dit zinnetje in deze volgorde al eens eerder gehoord, met precies dezelfde woordjes erin?

    Ja nee

Einde. Dankjewel voor het meedoen!

Instructions NL-NADL

Voor jullie hebben jullie allemaal een boekje, daar mogen jullie je naam zetten.

Dadelijk krijgen jullie een luisterfragment te horen. De vorige keer hebben jullie naar zinnetjes geluisterd uit een gekke taal, maar nu gaan jullie naar melodietjes luisteren. De melodietjes bestaan uit 3 klanken of muziektonen, en die drie klanken vormen samen een reeks. Eigenlijk is iedere klankenreeks een soort zinnetje, want de volgorde waarin de klanken staan is erg belangrijk, net als in onze taal. Als ik bijvoorbeeld zeg “Poes de mooie”, dan weten jullie vast wel dat dat zinnetje niet helemaal klopt, want het moet natuurlijk zijn: “De mooie poes”. De volgorde van de muzieknootjes in de reeks is ook belangrijk. Jullie gaan dus naar de klankenreeksen luisteren en ondertussen mogen jullie de kleurplaat in jullie boekje inkleuren. Als het luisterfragment afgelopen is, dan krijgen jullie 12 vragen. De vragen gaan we meteen na het luisteren beantwoorden. Is iedereen er klaar voor? Dan gaan we beginnen.

(start experiment)

We gaan nu de vragen beantwoorden. De vragen staan in jullie boekje. Sla je bladzijde maar om en kijk naar vraag 1. Iedere vraag gaat zo: heb je deze klankenreeks in deze volgorde al eens eerder gehoord? Je hoeft alleen maar ja of nee te antwoorden, dat mag je omcirkelen in het boekje, en het gaat bij dit onderzoekje echt om wat jij gehoord hebt! Dus niet wat je buurman of buurvrouw gehoord heeft. Er zijn geen goede of foute antwoorden, dus werk gewoon lekker voor jezelf. Ik speel ieder melodietje maar één keer af, dus let goed op. Als er geen vragen zijn dan gaan we beginnen. Denk erom, de volgorde van de klanken in de reeks is belangrijk!

[You all received a booklet, you may write your name on it. In a few minutes you are going to listen to something. Last time you listened to funny words, but now you will listen to some melodies. They consist of three sounds, or notes, which together form a sequence. Actually, each sequence is like a sentence, because the order of the sounds in the sequence is really important, just like in our language. If I say: “cat the nice’, then you probably all know that that sentence is not correct, because the right answer is “the nice cat”. So, the order of the notes in the sequence is important too. You are going to listen to these sequences and meanwhile you may color a picture. As soon as the listening period is over, you will get 12 questions. We will answer them straight after listening. Is everybody ready? Let’s start.

(experiment starts)

We will now answer the questions. The questions are in your booklet. Turn the page and look at question 1. Every question goes as follows: did you hear this sequence in this order before? You only have to answer yes or no, which you can circle in your booklet. This experiment is all about what you heard! So what your neighbor answers is not important. There are no right or wrong answers, so just do your best. I play each melody only once, so pay close attention! If there are no questions we start. Remember, the order of the notes in the sequence is really important!]

Onderzoek Dyslexie

Universiteit Utrecht

Naam:..

School:..

Groep:...

Geboortedatum:...

Testdatum:...

Vragenlijstje

1. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

2. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

3. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

4. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

5. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

6. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

7. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

8. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

9. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

   Ja nee

10. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

    Ja nee

11. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

    Ja nee

12. Heb je deze klankenreeks in deze volgorde al eens eerder gehoord?

    Ja nee

Einde. Dankjewel voor het meedoen!