Language production and implicit statistical learning in typical development and children with acquired language disorders: an exploratory study

Xue Ting Joelle Mok[1], Siew Li Goh[1], James Douglas Saddy[2], Rosemary Varley[3], Vitor Zimmerer[3]

[1]KK Women's and Children's Hospital

[2]University of Reading

[3]University College London

Corresponding author:

Mok Xue Ting, Joelle

Speech Language Therapy Service

Rehabilitation Centre

KK Women's and Children's Hospital

100 Bukit Timah Road

Singapore 229899

E-mail: joelle.mok.xt@kkh.com.sg

Abstract

Statistical properties of language provide important cues for language learning and may be processed by domain-general cognitive systems. We investigated the relationship between implicit statistical learning (the unconscious detection of statistical regularities in input) and language production. Twenty typically developing (TD) children and nine children with acquired language disorders (ALD) (aged 6 to 18 years) took part in a Boston Cookie Theft picture description task. Using a computerized analysis, we investigated statistical properties, such as usage frequency of words and collocation strength of word combinations. Participants also completed a non-linguistic serial reaction time (SRT) task, which tested non-verbal, implicit statistical learning in the visual-motor modality. We determined age effects, and compared language production and SRT performance between both groups. Older TD children produced more connected language, more words, less frequent function words, more rare or novel combinations, and showed better statistical learning. Children with ALD produced less connected language, more weakly collocated combinations, displayed less lexical diversity and showed poorer statistical learning. Post-hoc analyses found correlations between statistical learning and statistical properties of spoken language. Given the rarity and heterogeneity of children with ALD, group size was small and the study should be considered exploratory. However, we note that results are compatible with the view that language production draws on statistical learning and that impairment of statistical learning can be related to language disorders.

Keywords: *acquired language disorder, typical development, paediatric, language production, sequential learning*

Introduction

Implicit statistical learning is the ability to detect probabilistic regularities from a given input without being consciously aware of these patterns (Christiansen & Chater, 2017; Plante & Gómez, 2018). It is fundamental to human behaviour and plays a role in many domains which involve sequencing of information or actions, such as music or identifying visual patterns in the environment (Turk-Browne et al., 2005). Evidence suggests that it is also relevant for language acquisition and use. In this study, we explored two aspects of statistical learning (1) whether, in typical language development, the effectiveness of statistical learning can be related to properties of the individual's spontaneous language production, including statistical properties, and (2) whether statistical learning can be disrupted in children with acquired language disorder (ALD). Also known as "childhood aphasia", ALD in children is an impairment of speech, language and/or communication following a neurological event, and after a pre-morbid period of typical language development (Dennis, 2010). ALDs occur due to neurological insults to the brain, such as a stroke, traumatic brain injury, tumour or infection. They often result in language deficits in both receptive and expressive language domains. Common characteristics of language deficits include difficulties with language comprehension (including inference making), learning new linguistic material, organising and structuring sentences, expressing more complex ideas, as well as features of word finding difficulties, such as perseverations and paraphasias. There are also often co-occurring cognitive deficits, such as problems with working memory, executive organisation, attention and processing speed (Gravel et al., 2007). The severity of the language impairment is largely variable based on the extent and location of the injury.

Statistical regularities of language input help explain the trajectory of language acquisition, and some propose that children's representation of grammar is to a substantial degree probabilistic

(Diessel, 2007; Erickson & Thiessen, 2015). Most of children's early constructions reflect what they hear frequently, and they are produced with little variation, suggesting that high-frequency collocations are learned in a very holistic fashion (Dąbrowska & Lieven, 2005). The child may produce combinations of frequently co-occurring words, such as 'I want', without necessarily having knowledge of their internal structure. After about 20 months, the child develops greater generative capacity, and analytic processing starts to emerge (Bannard & Lieven, 2012). However, these frequency effects remain important across the lifespan. In adults, more common combinations (e.g., 'I don't know why'; Arnon & Snider, 2010) are processed faster, suggesting that their representation is still holistic, or at least different from rare or novel combinations (Arnon & Snider, 2010; Conklin & Schmitt, 2008; Siyanova-Chanturia et al., 2017). More frequent words and word combinations are also more likely to be preserved in adults with aphasia (Bruns et al., 2019; Zimmerer et al., 2018).

The question arises whether the statistical learning network that is involved in language processing is specific to it, or domain general, and whether impairment of statistical learning can contribute to language impairment. Statistical learning has been observed in different stimulus modalities and experimental paradigms, including word segmentation (Saffran et al., 1996), visual pattern learning (Kidd, 2012), form-meaning mapping (Graf Estes et al., 2007), and serial reaction time (SRT) tasks (Robertson, 2007). Because of its involvement in several domains, it has been suggested that the statistical learning systems relevant for language are domain general at least to some degree (Christiansen & Chater, 2017; Conway, 2020). This view has been based on correlations between performance in non-linguistic statistical learning tasks, and language capacity (Erickson & Thiessen, 2015). Adults who more successfully learned statistical patterns in a visual, non-verbal artificial language learning task showed more sensitivity to word predictability in a speech perception task (Conway et al., 2010). Learning of statistical patterns in visual sequences was also associated with reading ability in neurotypical children and adults (Arciuli & Simpson, 2012).

Important evidence for a link between implicit statistical learning in non-verbal modalities and language production also comes from populations with acquired or developmental disorders. In adults with aphasia, particularly those with profound grammatical impairment, sequence learning in non-verbal artificial language tests was found to be impaired or substantially different (Christiansen et al., 2010; Zimmerer et al., 2014). Other evidence comes from studies using serial reaction time (SRT) tasks, which test implicit statistical learning in a visual-motor modality: Goschke et al. (2001) and Schuchard et al. (2017) concluded that statistical learning in adults with aphasia was generally intact, although learning effects in adults with aphasia were smaller than for controls. Correspondingly, Vadinova et al. (2020) found that implicit statistical learning mechanisms were present, but impaired, in aphasia, and found a correlation between implicit statistical learning mechanisms and the degree of syntactic impairment.

Further evidence comes from studies on the paediatric population. In an artificial language paradigm, children with developmental language disorder (DLD) demonstrated poorer implicit learning than controls (Evans et al., 2009), which is consistent with the results of a meta-analysis of eight SRT studies (Lum et al., 2014). A follow-up analysis including SRT, but also other statistical learning studies, came to the conclusion that children with DLD have impaired statistical learning, which may account for phonological and syntactic deficits (Obeid et al., 2016). In other language-linked developmental disorders such as developmental dyslexia, children have been found to demonstrate poorer sensitivity to transitional probability structures in both linguistic and non-linguistic stimuli (Gabay et al., 2015). Since then, new studies have found no significant associations in children with DLD (Lammertink et al., 2020) and dyslexia (van Witteloostujin et al., 2019).

Beyond heterogeneity within different paediatric populations, one issue which has made conclusions difficult is that studies use different empirical approaches while claiming to investigate the same subject. Implicit statistical learning has been investigated in different sensory modalities (auditory, visual, visual-motor), different methods of engaging with stimuli (e.g. grammaticality judgments, motor response), and with sequences that contain different types of dependencies (adjacent, non-adjacent, hierarchical) and complexity. It should not be assumed that all methods probe the same learning network (Conway, 2020).

Neuropsychological studies suggest an overlap in the neural bases of statistical learning in language processing and statistical learning in other domains. Findings in statistical learning studies can vary substantially depending on which experimental paradigm is employed. Studies using speech-segmentation paradigms have identified the left inferior frontal gyrus and left superior temporal gyrus (Karuza et al., 2013; McNealy et al., 2006; McNealy et al., 2010; Plante et al., 2017). Artificial grammar learning studies in different sensory modalities using probabilistic finite-state grammars have identified, beyond activation in sensory-modality specific areas, frontal cortical activation, including left inferior frontal gyrus, and subcortical activation, particularly basal ganglia (Conway & Pisoni, 2008; Newman-Norlund et al., 2006; Petersson et al., 2012).

A meta-analysis of 20 SRT studies using fMRI or PET found that only basal ganglia were significantly associated with learning (Janacsek et al., 2020), while in a lesion study, participants with damage to the basal ganglia displayed less evidence of SRT learning (Vakil et al., 2000).

While cortical areas such as left inferior frontal and superior temporal gyrus are frequently considered language areas, involvement of the basal ganglia in language has received less attention. Damage to the basal ganglia increases aphasia severity (Brunner et al., 1982), and has been associated with grammatical change in neurodegenerative conditions (Hinzen et al., 2018). In Parkinson's disease, there is report of a loss of ability to use common, holistic expressions, such as daily prayers or social formula (Van Lancker Sidtis, 2012; Van Lancker Sidtis et al., 2015). Ullman et al. (2020) suggest that the basal ganglia, together with the left inferior frontal gyrus, form the basis of procedural processing in language, and are therefore crucial for grammatical processing.

Current study

In the current study, we explored the relationship between implicit statistical learning and language production in a paediatric population, using recent concepts in language analysis and statistical learning paradigms. In TD children, we were interested in whether development of grammar is related to maturation of statistical learning. Previously, Meulemans and Van der Linden (1998) found no age-related differences between 6 year olds, 10 year olds and adults in a serial reaction time task, while Finn et al. (2016) found that the performance of 10 year olds on an artificial grammar learning task was comparable to that of adults. In contrast, Arciuli & Simpson (2011) found that visual statistical learning improved with age. With regard to learning across the lifespan, a large study conducted by Janacsek et al. (2012) comparing statistical learning performance in healthy individuals aged 4 to 85 found that learning was uniform until around 12 years, which was followed by a decrease and subsequently remaining uniform until about age 60, which was then followed by another decrease. We were interested not only in age-related effects on statistical learning, but also in correlations between statistical learning and properties of language production.

We also examined differences between TD children and children with ALD. The scarcity of this population results in it being largely understudied. While there has been research on adult clinical populations and developmental disorders in children, to our knowledge there has been no evidence on implicit statistical learning in children with ALD, or the relationship between implicit statistical learning and language production in this population. With evidence from other populations such as DLD, as well as from neuroimaging and lesion studies, linking non-verbal statistical learning and language processing, we found it plausible to hypothesise that statistical learning could be affected in children with ALD.

We employed a SRT task (Nissen & Bullemer, 1987) to test statistical learning in both groups. SRT studies test implicit statistical learning of adjacent dependencies in the visual-motor modality. In many SRT tasks, participants are first familiarized with a sequence order. Learning is tested by measuring reaction time (RT) increase when the order becomes random. In this study however, we used a Lindenmayer grammar (Geambasu et al., 2020; Prusinkiewicz & Hanan, 1989) which generates strings with stimulus transitions which can be defined probabilistically, with some being more predictable than others. Rather than contrasting structured with random trail blocks, we examined the degree to which RT patterns mirrored the statistical predictability of the generated structure. A learning participant should display lower RTs where transitions between stimuli are more predictable. Better 'fits' of RT data to the statistical properties of the strings were interpreted as indicators of successful statistical learning. The design allows testing of learning of different complexities: the probability of a symbol appearing in any given position, the probability of one symbol given the preceding symbol (first order), or the probability of a symbol given the preceding two symbols (second order). We established models to test learning of each level of complexity and hypothesized that groups could differ with regards to what they learn (which model fits best), and how well they learn (how well the model fits to the RT data).

We investigated statistical properties of language production using the Frequency in Language Analysis Tool (FLAT), which has previously been applied in investigations of adults with aphasia (Bruns et al., 2019; Zimmerer et al., 2020; Zimmerer et al., 2018). The FLAT analyses orthographic transcripts by extracting each word, bigram (two-word combination) and trigram (three-word combination) and determines their usage frequency using the 10 million word spoken corpus of the British National Corpus (BNC; The British National Corpus, 2007). The BNC reflects typical language use as it is based on utterances collected from a range of speakers with different backgrounds. Based on frequency, FLAT also computes the collocation strength of each word combination, which indicates how strongly words co-occur relative to their usage frequency. Similar to the SRT task, this measure therefore describes the statistical association between adjacent units. In adults with aphasia, increased values were interpreted as a sign of language impairment as individuals relied more on familiar forms and word combinations that are easier to process. It is, to our knowledge, the only software that allows analysis of word combinations in such a manner. However, it was designed primarily for adult language and the BNC is an adult corpus. We think that for children, higher values may reflect successful acquisition and alignment to adult patterns as children pick up the statistical properties of their language. The FLAT software also provides word counts, measures for word and combination diversity (type/token ratios; TTR), the proportion of content words in each sample, and a measure of fluency, so we also examined the effect of age and group on these variables.

Research questions and hypotheses

Our main research questions were:

1. Are there any age effects of general production variables (word count, fluency, content word ratio, type/token ratio), as well as frequency and collocation strength of produced language forms?
2. How does language production differ between TD children and children with ALD?
3. Are there any age effects in implicit statistical learning abilities?
4. How does implicit statistical learning differ between TD children and children with ALD?
5. Is there a relationship between language production and implicit statistical learning?

Our hypotheses were nondirectional as we expected changes in age and between groups, but had insufficient grounds to predict a direction given that we used novel methods and there is very little research on children with ALD.

Methods

Participants

This study was approved by the University College London Research Ethics Committee and the SingHealth Centralised Institutional Review Board. It took place in Singapore and informed consent for participation in the study was obtained from all participants.

Twenty Singaporean TD children were recruited via a recruitment poster. They had no reported history or diagnosis of speech or language impairment, including dyslexia. There were eight males and 12 females, aged 6 to 18 years (mean = 11.5; standard deviation (*SD*) = 2.56), and 19 were right-handed and one left-handed. All had English as their first language. 17 children were bilinguals who spoke English and Mandarin Chinese, one of which also spoke basic Malay in addition to English and Mandarin Chinese. The remaining three children were monolinguals and only spoke English.

Nine Singaporean children with ALD secondary to neurological events were recruited through a Speech-Language Pathology neurology clinic in a hospital. The group consisted of six males and three females aged 7 to 15 years (mean = 11.6; *SD* = 4.04). Six were right-handed and three were left-handed. English was the first language for all but one participant, P4, who spoke English as a sequential bilingual, with Mandarin Chinese being her first language. There were a total of six bilingual children (three spoke English and Mandarin Chinese, one spoke English and Malay and the remaining child spoke English and Tagalog). One of the bilingual children also spoke very basic dialect (Teochew). There was a total of three monolingual children, one of whom spoke very basic Mandarin Chinese. The groups did not differ significantly by age, $t(27) = 0.401$, $p = 0.691$, $d = 0.15$.

Four of the children with ALD suffered from a traumatic brain injury (TBI), three had cancer and two had ruptured arteriovenous malformations (AVMs). The age of onset of the neurological events ranged from one year to approximately 14 years. As at the time of the study being conducted, the post-onset duration ranged from less than a month to seven years. The participants' language abilities had been assessed prior by Speech-Language Pathologists via various formal and/or informal assessment tools which may include formal assessments such as the Pediatric Test of Brain Injury (Hotz et al., 2010) and Clinical Evaluation of Language Fundamentals (4^{th} edition; Semel et al., 2003), and/or informal assessments (e.g. picture description tasks, narrative language samples). Information on participant characteristics for the ALD group can be found in Table 1.

Participant ID	Gender	Age (years)	Right (R) / Left (L) Handed	Primary Neurological Diagnosis	Age at Onset	Post-Onset Duration (as on Test Date)	Severity of ALD	Additional remarks regarding ALD
P1	M	14	R	Traumatic Brain Injury (TBI; Left Temporo-Parietal Haematoma)	14 years 4 months	Less than 1 month	Moderate	Non-Fluent Aphasia, with cognitive difficulties
P2	M	9	R	TBI (Right Subdural/ Subarachnoid Haemorrhage)	6 years 5 months	1 year 5 months	Moderate	Reduced receptive and expressive language skills, with cognitive difficulties
P3	F	10	R	TBI (Diffuse Axonal Brain Injury)	8 years 2 months	2 years 6 months	Moderate	Reduced expressive language skills, with cognitive difficulties
P4	F	13	L	Cerebellar Medullablastoma	6 years 11 months	6 years 11 months	Moderate	Reduced receptive and expressive language skills, with reduced memory

P5	M	15	L	TBI (Left Subdural Haemorrhage)	15 years 3 months	6 months	Mild	Reduced receptive and expressive language skills, with reduced memory (Mild Dysarthria also present)
P6	M	8	R	Ruptured AVM (Right Parietal Lobe)	7 years 8 months	1 year 3 months	Mild	Mildly reduced receptive and expressive language skills, with reduced attention, motivation and cognitive difficulties
P7	M	7	R	Anaplastic Astrocytoma (Grade III) Brain Tumor in Left Parieto-Occipital Lobe	1 year	6 years	Moderate	Reduced receptive and expressive skills
P8	M	10	R	Posterior Reversible Encephalopathy Syndrome secondary to	4 years	7 years	Mild	Functional receptive skills and mildly reduced expressive skills

				Acute Lymphoblastic Leukemia				
P9	F	13	L	Ruptured AVM (Left Sylvian Fissure)	6 years	7 years	Moderate	Fair receptive skills and moderately reduced expressive language skills

Table 1: *Background information for Children with ALD*

Stimuli and procedure

Participants completed two tasks: spoken narrative production (Boston Cookie Theft picture description, Goodglass et al., 2001) and a SRT task. Methods and protocols in our study were simple, short, and suitable for a paediatric population. Due to experimenter error, five TD children completed the SRT task first, while all other participants completed the picture description task first.

Boston Cookie Theft picture description

The Boston Cookie Theft picture description from the Boston Diagnostic Aphasia Examination is a widely used tool to elicit discourse in individuals with disorders of language function or aphasic syndromes (Goodglass et al., 2001). It depicts a mother washing dishes at an overflowing sink, while two children are attempting to get cookies from a cookie jar up on a shelf, with the boy about to fall from a stool. It was chosen as it is a short, simple task that caters to our ALD group, as well as younger participants in both groups.

The experimenter showed the picture to the participant and gave the prompt: ‘I am going to show you a picture. Can you tell me what’s happening in the picture?’ Descriptions were audio-recorded. At the end of the description, the experimenter asked the participant if they had anything else to add. Additional prompts were only given if the participant showed obvious difficulty in production. In these cases, the experimenter would ask the participant a prompting question similar to the initial question (e.g. ‘What’s happening?’, ‘Anything else?’). Spoken language samples were subsequently orthographically transcribed and processed by the FLAT according to transcription guidelines described in Zimmerer et al. (2018).

Table 2 describes the measures that were obtained. They include frequency in relation to our questions about statistical features of language production, and other measures which investigate quantity of output, fluency and lexical diversity. Combination ratio (i.e. trigram count divided by word count) was used as a measure of connected language. For word combinations, we chose collocation strength (*t*-scores), which quantifies how often words appear together considering their individual usage frequency. Differences in collocation strength therefore more likely reflect real differences at the level of combinations instead of single words. We focused on bigrams rather than trigrams, as the former yielded larger effect sizes in comparisons between clinical and control groups (Bruns et al., 2019; Zimmerer et al., 2020; Zimmerer et al., 2018), and included bigrams with a frequency of one or higher. For frequency and collocation strength measures we averaged values for each type, not token, produced by a participant. The reason for this decision is that token averages are confounded by repetitions within samples, which we captured independently by determining type/token ratios (TTR) for words and bigrams.

While we excluded variables with a frequency of zero from the collocation strength measure, the production of these is useful for profiling language production. As in previous FLAT studies, we computed the proportion of combinations that occur in the BNC ('BNC ratio'). This variable too is an indicator of how much an individual relies on familiar, as opposed to rare or novel, language forms. A higher value indicates reliance on more familiar combinations. Bigrams that are ungrammatical were excluded from analysis, with the exception of missing third person marking in verbs (e.g. 'the water fall'), which we considered dialectically correct.

Measure	Description
Word count	Quantity of verbal output.
Combination ratio	Measure of connected language (trigram count divided by word count).
Content word ratio	Proportion of content (open class) words as part of the total word count.
Content word frequency	Usage frequency of content (open class) words, per million words.
Function word frequency	Usage frequency of function (closed class) words, per million words.
Bigram t-score	Collocation strength of combinations with frequency > 0.
Bigram BNC ratio	Proportion of bigrams that occur in the BNC, i.e. have a frequency > 0.
Word/Bigram TTR	Measures of diversity of linguistic forms.

Table 2: *Measures determined by automated FLAT analysis. Frequency, t-scores and BNC ratios measured based on the spoken subsection of the BNC.*

SRT task

The SRT task was written using DMDX (Forster & Forster, 2003). SRT stimuli were dots that appeared on the left (‘A’) or right (‘B’) side of a computer screen. The screen was split into sections by a large ‘X’ (Figure 1). Stimulus A was mapped to the left button on the right hand pad of a Logitech Precision Gamepad, stimulus B to the right button. The stimuli remained on screen until the participant pressed a response button. Button selection and RT was recorded. Participants were shown a visual representation of the button-stimulus mapping and asked to press the corresponding button as quickly as possible when the stimulus appeared. The interstimulus interval was 500ms, during which a blank grey screen was presented. The experiment contained seven trial blocks with a self-timed break between blocks.

Each block contained 55 stimuli and lasted about two minutes. Of the 55 stimuli, 21 were stimulus A and 34 stimulus B. In the first block, the order was random. In the other blocks the order was determined by a Fibonnaci grammar. The Fibonacci grammar belongs to the family of Lindenmayer systems (Prusinkiewicz & Hanan, 1989). Typically, it contains two symbols A and B, which were matched to stimuli A and B respectively, and consists of two rewrite rules: A -> B, B -> BA. Starting with the symbol A, the grammar generates the following sequences with each generation: A, B, BA, BAB, BABBA, BABBABAB, etc. (the length increase follows the Fibonacci number series). The structure for trial blocks 2-7 was generated by randomly extracting a section of 55 stimuli from a 26th generation Fibonacci sequence (75025 symbols long).

In longer sequences, statistical regularities quickly emerge (Appendix A). These regularities are the basis for determining which information about the sequence structure a participant extracted, and how successfully it was implemented. The simplest statistical model focuses only on a single symbol. After some exposure, it can learn that the probability of any symbol being A is ca. 38.1%, and the probability of it being B is ca. 61.8%. The next complex model is a bigram (or first order) model. A bigram model can learn that if a symbol is A, it will always be followed by B (i.e. the representation of probability during exposure approaches 100%), and that if the symbol is B, it is followed by A with a probability of ca. 61.8%. A trigram (second order) model can also learn that a bigram sequence AB is followed by A with a ca. 38.1% probability, and that the bigram BB is followed by A with a probability approaching 100%.

As the statistical models become more complex, they become able to predict the next symbol with greater certainty on the basis of preceding symbols. Since RT can be expected to be faster the closer the prediction gets to certainty, RT patterns can be modelled to determine the most likely probabilistic representation of a given participant, and how successful the model is. We therefore determined which model was the best fit, and compared groups on the fit of the model. RT patterns mirroring more complex representations (bigrams, trigrams) can be seen as evidence for better learning. To our knowledge, this novel method has not been used with children. However, a pilot study (Zimmerer et al., 2013) suggested that for adults, trigram models were the best representation of statistical learning.

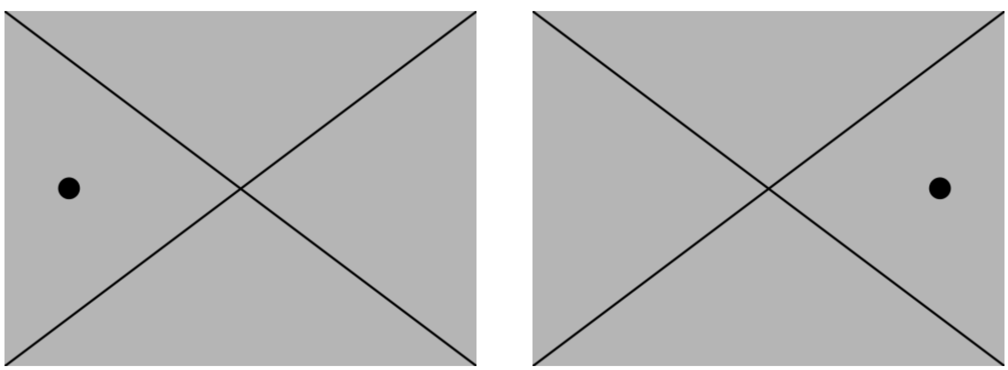

Figure 1: *SRT stimuli A (left) and B (right)*

Results

Shapiro-Wilk tests for normality were carried out on all variables. All data were not normally distributed, except for function word frequency in the TD group and word count in the ALD group. We categorized effect sizes according to Cohen's (1988) criteria for interpretation (small effect: $d \geq 0.2$; intermediate effect: $d \geq 0.5$; large effect: $d \geq 0.8$). We associated each outcome variable with a different property of language, and therefore a separate hypothesis, with the exception of bigram *t*-score and bigram BNC ratio which both tested familiarity of word combinations. In the latter case one could argue for Bonferroni adjusted thresholds, and we report them to signal that the two variables tested the same hypothesis. However, we advise caution since Bonferroni adjustments have been criticized to introduce an 'unacceptable' risk of Type II errors especially in exploratory studies with limited *n*, in which effect sizes may be more informative than *p* value statistics (Nakagawa, 2004; Perneger, 1998).

Table 3 displays results from both groups. For each variable we report the effect of both age and group separately, followed by a rank analysis of covariance (a variant of the ANCOVA for non-parametric data; Quade, 1967), in which we residualized the effect of group over age. Because groups did not differ significantly on age, the ranked ANCOVA serves as a way of noise-reduction. The effects of group and group residualized over age were similar, which suggest that variances explained by age and group do not overlap. For word and bigram TTRs we additionally entered word and bigram count (respectively) as covariates, since TTRs can be confounded by differences in sample size. We chose this solution over alternatives to TTR (such as *vocd*; Harris Wright et al., 2003) because samples were overall short and for bigrams, such alternatives have not been developed.

Age had a significant effect on several variables as older children had a higher word count and combination ratio, produced function words with lower frequency and had a lower bigram BNC ratio, i.e. older children produced more words, had more connected output, used rarer function words and more combinations which were rare or novel. Figure 2 shows that the relationship between age and these variables was driven by patterns within the TD group.

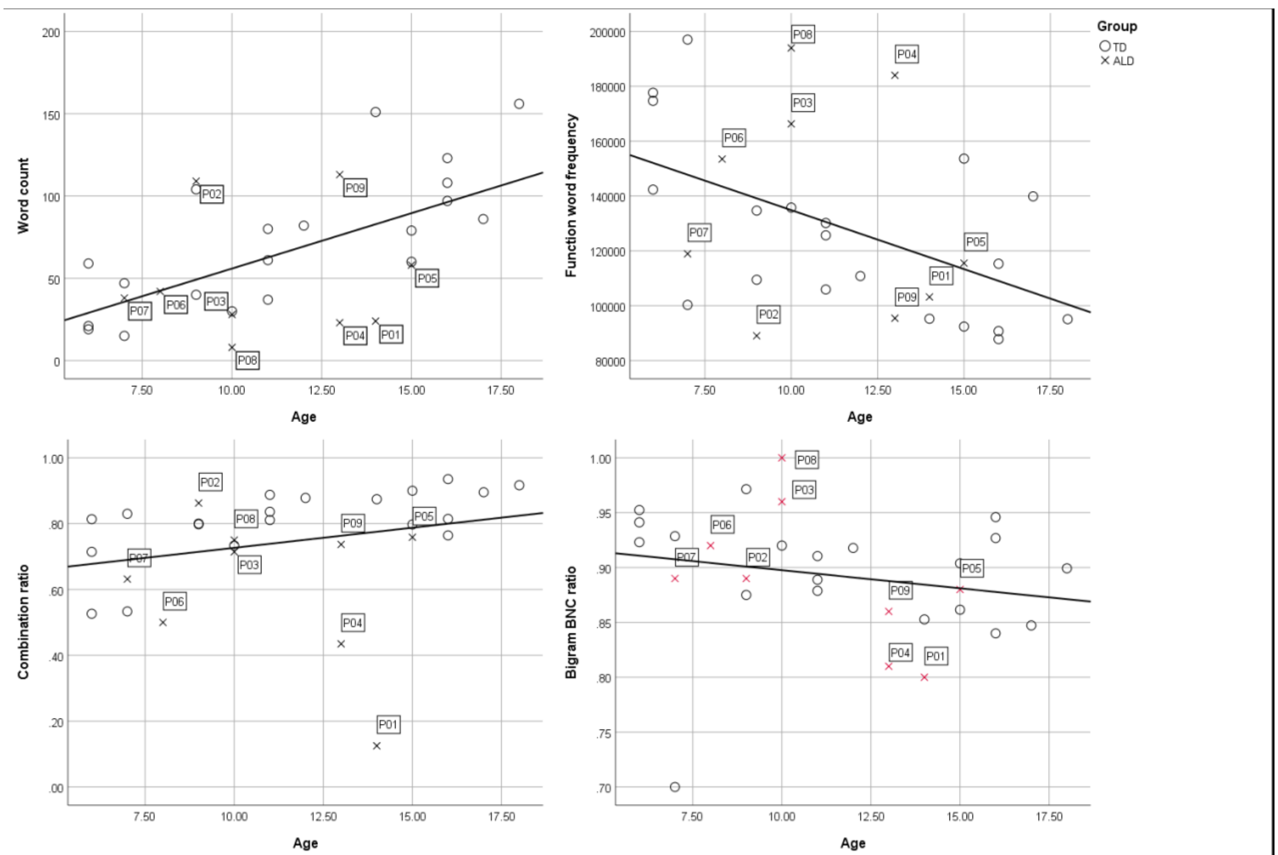

Figure 2: *The effect of age on word count, function word frequency, combination ratio (a measure of connected language) and Bigram BNC ration (the proportion of bigrams with a frequency > 0). The linear best fit line is based on data from both ALD and TD groups. LOESS curves for each group, which are more sensitive to outliers, show that age effects were driven by distributions within the TD group.*

After residualizing over age, TD children and children with ALD differed significantly on a number of variables. Children with ALD had a lower combination ratio, lower bigram *t*-scores and word TTR. The differences in *t*-scores would not be significant under a Bonferroni-adjusted threshold ($p = 0.025$; however, note our reservations above). Effect sizes were large. Note however that because of the smaller size of the ALD group, other large effects were not statistically significant, namely differences for content word ratio and bigram TTR; children with ALD produced more content words in relation to function words and showed less diversity for word combinations.

Variable	**TD (IQR)**	**ALD (IQR)**	**Age effect**	**Group effect**	**Group residualized over Age**
*Word count	70 (64.5)	38 (60)	$t(27) = 3.758$, $p = 0.001$, $\beta = 0.59$	$t(27) = 1.390$, $p = 0.176$, $d = 0.57$	$t(27) = 1.454$, $p = 0.157$, $d = 0.67$
*,†Combination ratio	0.81 (0.11)	0.75 (0.26)	$t(27) = 2.330$, $p = 0.028$, $\beta = 0.41$	$t(27) = 3.325$, $p = 0.003$, $d = 1.28$	$t(27) = 2.728$, $p = 0.011$, $d = 1.26$
Content word ratio	0.39 (0.05)	0.43 (0.13)	$t(27) = 1.123$, $p = 0.271$, $\beta = 0.21$	$t(27) = 1.926$, $p = 0.065$, $d = 0.81$	$t(27) = 1.861$, $p = 0.074$, $d = 0.82$

Content word frequency	341 (284) per million	316 (474) per million	$t(27) = 0.291$, $p = 0.773$, $\beta = 0.06$	$t(27) = 0.559$, $p = 0.581$, $d = 0.25$	$t(27) = 0.534$, $p = 0.597$, $d = 0.24$
*Function word frequency	120 (45) per million	119 (76) per million	$t(27) = -2.804$, $p = 0.009$, $\beta = -0.48$	$t(27) = 0.606$, $p = 0.550$, $d = -0.25$	$t(27) = -0.447$, $p = 0.658$, $d = -0.20$
†Bigram t-score	19.63 (9.4)	15.57 (12.26)	$t(27) = 0.563$, $p = 0.578$, $\beta = 0.11$	$t(27) = 2.097$, $p = 0.046$, $d = 0.86$	$t(27) = 2.051$, $p = 0.05$, $d = 0.84$
*Bigram BNC ratio	0.91 (0.06)	0.89 (0.11)	$t(27) = -2.083$, $p = 0.047$, $\beta = -0.37$	$t(27) = 0.441$, $p = 0.662$, $d = 0.19$	$t(27) = 0.656$, $p = 0.517$, $d = 0.27$
†Word TTR	0.63 (0.19)	0.61 (0.17)	$t(27) = -0.114$, $p = 0.91$, $\beta = -0.02$	$t(27) = 2.608$, $p = 0.015$, $d = 1.21$	$t(27) = 2.596$, $p = 0.015$, $d = 1.20$
Bigram TTR	0.94 (0.07)	0.85 (0.19)	$t(27) = 0.268$, $p = 0.791$, $\beta = 0.58$	$t(27) = 1.730$, $p = 0.095$, $d = 0.81$	$t(27) = 1.770$, $p = 0.088$, $d = 0.83$

Table 3: *Medians and interquartile ranges (IQRs) of FLAT variables for each group, as well as inferential measures of group difference and age effect. Because the majority of distributions were non-parametric, we ranked all variables to maintain comparability. We calculated the effects of age and group separately, as well as the effect of group residualized over age to*

determine how much the variances explained by the variables are shared. Where the effect of group and group residualized over age are similar, there is little shared variance. For word and bigram TTRs, we additionally residualized over word and bigram count (respectively).

**significant age effect, †significant residualized group difference; α-level = 0.05.*

For RT analysis we excluded wrong button presses and RT values which we identified as outliers i.e., values higher or lower than 2 SDs from the individual's mean. In total, 8.86% of responses were excluded after applying these criteria (6.86% in the TD group, 10.36% in the ALD group). To determine the complexity of statistical learning we assigned each symbol in the SRT sequence probabilities based on single item, bigram and trigram models (Appendix A). For example, the symbol A in a trigram BBA has a probability of 38.1% in a single symbol model (since As occur 38.1% of the time), a probability of 68.1% in a bigram model (since A follows B 68.1% of the time) and a probability of 100% in a trigram model (since A always follows BB). We then conducted for each model a linear regression for each participant for blocks 2-7 with the probabilities as dependent and RTs as the independent variable. Because we assumed a negative correlation between probabilities and RTs (i.e. fastest for 100% certainty), negative *t* values indicated strong evidence that a given model represented the statistical representations of an individual. While we focussed on response times, we also checked whether statistical models could predict whether trials were more likely to be excluded. For each group, we computed the percentage of accurate responses within 2 SDs of the individual's average RT for a given trial stimulus and used linear regressions with single symbol, bigram and trigram probabilities as independent, and the percentage of accurate responses as the dependent variable.

In the ALD group, we excluded participant P7 because he seemed very distracted during the task and displayed an abnormally high error rate (54.5%). Shapiro-Wilk tests for normality found non-parametric distributions for mean RT in the TD group and number of errors in the ALD group. Distributions were parametric for fit to single symbol, bigram and trigram models. For greater comparability, we again used non-parametric tests on all variables. Success in learning the pattern was measured using two variables, namely bigram and trigram models. We therefore note that Bonferroni adjustments can be applied in comparisons of learning, however, see above for reservations regarding use of Bonferroni in exploratory studies.

Table 4 shows SRT data from both groups, as well as inferential group comparisons and investigations of age effects. The data show that the bigram model was the best fit for both groups, suggesting that participants most reliably learned the probability of a stimulus based on the previous stimulus. There was no substantial overlap between variances explained by age and group.

Older children had faster responses and better statistical learning. Figure 3 shows the relationship between age and SRT learning in both groups, and suggests that the age effect was driven by TD children.

There were large differences between groups as children with ALD were slower and showed worse bigram learning. Both effects were large, and the latter withstands correction for multiple comparisons (see above; adjusted threshold $p = 0.025$).

We were interested in whether the slower RT of children with ALD was an effect of less efficient learning, or of possible motor coordination impairment as the result of neurological damage. We compared average RTs from block 1 (in which stimulus order was randomized). TD children's average RT was 468ms (SD = 198), while the average RT for children with ALD was 530ms (SD = 135). Groups did not differ significantly in block 1, $t(27) = -.856$, $p = .40$, $d = .31$, which suggests that differences in learning capacity contributed to RT differences in the other trial blocks.

We further examined learning within each group in each trial block. The aim was to explore whether group comparisons revealed differences in the ability to attain a certain level of statistical representation, or were rather caused by differences in maintaining performance. We considered that the ALD group could have taken longer to reach a performance peak, or that performance dropped off earlier because of difficulties with maintaining attention. Figure 4 shows that learning performance in the ALD never reaches the level of the TD group. Further, performance in the ALD group peaks in Block 5 and then drops off, while the TD group performance peaks in Block 6.

Variable	**TD (IQR)**	**ALD (IQR)**	**Age effect**	**Group effect**	**Group residualized over Age**
*,†Mean RT	332ms (237)	528ms (147)	$t(26) = -4.074$, $p < 0.001$, $\beta = -0.624$	$t(26) = -2.043$, $p = 0.051$, $d = -0.53$	$t(26) = -2.627$, $p = 0.014$, $d = -1.23$
Error count	10 (0.95)	9.5 (23.25)	$t(26) = -3.400$, $p = 0.002$, $\beta = -0.555$	$t(26) = -0.451$, $p = 0.655$, $d = 0.11$	$t(26) = -0.466$, $p = 0.645$, $d = -0.20$

Single symbol learning	-0.35 (4.75)	-0.78 (3.03)	$t(26) = 0.268$, $p = 0.791$, $\beta = 0.052$	$t(26) = -0.100$, $p = 0.921$, $d = -0.06$	$t(26) = -0.106$, $p = 0.916$, $d = -0.04$
*,†Bigram learning	-9.19 (6.05)	-3.31 (4.36)	$t(26) = -2.678$, $p = 0.013$, $\beta = -0.465$	$t(26) = -3.699$, $p = 0.001$, $d = -1.89$	$t(26) = -4.378$, $p < 0.001$, $d = -1.89$
Trigram learning	1.53 (3.26)	0.82 (2.42)	$t(26) = -0.329$, $p = 0.744$, $\beta = -0.329$	$t(26) = 1.453$, $p = 0.158$, $d = -0.71$	$t(26) = 1.465$, $p = 0.155$, $d = -0.59$

Table 4: *Medians and IQRs of SRT variables for each group, as well as inferential measures of group difference and age effect. Negative values for SRT models indicate a better fit, and better learning of the respective representation. We calculated the effects of age and group separately, as well as the effect of group residualized over age to determine how much the variances explained by the variables are shared. Where the effect of group and group residualized over age are similar, there is little shared variance.*

**significant age effect, †significant residualized group difference; α-level = 0.05.*

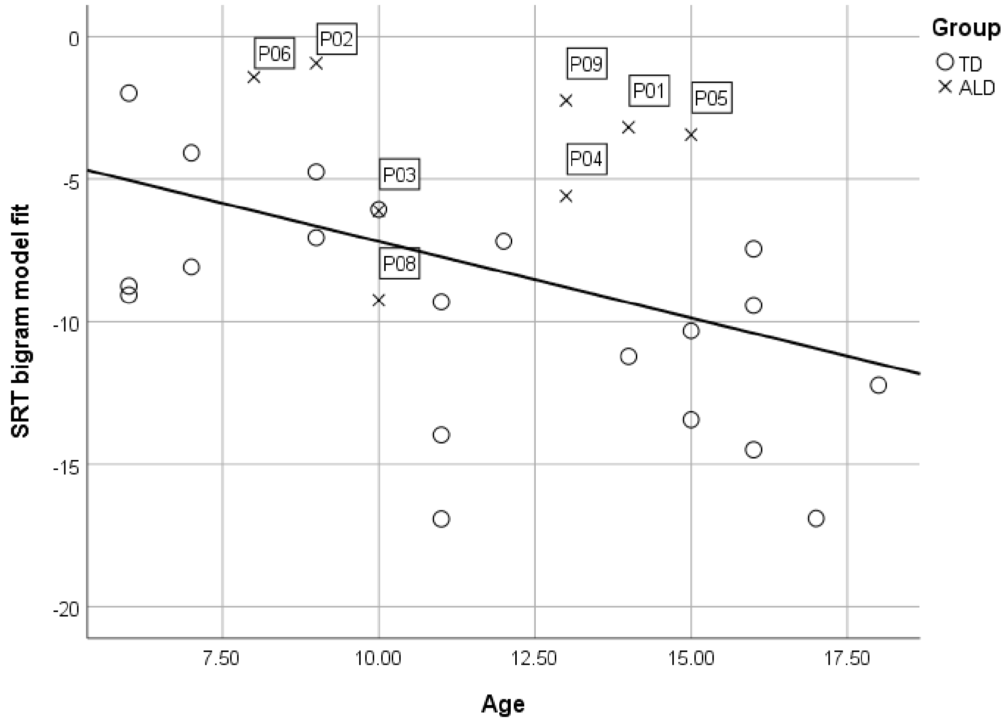

Figure 3: *Relationship between age and SRT learning (bigram model) in both participant groups. Negative values imply better learning. The linear best fit line is based on data from both ALD and TD groups.*

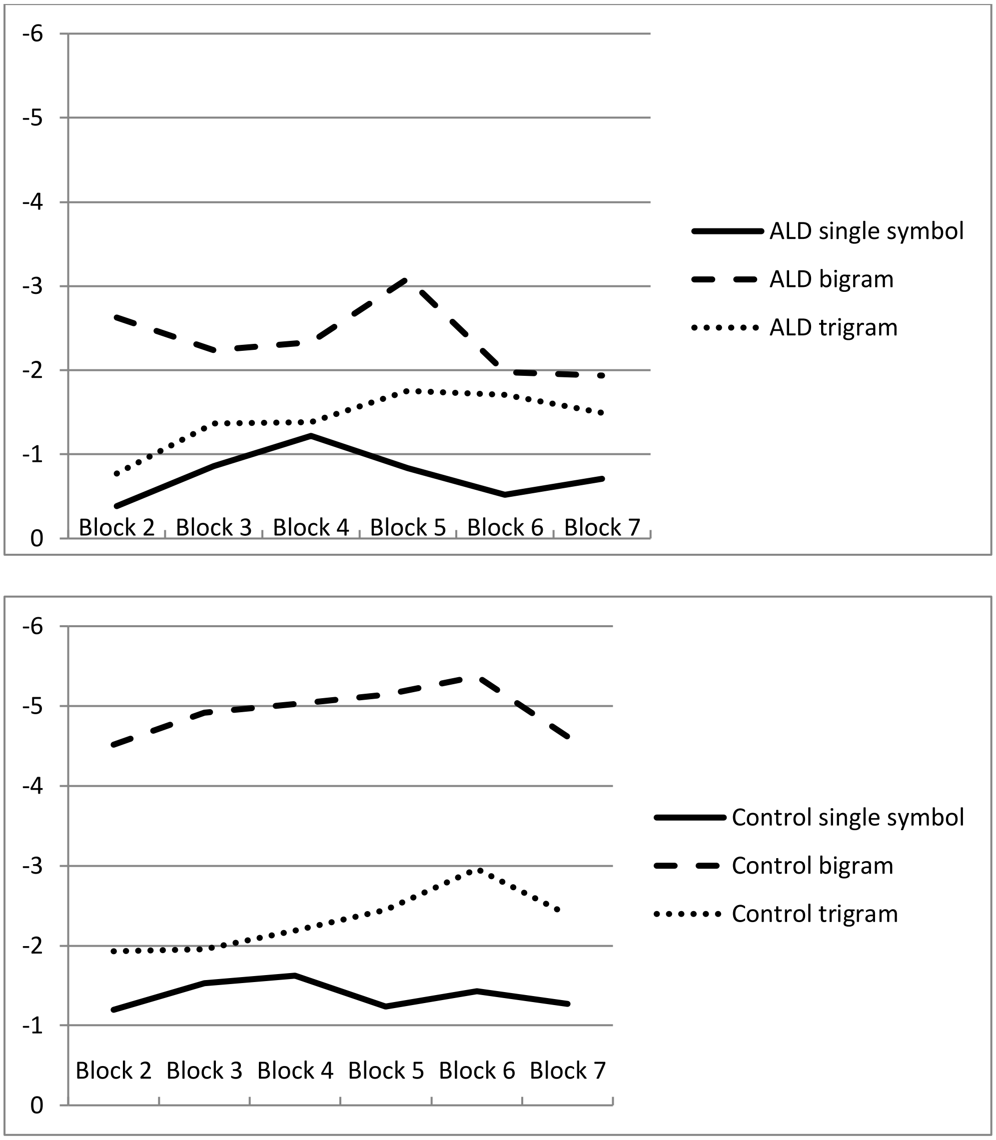

Figure 4: *SRT learning over trial blocks of each 55 stimuli, divided by group. The x-axis denotes trial blocks (Block 1 was a training block and is not shown). The y-axis denotes fit to learning models of different complexities (single symbol, bigram, trigram). The axis is inverted and lower values reflect better fit to the learning model.*

Accuracy data further suggest that the bigram representation is the most likely predictor of participant performance. Using the percentage of accurate trials within the group as independent variable in a linear regression, it was the best fit for controls, $\beta = .45$, $p < .001$, and children with ALD, $\beta = .331$, $p < .001$. In both groups, higher certainty was associated with greater accuracy.

After finding that age affected both language variables and statistical learning, and that groups differed at both levels, we looked for more direct relationships between language production and SRT. We selected the bigram model as it best captured SRT performance and correlated individual's *t* values with their language variables, using the Kendall's rank correlation. Note that this analysis must be considered a post-hoc exploration since we made this selection after the previous analysis step.

As a joint group, SRT bigram learning correlated significantly only with one variable, combination ratio, $\tau = -0.358$, $p = 0.008$, meaning that children who showed better learning produced more connected language. Split into separate groups, TD children who showed better learning had a higher bigram BNC ratio, $\tau = 0.432$, $p = 0.008$, i.e. produced more rare or novel combinations. Children with ALD who showed better SRT learning produced lower-frequency function words, $\tau = -0.714$, $p = 0.013$, and produced more words $\tau = 0.571$, $p = 0.048$.

Discussion

Language acquisition and processing may be supported by domain-general statistical learning. In this exploratory study, we investigated statistical properties of language production and statistical learning capacities in TD children and children with ALD. We were interested in age effects suggesting a developmental trajectory, and in whether performance in both domains is correlated. We also hypothesized that children with ALD would not only differ in language production, but also in statistical learning. We used a SRT paradigm which tested implicit learning of adjacent statistical dependencies in the visual motor modality. Our results touch upon a range of issues concerning child development and language, some of which, especially with regards to ALD, have barely been examined.

With regard to language production, we found that older TD children produced more connected language, less frequent function words and more word combinations which were rare or novel. The results are consistent with greater grammatical expressivity in older children and an increased use of generative processes. Children with ALD produced less connected language. Their samples also showed less lexical diversity. While both groups produced similar proportions of rare or novel combinations, when children with ALD produced combinations that occur in the BNC, the combinations were more weakly collocated. This points towards a decrease in use of holistic forms. While there has been much attention to generative aspects of language and its ability to produce new utterances, common collocations often have important pragmatic functions (e.g. conversational formulas; Wray, 2012), and a disrupted ability to acquire or process these may have a substantial effect on the ability to communicate effectively.

With regards to statistical learning, we found that children most likely extracted bigram (first order) probabilities from the SRT stimuli. A correlation between age and learning model fit in our SRT design provides corroborating evidence that statistical learning improves during childhood and adolescence (Arciuli & Simpson, 2011). If the capacities relevant for learning in this non-verbal, visual motor task are used in language, this would mean that language maturation would not only be driven by the emergence of more abstract knowledge (e.g. phrase structure rules), but also by the improved ability to track statistical regularities. Comparisons across blocks suggest that children quickly extracted statistical regularities and maintained their representations until the end of the experiment, as witnessed by the fit to learning models plateauing quickly. This very rapid learning has been observed in previous SRT experiments (Kóbor et al., 2018; Simor et al., 2019), however, one should not conclude that representations of probabilities could not have changed (e.g. to a more complex trigram/second order model) if the experiment had been longer.

Children with ALD also most likely learned bigram representations, but displayed weaker statistical learning in the SRT task, reflected in SRT data fitting less to the learning models, and in a performance drop-off earlier in the experiment. Overall poorer learning was predicted and our SRT results fit with impaired learning in SRT task performance with children with specific language impairment or DLD. However, in this population, recent evidence has also shown that learning can be achieved at a comparable rate to controls when given more time (Lum et al., 2014). Given the importance of implicit statistical learning and visual motor learning in skill acquisition, SRT and similar paradigms could be important tools for determining strengths and weaknesses of children with neurological damage, and supporting their development, for example in education.

A post-hoc analysis found some correlations between statistical learning and properties of language production, supporting the general notion that statistical learning capacity, even in a non-verbal domain, is involved in language processing. However, the picture is not clear since correlations in the TD group, where better learners produced more rare and novel combinations, differed substantially from correlations in the ALD group, where better learners produced rarer function words and more words. Correlational analyses are particularly vulnerable to heterogeneity within groups, especially in children with ALD. The general impression from the data is that in development, better statistical learning is associated with the ability to produce rarer forms, perhaps reflecting the individual's capacity to acquire more than just the strongest patterns, but also form novel combinations.

The debate about the involvement of statistical learning in language is complex. Statistical learning can involve implicit or explicit learning paradigms (where the latter involves feedback or conscious choices), different modalities, different structures, and varying levels of attention and conscious processing. Different statistical learning can be at work dependent on task demand (Conway, 2020). Language, too, involves different types of structures and probabilistic relationships, such as the relationship between speech sounds, between lexical and semantic representations, or between words, word categories and phrases. More research is required to determine which aspects of statistical learning are related to which aspects of language (Erickson & Thiessen, 2015).

Our study is limited by its group size and within-group heterogeneity, especially in the ALD group with regards to type of neurological damage, time post-onset and the effects of bilingualism, which is very common in Singapore. The effect of task order (SRT and language elicitation) also could not be determined. These limitations not only mean that results call for replication, but also that more complex relationships, such as the interaction between neurology, the trajectory of language acquisition and statistical learning could not be investigated.

There is a general worry about the impact of comorbidities in SRT studies. Perceptual and motor limitations may affect RTs and reduce the reliability in measuring learning (Ostegaard, 1998: West et al., 2018). In our study, such confounds would be more expected in the ALD group, however, data suggest that, with the exception of one participant whose data we excluded from this part of the study, our group performed well in the SRT task. Children with ALD did not make significantly more errors than TD controls, and we found no RT differences between groups in our first, randomized trial block. This suggests that in our study, RT differences in the latter blocks reflect differences in learning as detected by our models. However, to address confounding factors in SRT tasks, Bogaerts et al. (2020) argue for the need of separate control tasks.

With these limitations in mind we suggest a careful conclusion: The comparison between TD and ALD is also consistent with a proposed relationship between the two domains. However, the effect of ALD in childhood on statistical properties of language appears to be very different from the effect of aphasia in adults. In adults with aphasia, production is marked by more frequent and strongly collocated forms (Bruns et al., 2019; Zimmerer et al., 2018), indicating that speakers are restricted to more formulaic expressions (Van Lancker Sidtis & Postman, 2006). Children with ALD on the other hand produced combinations which were more weakly collocated. This may reflect difficulties acquiring language formulas, which may be related to impaired statistical learning.

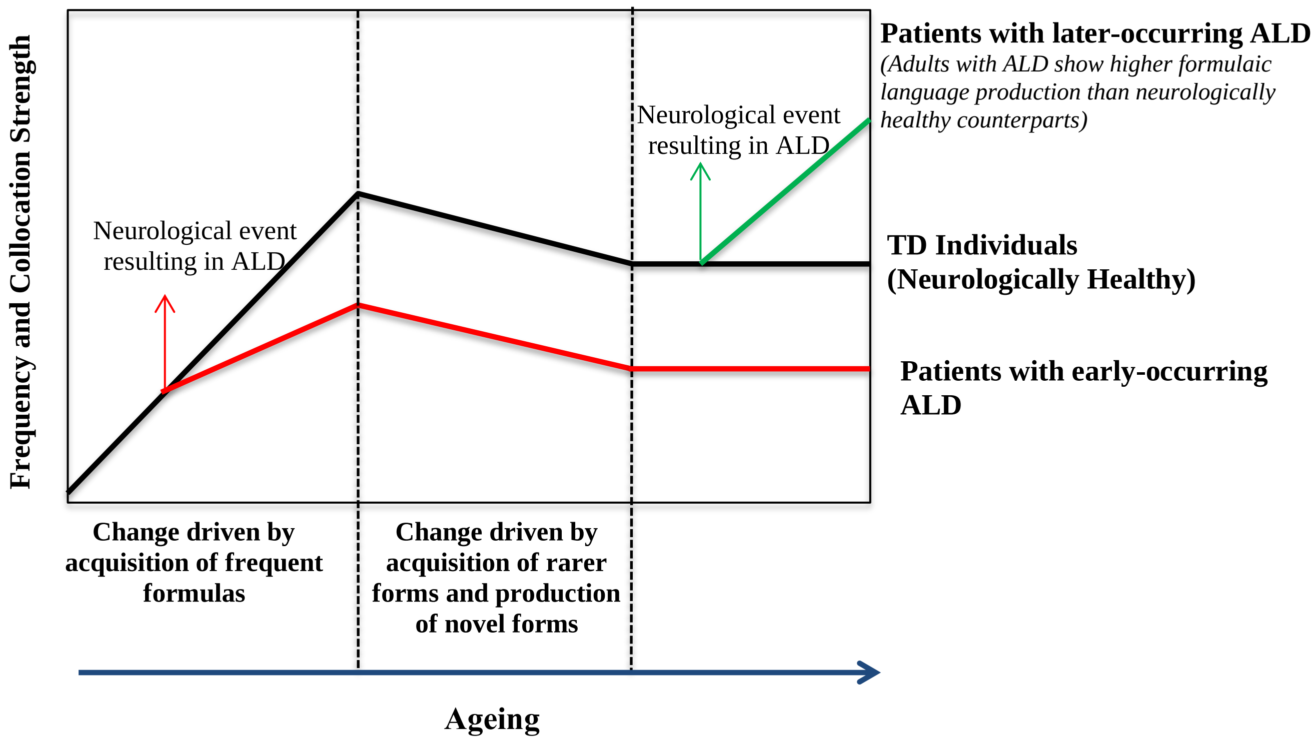

Figure 5: *Possible model of frequency and collocation strength over the lifespan and the impact of ALD.*

Figure 5 presents a possible model for language production over the lifespan, as captured by frequency and collocation strength, and the impact of early ALD and aphasia in adults. Two processes affect this aspect of language use: the learning of familiar, highly collocated forms, and the more generative ability to produce rare and novel forms. In a first stage, language development is driven by the former, in a second stage by the latter. Our results suggest that ALD in children interferes with acquisition of familiar forms, while aphasia in adults results in the language system being limited to these forms. We emphasize that this model is to a large degree speculative, and that more data are needed in order to test it.

While individuals with language impairment often show decreased capacity for statistical learning, statistical learning is commonly found. Our results show that this is also the case for children with ALD. Clinically, it is possible to employ implicit statistical learning in language intervention (Plante & Gómez, 2018), and studies on individuals with DLD have shown promising results. In some populations such as Williams syndrome, 'rule-based' generalizations may be impaired, meaning that statistical processing becomes even more important for acquisition (Stojanovik et al., 2018).

This exploratory study has explored learning mechanisms in a relatively understudied population. There are corroborations with current literature in developmental disorders of language with regard to statistical learning abilities. Yet, the data indicate possible differences in performance abilities across time and differences in language production as compared to adults with acquired language disorders. Future research should not only consider recruiting more children with ALD (which is a challenge given the low prevalence), but also collect samples from TD participants from across the lifespan. Correlations between deficits and the lesion site also require further investigation. The complexity of statistical processing should also be continued to be explored. For instance, the FLAT investigates frequencies of specific word forms and combinations thereof, but co-occurrence patterns can be described at other levels such as morphemes, lemmas or word classes, and multilevel models may better capture age effects and differences between clinical and non-clinical populations. Future studies could also employ tools that use child corpora for reference and more accurately reflect a child's linguistic environment. Despite limitations, the findings provide direction for further research into the relationship between statistical learning and typical and impaired language development, as well as for the development of tools for linguistic and cognitive assessment.

Acknowledgements

The manuscript is based on a dissertation project by Mok Xue Ting, Joelle, supervised by Vitor Zimmerer and Rosemary Varley. The project was conceived by Vitor Zimmerer and Mok Xue Ting, Joelle. Methods were designed by Vitor Zimmerer, Mok Xue Ting, Joelle, Rosemary Varley and James Douglas Saddy, and the recruitment was carried out by Mok Xue Ting, Joelle and Goh Siew Li. Data collection was carried out by Mok Xue Ting, Joelle. Data analysis was conducted by Mok Xue Ting, Joelle and Vitor Zimmerer. The first draft of the manuscript was written by Mok Xue Ting, Joelle and Vitor Zimmerer. Rosemary Varley contributed to further drafts.

We would like to thank all the participants and their parent(s)/guardian(s) for their time in participating in this study. We would also like to thank Leo Wei Zhi, for helping in recruitment of the patients, as well as Elizabeth Wonnacott and Courtenay Norbury for helpful comments when this work was presented at a lab meeting. We also thank our anonymous reviewers for their helpful suggestions.

Declaration of Interest Statement

The authors declare no conflict of interest.

References

Arciuli, J., & Simpson, I. C. (2011). Statistical learning in typically developing children: The role of age and speed of stimulus presentation. *Developmental Science, 14*(3), 464-473. https://doi.org/10.1111/j.1467-7687.2009.00937.x

Arciuli, J., & Simpson, I. C. (2012). Statistical learning is related to reading ability in children and adults. *Cognitive Science, 36*(2), 286-304. http://doi.org/10.1111/j.1551-6709.2011.01200.x

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language, 62*(1), 67-82. http://doi.org/10.1016/j.jml.2009.09.005

Bannard, C., & Lieven, E. (2012). Formulaic language in L1 acquisition. *Annual Review of Applied Linguistics, 32*, 3-16. https://doi.org/10.1017/S0267190512000062

Bogaerts, L., Siegelman, N., & Frost, R. (2020). Statistical learning and language impairments: Toward more precise theoretical accounts. *Perspectives on Psychological Science, 16(2),* 319-337. https://doi.org/10.1177%2F1745691620953082

Brunner, R. J., Kornhuber, H. H., Seemüller, E., Suger, G., & Wallesch, C. W. (1982). Basal ganglia participation in language pathology. *Brain and Language, 16*(2), 281–299. https://doi.org/10.1016/0093-934X(82)90087-6

Bruns, C., Varley, R. A., Zimmerer, V. C., Carragher, M., Brekelmans, G., & Beeke, S. (2019). I don't know: A usage-based approach to familiar collocations in non-fluent aphasia. *Aphasiology, 33*(2), 140-162. http://doi.org/10.1080/02687038.2018.1535692

Christiansen, M. H., & Chater, N. (2017). *Creating language. intergrating evolution, acquisition,* *and processing*. MIT Press.

Christiansen, M. H., Louise Kelly, M., Shillcock, R. C., & Greenfield, K. (2010). Impaired artificial grammar learning in agrammatism. *Cognition*, *116*(3), 382-393. https://doi.org/10.1016/j.cognition.2010.05.015

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). Erlbaum.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*, 72-89. http://doi.org/10.1093/applin/amm022

Conway, C. M. (2020). How does the brain learn environmental structure? Ten core principles for understanding the neurocognitive mechanisms of statistical learning. *Neuroscience & Biobehavioral Reviews*, *112*, 279-299. https://doi.org/10.1016/j.neubiorev.2020.01.032

Conway, C. M., Bauernschmidt, A., Huang, S. S., & Pisoni, D. B. (2010). Implicit statistical learning in language processing: Word predictability is the key. *Cognition*, *114*(3), 356-371. https://doi.org/10.1016/j.cognition.2009.10.009

Conway, C. M., & Pisoni, D. B. (2008). Neurocognitive basis of implicit learning of sequential structure and its relation to language processing. *Annals of the New York Academy of Sciences*, *1145*, 113–131. https://doi.org/10.1196/annals.1416.009

Dąbrowska, E., & Lieven, E. (2005). Towards a lexically specific grammar of children’s question constructions. *Cognitive Linguistics*, *16*(3), 437-474. http://doi.org/https://doi.org/10.1515/cogl.2005.16.3.437

Dennis, M. (2010). Language disorders in children with central nervous system injury. *Journal of*

Clinical and Experimental Neuropsychology, *32*(4), 417–422. https://doi.org/10.1080/13803390903164355

Diessel, H. (2007). Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology*, *25*(2), 108-127. http://doi.org/10.1016/J.NEWIDEAPSYCH.2007.02.002

Erickson, L. C., & Thiessen, E. D. (2015). Statistical learning of language: Theory, validity, and predictions of a statistical learning account of language acquisition. *Developmental Review*, *37*, 66–108. https://doi.org/10.1016/j.dr.2015.05.002

Evans, J. L., Saffran, J. R., & Robe-Torres, K. (2009). Statistical learning in children with specific language impairment. *Journal of Speech, Language, and Hearing Research*, *52*(2), 321-35. http://doi.org/10.1044/1092-4388(2009/07-0189)

Finn, A. S., Kalra, P. B., Goetz, C., Leonard, J. A., Sheridan, M. A., & Gabrieli, J. D. E. (2016). Developmental dissociation between the maturation of procedural memory and declarative memory. *Journal of Experimental Child Psychology*, *142*, 212-220. https://doi.org/10.1016/j.jecp.2015.09.027

Forster, K. I., & Forster, J. C. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods, Instruments & Computers*, *35*(1), 116-124. https://doi.org/10.3758/BF03195503

Gabay, Y., Thiessen, E. D., & Holt, L. L. (2015). Impaired statistical learning in developmental dyslexia. *Journal of Speech Language and Hearing Research*, *58*(3), 934. http://doi.org/10.1044/2015_JSLHR-L-14-0324

Geambaşu, A., Toron, L., Ravignani, A., & Levelt, C. C. (2020). Rhythmic Recursion? Human Sensitivity to a Lindenmayer Grammar with Self-similar Structure in a Musical Task. *Music & Science, 3*, 205920432094661. https://doi.org/10.1177/2059204320946615

Goodglass, H., Kaplan, E., & Barresi, B. (2001). *Boston diagnostic aphasia examination.* (3rd ed.). Pro-Ed.

Goschke, T., Friederici, A. D., Kotz, S. A., & van Kampen, A. (2001). Procedural learning in Broca's aphasia: Dissociation between the implicit acquisiton of spatio-motor and phoneme sequences. *Journal of Cognitive Neuroscience, 13*(3), 370-388. http://dx.doi.org/10.1162/08989290151137412

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words? *Psychological Science, 18*, 254-260. https://doi.org/10.1111/j.1467-9280.2007.01885.x

Gravel, J. S., Burstein, J. R., Donaher, J. G., Franck, K. H., Knightly, C., Solot, C., & White, A. (2007). Pediatric Communication Disorders. In R. F. Wetmore & L. M. Bell (Eds.), *Pediatric Otolaryngology: The Requisites in Pediatrics* (pp. 29-59). Mosby. https://doi.org/10.1016/B978-0-323-04855-2.50008-3

Harris Wright, H., Silverman, S., & Newhoff, M. (2003). Measures of lexical diversity in aphasia. *Aphasiology*, *17*(5), 443-452. http://doi.org/10.1080/02687030344000166

Hinzen, W., Rosselló, J., Morey, C., Camara, E., Garcia-Gorro, C., Salvador, R., & de Diego-Balaguer, R. (2018). A systematic linguistic profile of spontaneous narrative speech in pre-symptomatic and early stage Huntington's disease. *Cortex*, *100*, 71–83.

https://doi.org/10.1016/j.cortex.2017.07.022

Hotz, G., Helm-Estabrooks, N., Nelson, N., & Plante, E. (2010). *Pediatric Test of Brain Injury*. Brooks Publishing.

Janacsek, K., Fiser, J., & Nemeth, D. (2012). The best time to acquire new skills: Age-related differences in implicit sequence learning across the human life span. *Developmental Science*, 15, 496–505. https://doi.org/10.1111/j.1467-7687.2012.01150.x

Janacsek, K., Shattuck, K. F., Tagarelli, K. M., Lum, J. A. G., Turkeltaub, P. E., & Ullman, M. T. (2020). Sequence learning in the human brain: A functional neuroanatomical meta-analysis of serial reaction time studies. *NeuroImage, 207*, 116387, https://doi.org/10.1016/j.neuroimage.2019.116387

Karuza, E. A., Newport, E. L., Aslin, R. N., Starling, S. J., Tivarus, M. E., & Bavelier, D. (2013). The neural correlates of statistical learning in a word segmentation task: An fMRI study. *Brain and Language, 127(1),* 46-54. https://doi.org/10.1016/j.bandl.2012.11.007

Kóbor, A., Takács, Á., Kardos, Z., Janacsek, K., Horváth, K., Csépe, V., & Nemeth, D. (2018). ERPs differentiate the sensitivity to statistical probabilities and the learning of sequential structures during procedural learning. *Biological Psychology*, *135*, 180–193. https://doi.org/10.1016/j.biopsycho.2018.04.001

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*(1), 171-184. http://dx.doi.org/10.1037/a0025405

Lammertink, I., Boersma, P., Rispens, J., & Wijnen, F. (2020). Visual statistical learning in children with and without DLD and its relation to literacy in children with DLD. *Read*

Writing, 33, 1557–1589. https://doi.org/10.1007/s11145-020-10018-4

Lum, J. A. G., Conti-Ramsden, G., Morgan, A. T., & Ullman, M. T. (2014). Procedural learning deficits in specific language impairment (SLI): A meta-analysis of serial reaction time task performance. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *51*, 1-10. http://doi.org/10.1016/j.cortex.2013.10.011

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2006). Cracking the language code: Neural mechanisms underlying speech parsing. *Journal of Neuroscience*, *26*(29), 7629–7639. https://doi.org/10.1523/JNEUROSCI.5501-05.2006

McNealy, K., Mazziotta, J. C., & Dapretto, M. (2010), The neural basis of speech parsing in children and adults. *Developmental Science, 13*, 385-406. https://doi.org/10.1111/j.1467-7687.2009.00895.x

Meulemans, T., & Van der Linden, M. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, *69*, 199-221. https://doi.org/10.1006/jecp.1998.2442

Nakagawa, S. (2004). A farewell to Bonferroni: The problems of low statistical power and publication bias. *Behavioral Ecology*, *15*(6), 1044–1045. https://doi.org/10.1093/beheco/arh107

Newman-Norlund, R. D., Frey, S. H., Petitto, L. A., & Grafton, S. T. (2006). Anatomical substrates of visual and auditory miniature second-language learning. *Journal of Cognitive Neuroscience*, *18*(12), 1984–1997. https://doi.org/10.1162/jocn.2006.18.12.1984

Nissen, J. N., & Bullemer, P. (1987). Attentional requirements of learning: evidence from performance measures. *Cognitive Psychology*, *19*, 1-32. https://doi.org/10.1016/0010-0285(87)90002-8

Obeid R., Brooks, P. J., Powers, K. L., Gillespie-Lynch K., & Lum J. A. G. (2016). Statistical Learning in Specific Language Impairment and Autism Spectrum Disorder: A Meta-Analysis. *Frontiers in Psychology, 7,* 1245. https://doi.org/10.3389/fpsyg.2016.01245

Ostergaard, A.L. (1998). The effects on priming of word frequency, number of repetitions, and delay depend on the magnitude of priming. *Memory and Cognition*, *26*, 40–60. https://doi.org/10.3758/BF03211369

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, *316*(7139). https://doi.org/10.1136/bmj.316.7139.1236

Petersson, K. M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, *120*(2), 83–95. Retrieved from http://www.sciencedirect.com/science/article/pii/S0093934X10001380

Plante, E., & Gómez, R. L. (2018). Learning without trying: The clinical relevance of statistical learning. *Language, Speech, and Hearing Services in Schools, 49,* 710–722. https://doi.org/10.1044/2018_LSHSS-STLT1-17-0131

Plante, E., Patterson, D., Sandoval, M., Vance, C. J., & Asbjørnsen, A. E. (2017). An fMRI study of implicit language learning in developmental language impairment. *NeuroImage: Clinical, 14,* 277–285. https://doi.org/10.1016/j.nicl.2017.01.027.

Prusinkiewicz, P., & Hanan, J. (1989). *Lindenmayer Systems, Fractals, and Plants*. *Lecture Notes in Biomathematics.* (Vol. 79). Springer-Verlag.

Quade, D. (1967). Rank analysis of covariance. *Journal of the American Statistical Association*, *62*(320), 1187–1200.

Robertson, E. M. (2007). The serial reaction time task: Implicit motor skill learning? *The Journal of Neuroscience*, *27*(38), 10073–10075.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. http://doi.org/10.1126/science.274.5294.1926

Schuchard, J., Nerantzini, M., & Thompson, C. K. (2017). Implicit learning and implicit treatment outcomes in individuals with aphasia. *Aphasiology*, *31*(1), 25-48. http://doi.org/10.1080/02687038.2016.1147526

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). The Psychological Corporation A Harcourt Assessment Company.

Simor, P., Zavecz, Z., Horváth, K., Éltető, N., Török, C., Pesthy, O., Gombos, F., Janacsek, K., & Nemeth, D. (2019). Deconstructing Procedural Memory: Different Learning Trajectories and Consolidation of Sequence and Statistical Learning. *Frontiers in Psychology*, *9*(JAN), 2708. https://doi.org/10.3389/fpsyg.2018.02708

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, *175*, 111-122. http://doi.org/10.1016/j.bandl.2017.10.004

Stojanovik, V., Zimmerer, V. C., Hudson, K., Saddy, J. D., & Setter, J. (2018). Artificial grammar learning in Williams syndrome and in typical development: The role of rules, familiarity and prosodic cues. *Journal of Neurolinguistics*, *39*(2), 327-353. http://doi.org/10.1017/S0142716417000212

The British National Corpus, version 2 (BNC XML Edition). (2007).

http://www.natcorp.ox.ac.uk

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology. General, 134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

Ullman, M. T., Earle, F. S., Walenski, M., & Janacsek, K. (2020). The neurocognition of developmental disorders of language. *Annual Review of Psychology*, 71(1), 389-417. https://doi.org/10.1146/annurev-psych-122216-011555

Vadinova, V., Buivolova, O., Dragoy, O., van Witteloostuijn, M., & Bos, L. S. (2020). Implicit-statistical learning in aphasia and its relation to lesion location. *Neuropsychologia, 147,* 107591. doi: 10.1016/j.neuropsychologia.2020.107591

Vakil, E., Kahan, S., Huberman, M., & Osimani, A. (2000). Motor and non-motor sequence learning in patients with basal ganglia lesions: the case of serial reaction time (SRT). *Neuropsychologia*, *38*(1), 1-10. https://doi.org/10.1016/S0028-3932(99)00058-5

Van Lancker Sidtis, D. (2012). Formulaic Language and Language Disorders. *Annual Review of Applied Linguistics*, *32*, 62–80. https://doi.org/10.1017/S0267190512000104

Van Lancker Sidtis, D., Choi, J., Alken, A., & Sidtis, J. J. (2015). Formulaic Language in Parkinson's Disease and Alzheimer's Disease: Complementary Effects of Subcortical and Cortical Dysfunction. *Journal of Speech, Language, and Hearing Research : JSLHR*, *58*(5), 1493–1507. https://doi.org/10.1044/2015_JSLHR-L-14-0341

Van Lancker Sidtis, D., & Postman, W. A. (2006). Formulaic expressions in spontaneous speech of left- and right-hemisphere-damaged subjects. *Aphasiology*, *20*(5), 411-426.

http://doi.org/10.1080/02687030500538148

van Witteloostujin, M., Boersma, P., Wijnen, F., & Rispens, J. (2019). Statistical learning abilities of children with dyslexia across three experimental paradigms. *PLOS ONE, 14(8)*, e0220041. https://doi.org/10.1371/journal.pone.0220041

West, G., Vadillo, M. A., Shanks, D. R., & Hulme, C. (2018). The procedural learning deficit hypothesis of language learning disorders: we see some problems. *Developmental Science*, *21*(2), e12552. https://doi.org/10.1111/desc.12552

Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, *32*, 231–254. https://doi.org/10.1017/S026719051200013X

Zimmerer, V. C., Cowell, P. E., & Varley, R. A. (2014). Artificial grammar learning in individuals with severe aphasia. *Neuropsychologia*, *53*, 25-38. http://doi.org/https://doi.org/10.1016/j.neuropsychologia.2013.10.014

Zimmerer, V. C., Hardy, C. J. D., Eastman, J., Dutta, S., Varnet, L., Bond, R. L., Russell, L., Rohrer, J. D., Warren, J. D., & Varley, R. A. (2020). Automated profiling of spontaneous speech in primary progressive aphasia and behavioral-variant frontotemporal dementia: An approach based on usage-frequency. *Cortex*, *133*, 103–119. https://doi.org/10.1016/j.cortex.2020.08.027

Zimmerer, V. C., Newman, L., Thomson, R., Coleman, M. J., & Varley, R. A. (2018). Automated analysis of language production in aphasia and right hemisphere damage: Frequency and collocation strength. *Aphasiology*, *32*(11), 1267-1283.

http://doi.org/10.1080/02687038.2018.1497138

Zimmerer, V. C., Shirley, E., & Saddy, J. D. (2013). Serial reaction time learning of two Lindenmayer grammars. Unpublished manuscript.

Appendix A. Probabilities for each stimulus according to models of varying complexity.

Example sequence	One symbol model	Bigram model	Trigram model
A	0.381	0.381	0.381
B	0.618	1	1
B	0.618	0.381	0.618
A	0.381	0.618	1
B	0.618	1	1
A	0.381	0.618	0.381
B	0.618	1	1
B	0.618	0.381	0.618
A	0.381	0.618	1
B	0.618	1	1
A	0.381	0.618	0.381
B	0.618	1	1
B	0.618	0.381	0.618
A	0.381	0.618	1
B	0.618	1	1
B	0.618	0.381	0.618
A	0.381	0.618	1
B	0.618	1	1
A	0.381	0.618	0.381