

ORIGINAL ARTICLE

The effect of children’s prior knowledge and language abilities on their statistical learning

Katja Stärk^{1,*} , Evan Kidd^{1,2,3}  and Rebecca L. A. Frost^{1,4} 

¹Language Development Department, Max-Planck-Institute for Psycholinguistics, Nijmegen, The Netherlands, ²Research School of Psychology, The Australian National University, Canberra, Australia, ³ARC Centre of Excellence for the Dynamics of Language, Canberra, Australia and ⁴Department of Psychology, Edge Hill University, Ormskirk, UK

*Corresponding author. Email: Katja.Staerk@mpi.nl

(Received 12 December 2021; revised 2 August 2022; accepted 21 August 2022; first published online 26 September 2022)

Abstract

Statistical learning (SL) is assumed to lead to long-term memory representations. However, the way that those representations influence future learning remains largely unknown. We studied how children’s existing distributional linguistic knowledge influences their subsequent SL on a serial recall task, in which 49 German-speaking seven- to nine-year-old children repeated a series of six-syllable sequences. These contained either (i) bisyllabic words based on frequently occurring German syllable transitions (naturalistic sequences), (ii) bisyllabic words created from unattested syllable transitions (non-naturalistic sequences), or (iii) random syllable combinations (unstructured foils). Children demonstrated learning from naturalistic sequences from the beginning of the experiment, indicating that their implicit memory traces derived from their input language informed learning from the very early stages onward. Exploratory analyses indicated that children with a higher language proficiency were more accurate in repeating the sequences and improved most throughout the study compared to children with lower proficiency.

Keywords: statistical learning; serial recall; entrenchment; language development; incremental learning

This study’s preregistration can be found on AsPredicted: <https://aspredicted.org/546hk.pdf>. All supporting materials regarding this research can be found in the following Open Science Framework (OSF) repository: <https://osf.io/t5qf4/>.

Katja Stärk is supported by a PhD studentship from the Language Development Department, Max-Planck-Institute for Psycholinguistics, and Evan Kidd is supported by the Australian Research Council (CE140100041).

We thank the members of the Language Development Department at the Max Planck Institute for Psycholinguistics, and especially our two anonymous reviewers, for their insightful comments on this work, which significantly improved the manuscript. Many thanks to Caroline De Becker for help with transcribing the data and to Andrew Jessop for his helpful guidance with coding in R and providing the TP calculation function. Thanks also to Liz Tollenaar for help preparing the data for transcription, to Julia Egger for providing the Python script to combine text files, and to Greta Kaufeld for recording the stimuli. Finally, we are extremely grateful to the school, their friendly staff, and of course, the children who participated in the study.

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited

Statistical learning (SL), the ability to use probabilistic co-occurrence to group elements present in the environment, has been argued to support development across multiple sensory and cognitive domains. One domain in which SL has been identified as playing a role is language development (Romberg & Saffran, 2010; Saffran *et al.*, 1996), where it has been linked to learning across the subdomains of phonology (e.g., Kuhl, 2004), segmentation and word learning (e.g., Hay *et al.*, 2011; Lany & Saffran, 2010), and grammar (Kidd & Arciuli, 2016). A common assumption of research in this field is that SL results in detailed long-term memory representations for learned stimuli; however, how these representations endure and how they influence children's performance in experiments is largely unknown. In the current paper, we explore this issue by asking how children's knowledge of their native language influences their performance on an auditory SL task, and whether variation in performance is linked to their language proficiency.

There is significant evidence in favor of the argument that participants' prior linguistic experience guides their SL of linguistic stimuli. Notably, many studies have shown that adult participants' segmentation of words and grammatical patterns (e.g., non-adjacent dependencies) from running speech is constrained both by language-specific phonotactic constraints (i.e., rules) and, within the class of legal sequences, phonotactic probabilities within and across languages (e.g., Bonatti *et al.*, 2005; Dal Ben *et al.*, 2021; Finn & Hudson Kam, 2008; Mersad & Nazzi, 2011; Onnis *et al.*, 2005; Trecca *et al.*, 2019). Although infants and children have much less experience with language, their prior knowledge also influences SL. For instance, Lew-Williams and Saffran (2012) found that nine- and ten-month-old infants' statistical speech segmentation was guided by their expectations about word length, which were forged through pre-exposure to words that were either consistent or inconsistent in length with items in the to-be-segmented stream. These findings are echoed in work by Thiessen *et al.* (2019), who demonstrated that 13-month-old, but not seven-month-old, native English-learning infants showed a preference for attending to backward versus forward transitional probabilities (TPs) in an ambiguous artificial language. That is, older infants attended to the direction which aligned with the structural properties of their native language (i.e., the head-initial nature of English, which favors backward processing; see Onnis & Thiessen, 2013, for related findings with adults). Taken together, these data provide converging evidence that learners' expectations about linguistic input emerge over experience with language and can permeate into the laboratory to shape participants' performance on SL tasks.

In the current paper, we ask a related though crucially different question: do children draw upon their distributional knowledge of syllable transitions when processing new statistically defined linguistic input? Specifically, does the presence of attested and therefore highly probable bigrams from participants' native language boost SL? Attention to TPs has featured prominently in explanations of SL, since this is how higher units of linguistic organization (i.e., "words" in studies of segmentation) have been defined (Saffran *et al.*, 1996). Successful learning on these tasks indicates that participants either track TPs or chunk co-occurring syllables into bigger units (Aslin *et al.*, 1998; Perruchet, 2019; Perruchet & Vinter, 1998). The underlying assumption that the output of SL – the acquisition of statistical structure as defined by TPs – leads to long-term representations that aid future learning is less

frequently tested. Accordingly, we ask whether participants use these prior expectations about TPs in their ambient language, which they have presumably acquired across development, to process new input.

There is growing evidence from adults in support of this proposal. In a series of studies investigating the interrelationships between visual SL and auditory SL for linguistic and non-linguistic stimuli, Siegelman et al. (2018) reported that, while visual SL and auditory SL for non-linguistic stimuli were related (suggesting a common underlying capacity for performance across modalities), auditory SL for linguistic stimuli was not related to the other two. Indeed, they reported poor internal consistency for their linguistic auditory SL task, suggesting that participants showed little overlap in performance across individual test items within the task. However, Siegelman et al. found that performance on the linguistic SL task was predicted by informant ratings of “wordness” for the test items. That is, the more closely a test word resembled a real word in the participants’ native language (Hebrew), the more likely they would be to segment it from the speech stream and recognize it as a word at test, suggesting that participants entered the experiment with entrenched linguistic knowledge affecting their performance on the SL tasks based on linguistic items.

Two more recent studies with adult populations provide more direct evidence for this entrenchment effect. In a between-participants design, Elazar et al. (2022) tested Spanish-speaking participants on an auditory SL task using one of two speech streams. In a “Spanish-like” condition, participants listened to a stream of continuous speech containing test words defined by TPs that were highly attested in Spanish corpora. In a “Spanish-unlike” condition, the test words were defined by TPs that were rarely attested. Overall, the authors reported a pattern of results that was suggestive of a learning advantage for Spanish-like stimuli, thus providing evidence in favor of the argument that participants build expectations of likely syllable transitions, which they use as priors when parsing new input.

In a study that we build upon in the current paper, Stärk et al. (2023) investigated whether native German-speaking adults drew on their prior knowledge of German syllable co-occurrences (i.e., syllable pairs with high TPs) to acquire new but distributionally consistent linguistic input. Unlike Elazar et al. (2022), the authors tested the influence of attested TPs in a within-participants design, a more stringent test of the entrenchment hypothesis since any across-condition differences could not be attributed to differences across participants. Doing so required a different method of measuring SL. Accordingly, participants heard and repeated three different kinds of sequences in a serial recall task: two of which were structured, while the other sequences were unstructured foils, with the dependent measure being recall accuracy. The structured sequences contained bisyllabic experimental words that were either based on likely German syllable transitions (naturalistic sequences) or were devoid of attested transitions (non-naturalistic sequences). The unstructured sequences were scrambled combinations of syllables, serving as baseline. Findings indicated that adults drew on their knowledge of German syllable transitions during the task and did so from the early phases of the experiment onward, showing higher recall accuracy and faster improvement for the naturalistic sequences over the other two sequence types. The participants also performed significantly better on recall of the non-naturalistic sequences in comparison to foils,

demonstrating that they were able to acquire knowledge of two sequences within the context of the task, despite the fact that there was likely a degree of interference across conditions because all sequences were constructed from the same syllable inventory.

Overall, the work of Elazar *et al.* (2022) and Stärk *et al.* (2023) provides converging evidence in support of the hypothesis that adults draw on their prior linguistic knowledge to process and learn subsequent input, which raises the question of whether children differ from adults in their weighing of prior knowledge and the input statistics, given their ability to pick up on new statistics present in a SL task. This may or may not differ from adults, depending on the paradigm (see Raviv & Arnon, 2018; Shufaniya & Arnon, 2018). Notable differences between adults and children have been found in studies of verbal learning. For instance, Smalle *et al.* (2017) showed that nine- to ten-year-old children require substantially less exposure to implicitly acquire phonotactic restrictions on novel words than adults do. Additionally, Smalle *et al.* (2018) have reported that eight- to nine-year-old children retain implicitly learned phonological sequences better than adults, as demonstrated through Hebbian Repetition Learning. This raises the possibility that, while still within the sensitive period for language acquisition, children differ from adults. Consequently, we investigate whether children draw on their prior distributional knowledge when processing subsequent linguistic input, as previously reported in adults (Elazar *et al.*, 2022; Stärk *et al.*, 2023).

Furthermore, we address the issue concerning whether and how performance in laboratory-based SL tasks relates to children's real-world language development. Typically developing children's language proficiency (as measured with assessments of their expressive and receptive vocabulary) has been found to be correlated with their SL ability (e.g., Evans *et al.*, 2009; Frost *et al.*, 2020; Kidd & Arciuli, 2016; Lany, 2014; though the relationship is not always observed and appears to be task-dependent, see West *et al.*, 2021). Here we take a slightly different approach and ask: do individual differences in language proficiency result in different learning trajectories throughout the course of SL (e.g., as measured over the course of our experiment)? If SL constitutes an individual ability that supports natural language learning (Siegelman *et al.*, 2017), we should expect children of different language abilities to be differentially sensitive to statistically defined sequences, under the assumption that superior SL abilities (and its component processes, Arciuli, 2017) have supported acquisition throughout development.

Thus, in the present study, we examined the effect of prior distributional knowledge on the learning of statistically defined linguistic input in seven- to nine-year-old German-speaking children. We utilized the repetition paradigm from Stärk *et al.* (2023) as described above, adjusting it for children's lower working memory capacity by shortening the sequence length. We hypothesized that children would recall naturalistic sequences better than non-naturalistic sequences and show faster improvement for the naturalistic than non-naturalistic sequences, similar to the findings with adults (Elazar *et al.*, 2022; Stärk *et al.*, 2023). Additionally, we investigated whether children's language proficiency was related to their SL. While past studies have reported correlational analyses between SL performance and language proficiency, including studies utilizing sequence repetition (e.g., Smalle *et al.*, 2018), we were interested in how differences in language proficiency influenced the

dynamics of learning throughout the task. Since no similar study exists, this analysis was exploratory.

Method

This study was preregistered on AsPredicted: <https://aspredicted.org/546hk.pdf>. All of our materials, data, analyses, and results are openly available on OSF: <https://osf.io/t5qf4/>.

Participants

Forty-nine seven- to nine-year-old native German-speaking children (29 female, 20 male; mean age = 8;7 years; months, SD = 0;6, range: 7;6–9;11) without any known hearing, speech, or language disorders were included in the final sample (N = 49). We had planned to test 60 children, but recruitment and testing were hindered by the COVID-19 pandemic. Despite not being able to recruit our originally planned sample, our power analysis suggests that a sample of 49 is sufficient to detect an effect size of 0.2 syllables recall difference between naturalistic and non-naturalistic sequences as well as between non-naturalistic and foil sequences (i.e., an effect size half the size of the one found in adults, compensating for children's cognitive development). The analysis was conducted in R 4.1.2 (R Core Team, 2022) using the package *simr* 1.0.6 (Green & MacLeod, 2016) and relied upon data from Stärk et al. (2023) whose design we adjusted for the current study (see the Analysis folder on OSF for a detailed description of this and two post hoc power analyses).

Participants were recruited from a German primary school in Leipzig, Germany. Invitations for participation were sent to the parents of all second- and third-graders (with German second-graders being 7–8 years old and third-graders being 8–9 years old). Of these, 55 consented to participating in our experiment. Data for six children were excluded from analyses: two did not fulfil the inclusion criteria regarding language proficiency (see Design section for more details), two found the serial recall task (i.e., our SL measure) too difficult to complete, and two were excluded due to technical failure. The study was approved by the Ethical Committee of the Faculty of Social Sciences, Radboud University, and was carried out in accordance with the World Medical Association Declaration of Helsinki. Children were free to withdraw at any time. They received a certificate and stickers for their participation.

Design

The experiment utilized a serial recall task, in which participants were presented with sequences of six syllables and repeated them out loud (see Figure 1). The task is based on studies of the Hebb repetition effect (Hebb, 1961; Page & Norris, 2009), and was modified from Stärk et al.'s (2023) related study with adults to be suitable for children. Specifically, we reduced the length of the sequences (sequences comprised 6 syllables, rather than 8), thereby making allowances for the fact that working memory (Cowan, 2016) and other cognitive processes supporting performance (e.g., processing speed, Kail & Salthouse, 1994) improve across childhood. The study had a within-participants design, in which all participants received all three

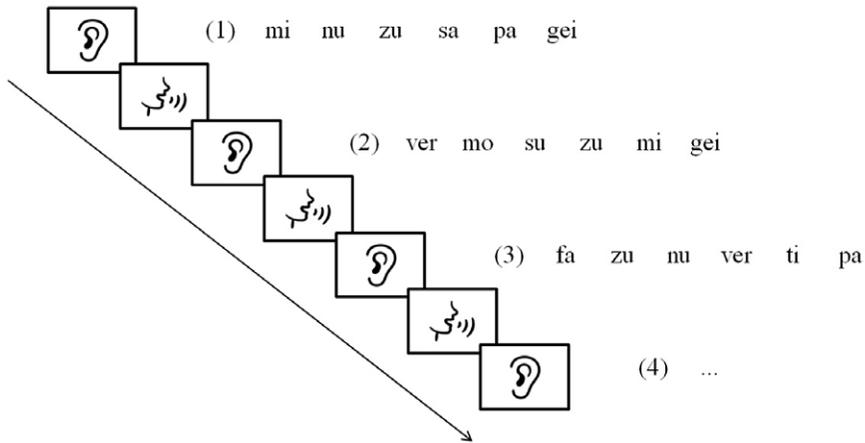


Figure 1. Three example experimental sequences: one naturalistic sequence (1), one unstructured foil sequence (2), and one non-naturalistic sequence (3). On each trial, participants listened to a six-syllable sequence and then repeated it. Design adapted from Stärk *et al.* (2023).

sequence types: naturalistic, non-naturalistic, and unstructured foils. The naturalistic and non-naturalistic sequences were structured, with each containing three bisyllabic experimental words, whereas unstructured sequences comprised the same syllables as the structured sequences but in a scrambled order (see the Materials section for further details). Participants' repetitions of the sequences were scored for accuracy, and we examined performance across the task to gain insights into learning over the course of exposure.

Materials

Language proficiency assessments

We assessed the children's language proficiency to ensure that both their lexical and morphosyntactic knowledge was representative of the average German-speaking seven- to nine-year-old. To measure their lexical knowledge, we used the German Peabody Picture Vocabulary Test version IV (PPVT; Lenhard *et al.*, 2015). This is a test of receptive vocabulary, in which children are shown four pictures per trial and have to identify the picture that best matches a word they were given. Their correct responses are counted and converted into an age-dependent score. Children whose performance was 1.5 SDs below the average of their age group were excluded from the main analyses.

To measure the children's morphosyntactic knowledge, we used the German LITMUS sentence repetition task (SRT; Abed Ibrahim *et al.*, 2018; Hamann *et al.*, 2013; Hamann & Abed Ibrahim, 2017). This test consists of 45 sentences with varying degrees of difficulty, which children listen to and repeat. Their correct responses are counted and form their score. If performance was 2.5 SDs below the average estimated from our own sample, children were excluded from the main analyses. A more lenient cut-off for the SRT was chosen because the test is not standardized.

Experimental stimuli

For the serial recall task (i.e., our SL measure), participants repeated sequences of six syllables. There were three different sequence types presented pseudo-randomly throughout the exposure: (i) naturalistic sequences, (ii) non-naturalistic sequences, and (iii) unstructured foil sequences. The first two sequence types were structured, meaning that they contained experimental words which participants could segment, while the unstructured foil sequences did not contain any learnable patterns. Importantly, the words contained in the naturalistic sequences differed from the words contained in the non-naturalistic sequences, in that the naturalistic words comprised frequently co-occurring German syllable pairs while the non-naturalistic words comprised unattested syllable pairs. That is, neither sequence type contained natural German words but the naturalistic words might appear familiar to German speakers because of their high TPs in natural German. They represent the outcome of participants' SL in natural German. In the current study, we investigate whether seven- to nine-year-old children draw on this prior knowledge (i.e., the outcome of previous SL in natural language) to process a new experimental language (as modeled by the six naturalistic words). This is compared to participants' SL in a language without prior knowledge (as modeled by the six non-naturalistic words).

The experimental words were created via a corpus analysis, which analyzed the TPs of the 1000 most frequent words of German child-directed speech on the CHILDES database (MacWhinney, 2000)¹. Six syllable pairs (i.e., bigrams) occurring with high within-word backward TPs were selected for use as the naturalistic words (naturalistic words: *gefa*, *minu*, *moti*, *pagei*, *versu*, *zusa*; see Stärk et al., 2022, for evidence that backward TPs are more reliable cues to 'wordness' than forward TPs in German speech). These were in turn used to create the naturalistic sequences. Importantly, the experimental naturalistic words did not comprise existing German words, but the high TPs (TP > 0.20, with an average TP = 0.69 and a range of 0.21–1) between the syllables within each pair make them potentially familiar to German speakers. There was no repetition of syllables across the words. The non-naturalistic words were created by switching the first and second syllables of the naturalistic words and combining each final syllable with a different first syllable, such that the syllable pairs neither form German words nor have high TPs (TP = 0) in natural German (non-naturalistic words: *fazu*, *geimi*, *nuver*, *samo*, *suge*, *tipa*). Unstructured foil sequences consisted of the same 12 syllables presented in a scrambled order, such that these sequences contained neither learnable patterns nor German or experimental words.

Within the context of the experiment, both types of structured sequences had perfect within-word TPs (*structured sequences*: within-word TPs = 1, between-word TPs \leq 0.25), while the TPs between syllables in the unstructured foil sequences were generally low (*unstructured sequences*: TPs \leq 0.17). Since participants repeated all three sequence types, with syllables being repeated across types, within-word TPs for both structured sequences were TP = 0.33, and TPs for all other syllable pairs were TPs \leq 0.14, calculated over the entire exposure. Forward and backward TPs were equal within the experimental sequences (see Stimuli file on OSF for the precise numbers and the details of the stimuli creation).

Each syllable was recorded in isolation by a female native speaker of German. The recordings were adjusted using Audacity (Audacity Team, 2018) to ensure uniformity of length, resulting in an average syllable duration of 377ms (range: 352ms–416ms). The experiment comprised 72 sequences in total, 24 of each sequence type. Each sequence consisted of six syllables (i.e., three experimental words in the structured sequences), separated by 500ms of silence. Thus, while we refer to our structured sequences as containing “words,” it is important to remember that to the participant, the stimuli were lists of syllables; thus, any grouping of the syllables based on TPs, be it from existing naturalistic knowledge or knowledge gained through the course of the task, is evidence of SL, which in this case can be interpreted as the chunking of co-occurring syllables. The final syllable of each sequence was followed by a beep, which indicated that participants could start their repetition.

The experiment was divided into 12 blocks of six sequences, which contained two items of each sequence type. Within each block, sequences were presented pseudo-randomly, with no direct repetition of a particular sequence type (e.g., there was no adjacent presentation of naturalistic sequences). There were four experimental lists, which differed in the order of blocks (and with which sequence type the experiment started). Participants were randomly assigned to one list. Across the entire experiment, each word occurred 12 times in total, appearing equally often in each position within a sequence (for more information on the stimuli and their creation see the Materials folder of the project on OSF).

Procedure

Children whose parents signed the consent form were invited for two experimental sessions at their after-school club. The children were tested over two sessions of approximately 30 minutes. In the first session, children completed the German PPVT (Lenhard *et al.*, 2015) and the SRT (Abed Ibrahim *et al.*, 2018; Hamann *et al.*, 2013; Hamann & Abed Ibrahim, 2017). In the second session, they completed the serial recall task (i.e., our main task, measuring children’s SL). Both sessions took place in a quiet, private room at the school. Stimuli were played over closed-cup headphones via a laptop, and the children repeated the sequences into a microphone, with repetitions being recorded for subsequent offline coding. The serial recall task was conducted using the software Presentation (Neurobehavioral Systems, Inc, 2014).

If the children fulfilled the language proficiency inclusion criteria in the first session, they were invited for the second session on the following day. The serial recall task was introduced as a game, which had the aim of helping an alien to repair their broken spaceship and fly back to their home planet. The children were told that they could help the alien by repeating what it says. First, they received six unstructured practice trials: three four-syllable sequences and three six-syllable sequences, comprising a different set of syllables than the experimental sequences (*ba, fun, gi, re, se, to*), before receiving the 72 experimental sequences. The children received a sticker and were told more components of the alien story after the practice trials, during the two breaks (after Sequences 24 and 48), and after completing the experiment. At the end of the session, the children were debriefed and received a certificate.

Data preparation

Language proficiency assessments

The PPVT was coded online by the experimenter, while the SRT was transcribed and coded offline upon completion by the experimenter and a trained assistant. The SRT data were coded for identical repetition following Hamann and Abed Ibrahim (2017). That is, children received a point for each sentence repeated entirely verbatim. Five participants' SRT recordings (10%) were transcribed by both coders for inter-transcriber reliability analyses, which revealed a strong reliability between the two transcribers (observed agreement (OA) = 94.7%; $\kappa = 0.83$ with 95% CIs of [0.74, 0.93]; following the more conservative interpretation of the kappa statistic suggested by McHugh, 2012). Furthermore, Cronbach's alpha (α) was calculated for the SRT to determine task-internal reliability, which revealed good internal consistency ($\alpha = .81$ [.74, .88]). This justifies its use as individual differences measure in our exploratory analysis (see OSF for almost identical values of McDonald's omega).

Serial recall task

Sequence repetitions were transcribed by the experimenter and a trained assistant. Participants' repetitions were scored at both the syllable and the bigram level, with a bigram being an experimental word in the two structured conditions and random syllable pairs in the unstructured condition (therefore referred to as "bigram level" rather than "word level" here). At the syllable level, participants received one point for each syllable repeated correctly in the correct position (max. 6 points per sequence). At the bigram level, participants received one point for each bigram (i.e., "word" or syllable pair) repeated correctly in the correct position (max. 3 points per sequence), providing valuable information about whether participants recalled sequences better because of learning the experimental words (i.e., bigrams) in the structured sequence types. In the unstructured sequences, the three bigrams per sequence were the random syllable pairs in Syllable positions 1 and 2, 3 and 4, and 5 and 6 (with syllable combinations varying between sequences).

Recall scores at the syllable and bigram level were highly correlated for all three sequence types (*naturalistic sequences*: $r(47) = .96$, $p < .001$; *non-naturalistic sequences*: $r(47) = .94$, $p < .001$; *unstructured foil sequences*: $r(47) = .88$, $p < .001$). Data for five participants were transcribed by both coders for inter-transcriber reliability analyses, revealing a moderate to strong reliability between the two transcribers (*syllable level*: OA = 79.2%; $\kappa = 0.74$ [0.69, 0.79]; *bigram level*: OA = 88.3%; $\kappa = 0.80$ [0.74, 0.86]; again following the more conservative interpretation of kappa suggested by McHugh, 2012). Furthermore, Cronbach's alpha (α) was calculated for the serial recall task as measure of task-internal reliability, revealing acceptable to excellent internal consistency (see Table 1, and see OSF for almost identical values of McDonald's omega).

We preregistered two different coding schemes for the serial recall task, the one described above and a serial-order coding scheme, which grants participants points more generously for repeating a syllable or bigram in the correct serial order rather than in the exact position within the sequence (see Isbilen et al., 2020, and Kidd et al., 2020). We report the results of the former coding scheme only, since the strict scheme is more conservative. However, the results of all coding schemes can be

Table 1. Task-internal consistency measures for the serial recall task

Sequence type	Syllable level		Bigram level	
	α	[95% CI]	α	[95% CI]
Overall	.95	[.93, .97]	.93	[.90, .96]
Naturalistic	.85	[.80, .91]	.81	[.74, .89]
Non-naturalistic	.89	[.84, .93]	.87	[.82, .92]
Unstructured foils	.85	[.79, .91]	.77	[.67, .86]

Notes: Statistic: α = Cronbach's alpha.

found in the Analysis folder of our OSF project. The results do not differ substantially between the different coding schemes.

Results

All of the analyses presented in this paper were performed in R 4.1.3 (R Core Team, 2022) using RStudio (RStudio Team, 2022). Data preprocessing and visualization were performed using the package *tidyverse* 1.3.1 (Wickham, 2017; Wickham *et al.*, 2019).

Language proficiency assessments

On the PPVT, the children scored on average 54.63 (SD = 9.07, range: 38–73), which is slightly higher than the normed average of 50. One child performed below our inclusion threshold of 1.5 SDs below the norming average (PPVT = 35) and was thus excluded from the analyses. On the SRT, the children scored on average 34.89 (SD = 5.36, range: 23–44). One child scored 2.2 SDs below the group average on the standard scoring criteria, but below the inclusion threshold of 2.5 SDs on more sensitive coding schemes (see preregistration, coding scheme, and analysis files on OSF). This participant was excluded from the analyses.

Serial recall task

Analysis by experimental block

We analyzed the data using generalized linear mixed-effects models (package *lmerTest* 3.1-3; Kuznetsova *et al.*, 2017; based on *lme4* 1.1-28; Bates *et al.*, 2015), with syllable recall and bigram recall as the dependent variables to test overall recall and recall of the experimental words, respectively. We specified a Poisson distribution with a log-link, since the dependent measures were count variables. Models were fitted with a fixed effect of sequence type using sliding contrasts (*naturalistic*: 0.5 vs. *non-naturalistic*: -0.5, and *non-naturalistic*: 0.5 vs. *foils*: -0.5) to examine whether recall differed across the experimental conditions, and a fixed effect of block was entered as a centered continuous variable to examine learning over

Table 2. Summary of the linear mixed-effects models investigating the influence of sequence type and block on the children's syllable and bigram recall, respectively

	Estimate	95% CI	SE	<i>t</i>	<i>p</i>
Syllable level					
(Intercept)	0.29	[0.14, 0.44]	0.08	3.74	< .001
Nat. vs. Non-nat.	0.10	[0.04, 0.17]	0.03	3.21	.001
Non-nat. vs. Foils	0.05	[-0.01, 0.10]	0.03	1.66	.10
Block	0.01	[-0.04, 0.05]	0.02	0.25	.80
Nat. vs. Non-nat. * Block	0.02	[-0.01, 0.06]	0.02	1.17	.24
Non-nat. vs. Foils * Block	0.01	[-0.03, 0.05]	0.02	0.49	.63
Bigram level					
(Intercept)	-1.15	[-1.39, -0.91]	0.13	-9.17	< .001
Nat. vs. Non-nat.	0.24	[0.11, 0.36]	0.06	4.07	< .001
Non-nat. vs. Foils	0.14	[0.04, 0.23]	0.05	2.90	.004
Block	0.01	[-0.06, 0.09]	0.04	0.37	.71
Nat. vs. Non-nat. * Block	0.03	[-0.04, 0.09]	0.03	0.82	.42
Non-nat. vs. Foils * Block	0.02	[-0.05, 0.09]	0.04	0.65	.52

Notes: Sequence type: Nat. = naturalistic sequences, Non-nat. = non-naturalistic sequences, Foils = unstructured foil sequences. Model fit syllable level: AIC = 10,961; BIC = 11,023; $R^2_m = 0.027$; $R^2_c = 0.088$; ICC = 0.063; RMSE = 1.248; $\sigma = 1$; model fit bigram level: AIC = 5661; BIC = 5742; $R^2_m = 0.046$; $R^2_c = 0.415$; ICC = 0.387; RMSE = 0.599; $\sigma = 1$.

the course of the experiment. Additionally, the interaction between the two factors was included as a fixed effect. We fitted the maximal model supported by the data (Barr et al., 2013), controlling for participants and items as random intercepts, with sequence type and block (as well as their interaction) as random slopes for participants, but not items as those differed between sequence types and blocks. The marginal and conditional R^2 effect sizes are also reported as goodness-of-fit estimates. These denote the proportion of the variance explained by the model both with (conditional R^2) and without (marginal R^2) controls for sources of random variance (Johnson, 2014; Nakagawa & Schielzeth, 2013).

There was a significant main effect of sequence type at both the syllable level and the bigram level, with participants displaying better recall for naturalistic than non-naturalistic sequences (see Table 2). At the bigram level, participants also showed better recall for the non-naturalistic than the unstructured foil sequences, indicating that the non-naturalistic words were also learnt, even though this did not facilitate the overall syllable recall for those sequences (for a visualization of participants' syllable and bigram recall accuracy, see Figures 2 and 3, respectively). There was no significant main effect of block, and no significant interaction between block and sequence type, suggesting that the conditional differences across sequence types reflected the overall performance of children across the experiment.

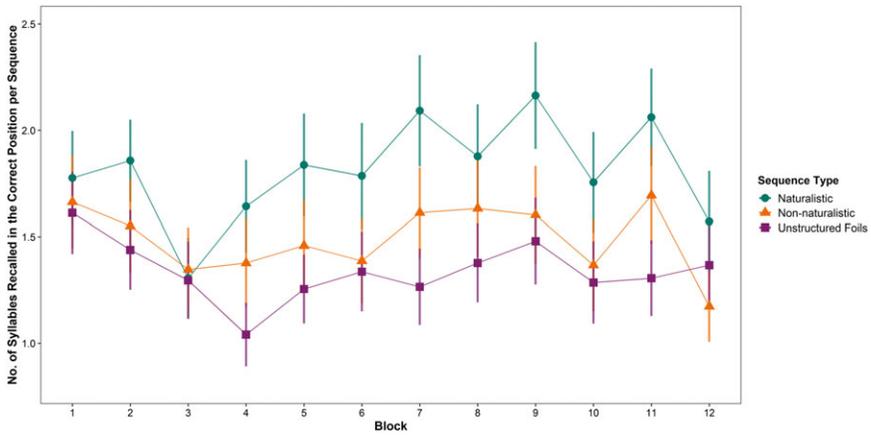


Figure 2. Mean syllable recall per sequence for the three sequence types (naturalistic, non-naturalistic, and unstructured foils), given by experimental block (1-12). Error bars indicate standard error.

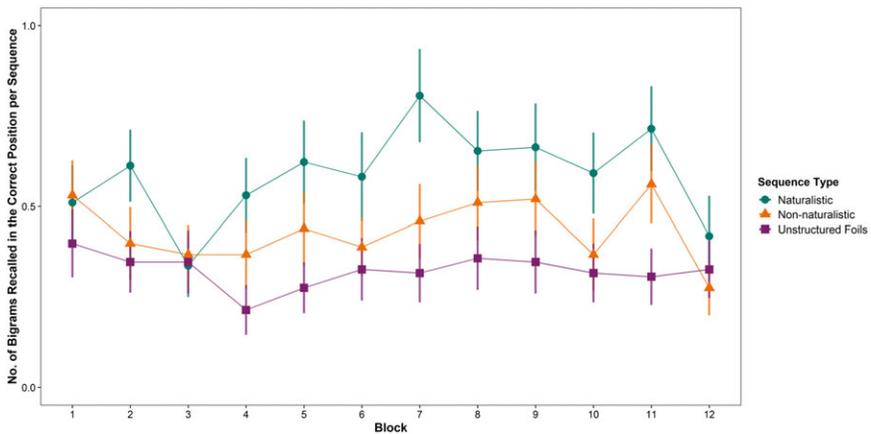


Figure 3. Mean bigram recall per sequence for the three sequence types (naturalistic, non-naturalistic, and unstructured foils), given by experimental block (1-12). Error bars indicate standard error.

Analysis by exposure phase

While there were no effects of block in the previous analyses, performance across the experiment was not totally even. Stärk *et al.* (2023) found that adult performance in the current task varied across phases of the experiment (early, intermediate, and late). We conducted the same analysis, dividing the exposure into three phases, separated by breaks during the data acquisition (*early exposure*: Blocks 1–4; *intermediate exposure*: Blocks 5–8; *late exposure*: Blocks 9–12). Accordingly, we added exposure phase instead of block as a fixed effect in the models examining children’s syllable and bigram recall. We fitted the maximal model supported by the data (Barr *et al.*, 2013) with sequence type (sliding contrast: *naturalistic*: 0.5 vs. *non-naturalistic*: –0.5, and *non-naturalistic*: 0.5 vs. *foils*: –0.5), exposure phase (sliding

Table 3. Summary of the linear mixed-effects models investigating the influence of sequence type and exposure phase on the children's syllable and bigram recall, respectively

	Estimate	95% CI	SE	<i>t</i>	<i>p</i>
Syllable level					
(Intercept)	0.29	[0.12, 0.44]	0.08	3.64	< .001
Nat. vs. Non-nat.	0.12	[0.05, 0.19]	0.03	3.60	< .001
Non-nat. vs. Foils	0.05	[-0.02, 0.10]	0.03	1.61	.11
Early vs. Intermediate	-0.01	[-0.06, 0.04]	0.02	-0.52	.61
Intermediate vs. Late	0.03	[-0.02, 0.07]	0.02	1.30	.19
Nat. vs. Non-nat. * Early vs. Intermediate	0.01	[-0.03, 0.05]	0.02	0.41	.68
Nat. vs. Non-nat. * Intermediate vs. Late	0.02	[-0.02, 0.06]	0.02	1.03	.30
Non-nat. vs. Foils * Early vs. Intermediate	0.02	[-0.03, 0.06]	0.02	0.89	.37
Non-nat. vs. Foils * Intermediate vs. Late	-0.02	[-0.07, 0.03]	0.02	-0.79	.43
Bigram level					
(Intercept)	-1.13	[-1.36, -0.88]	0.12	-9.47	< .001
Nat. vs. Non-nat.	0.17	[0.06, 0.28]	0.05	3.16	.002
Non-nat. vs. Foils	0.14	[0.04, 0.24]	0.05	2.92	.003
Early vs. Intermediate	0.05	[-0.03, 0.13]	0.04	1.21	.23
Intermediate vs. Late	0.01	[-0.06, 0.08]	0.04	0.20	.84
Nat. vs. Non-nat. * Early vs. Intermediate	0.03	[-0.05, 0.10]	0.04	0.76	.45
Nat. vs. Non-nat. * Intermediate vs. Late	0.01	[-0.07, 0.08]	0.04	0.21	.84
Non-nat. vs. Foils * Early vs. Intermediate	0.03	[-0.05, 0.12]	0.04	0.67	.50
Non-nat. vs. Foils * Intermediate vs. Late	0.00	[-0.09, 0.09]	0.05	-0.05	.96

Notes: Sequence type: Nat. = naturalistic sequences, Non-nat. = non-naturalistic sequences, Foils = unstructured foil sequences. Exposure phase: Early, Intermediate, Late. Model fit syllable level: AIC = 10,962; BIC = 11,085; $R^2_m = 0.024$; $R^2_c = 0.394$; ICC = 0.379; RMSE = 1.246; $\sigma = 1$; model fit bigram level: AIC = 5674; BIC = 5761; $R^2_m = 0.045$; $R^2_c = 0.084$; ICC = 0.041; RMSE = 0.598; $\sigma = 1$.

contrast: *early exposure*: -0.5 vs. *intermediate exposure*: 0.5, and *intermediate exposure*: -0.5 vs. *late exposure*: 0.5), and their interaction as fixed effects, and random intercepts and slopes for participants and items, where appropriate.

Comparable to the results of the previous analysis, we found significant main effects of sequence type but neither a main effect of exposure phase nor an interaction between the two factors (see Table 3). For an illustration of syllable and bigram recall accuracy over the three phases see Figures 4 and 5.

Relationship between language proficiency and SL

We next report our exploratory analyses investigating whether differences in children's performance on the SL task were related to their native-language proficiency.

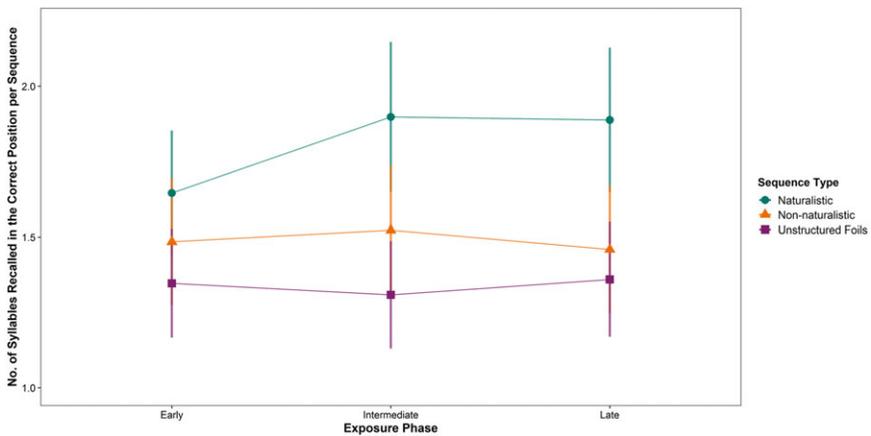


Figure 4. Mean syllable recall per sequence for the three sequence types (naturalistic, non-naturalistic, and unstructured foils), given by exposure phase (early, intermediate, late). Error bars indicate standard error.

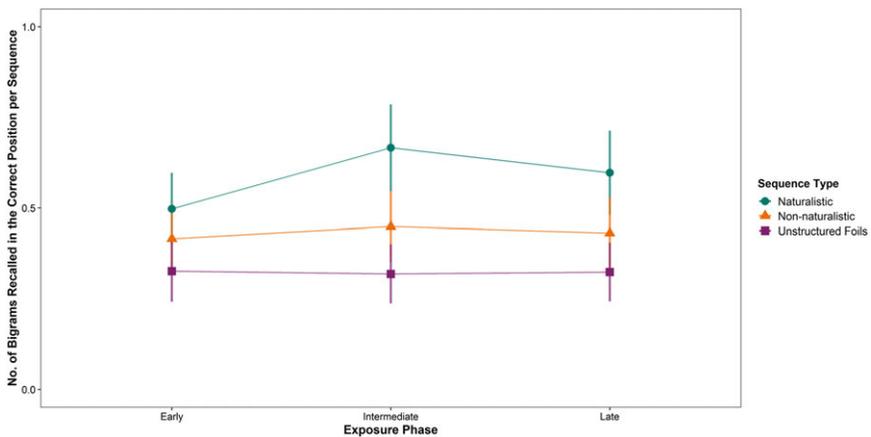


Figure 5. Mean bigram recall per sequence for the three sequence types (naturalistic, non-naturalistic, and unstructured foils), given by exposure phase (early, intermediate, late). Error bars indicate standard error.

We obtained two measures of language proficiency – vocabulary size, as measured by the PPVT, and morphosyntactic processing, as measured by the SRT. Table 4 shows the bivariate correlations between all variables, in addition to partial correlations between repetition of structured sequences and language proficiency, controlling for foil repetition. All bivariate correlations between repetition of syllable sequences and the SRT and PPVT were positive and all but two were significant. The partial correlations testing the relationship between syllable recall and language proficiency controlling for foil repetition were lower. Only the partial correlation between naturalistic repetition of bigrams and SRT performance was significant.

Table 4. Pearson bivariate correlations between children's recall in the three sequence types (naturalistic, non-naturalistic, and foil sequences) and their performance on the language proficiency assessments (SRT and PPVT) at the syllable (left) and bigram level (right). Partial correlations between language assessments and structured sequences, controlling for foil repetition, appear in brackets

	Syllable level				Bigram level			
	Non-nat.	Foils	SRT	PPVT	Non-nat.	Foils	SRT	PPVT
Nat.	.74 ***	.83 ***	.70 *** (.24)	.38 ** (.07)	.64 ***	.74 ***	.62 *** (.30 *)	.30 * (.05)
Non-nat.	–	.85 ***	.67 *** (.13)	.27 (–.16)	–	.81 ***	.58 *** (.17)	.16 (–.23)
Foils		–	.74 ***	.41 **		–	.63 ***	.36 *
SRT			–	.40 **			–	.40 **
PPVT				–				–

Notes: Sequence type: Nat. = naturalistic sequences, Non-nat. = non-naturalistic sequences, Foils = unstructured foil sequences. $N = 49$; * $p < .05$, ** $p < .01$, *** $p < .001$

While the simple correlations between sequence recall and PPVT or SRT were significant, we chose the latter measure as our individual differences variable because the recall task and the SRT both involve sequencing linguistic units in the same modality, but with different units (syllables versus morphemes). Finding that the two are systematically related would provide evidence in support of the idea that sequencing of syllables and morphemes share an underlying common and statistically sensitive mechanism (Isbilen et al., 2022). Thus, we next analyzed whether SRT performance influenced children's performance on the experiment. We fitted the maximal model supported by the data (Barr et al., 2013) with sequence type (sliding contrast: *naturalistic*: 0.5 vs. *non-naturalistic*: –0.5, and *non-naturalistic*: 0.5 vs. *foils*: –0.5), block (centered continuous variable) and SRT (centered continuous variable), as well as the interactions between the three factors, as fixed effects, and random intercepts and slopes for participants and items, where appropriate. In addition to the main effects of sequence type described above, there was a significant main effect of SRT score (*syllable level*: $\beta = 0.41$ [0.32, 0.50], $t = 8.34$, $p < .001$; *bigram level*: $\beta = 0.59$ [0.41, 0.76], $t = 6.55$, $p < .001$), with children who scored higher on the SRT displaying better syllable and bigram recall on the experimental task. Furthermore, there was a significant interaction between sequence type and SRT performance (*syllable level*: $\beta = -0.05$ [–0.09, 0.00], $t = -2.04$, $p = .04$; *bigram level*: $\beta = -0.11$ [–0.20, –0.02], $t = -2.36$, $p = .02$), with a greater recall difference between naturalistic and non-naturalistic sequences for children with lower SRT scores compared to their higher SRT scoring peers. At the syllable level, there was an additional significant interaction between block and SRT performance (*syllable level*: $\beta = 0.05$ [0.00, 0.09], $t = 2.25$, $p = .03$), with children with higher SRT scores showing greater improvement in syllable recall throughout the experiment, although the effect was not significant at the bigram level (see analysis file on OSF for full details).

Table 5. Summary of the linear mixed-effects model investigating the influence of sequence type, exposure phase, and SRT score on the children's syllable recall

	Estimate	95% CI	SE	<i>t</i>	<i>p</i>
(Intercept)	0.29	[0.18, 0.39]	0.05	5.46	< .001
Nat. vs. Non-nat.	0.12	[0.05, 0.19]	0.03	3.65	< .001
Non-nat. vs. Foils	0.05	[-0.02, 0.10]	0.03	1.52	.13
Early vs. Intermediate	-0.01	[-0.05, 0.04]	0.02	-0.49	.63
Intermediate vs. Late	0.02	[-0.02, 0.07]	0.02	1.04	.30
SRT	0.41	[0.31, 0.50]	0.05	8.33	< .001
Nat. vs. Non-nat. * Early vs. Intermediate	0.01	[-0.04, 0.05]	0.02	0.21	.83
Nat. vs. Non-nat. * Intermediate vs. Late	0.03	[-0.02, 0.08]	0.02	1.35	.18
Non-nat. vs. Foils * Early vs. Intermediate	0.02	[-0.04, 0.06]	0.02	0.66	.51
Non-nat. vs. Foils * Intermediate vs. Late	0.00	[-0.06, 0.05]	0.03	-0.17	.87
Nat. vs. Non-nat. * SRT	-0.05	[-0.10, 0.00]	0.02	-2.06	.04
Non-nat. vs. Foils * SRT	0.01	[-0.03, 0.05]	0.02	0.39	.69
Early vs. Intermediate * SRT	0.06	[0.02, 0.11]	0.02	2.85	.004
Intermediate vs. Late * SRT	0.00	[-0.04, 0.04]	0.02	-0.05	.96
Nat. vs. Non-nat. * Early vs. Intermediate * SRT	0.01	[-0.04, 0.05]	0.02	0.47	.64
Nat. vs. Non-nat. * Intermediate vs. Late * SRT	-0.03	[-0.07, 0.02]	0.02	-1.23	.22
Non-nat. vs. Foils * Early vs. Intermediate * SRT	0.01	[-0.04, 0.06]	0.02	0.38	.71
Non-nat. vs. Foils * Intermediate vs. Late * SRT	-0.04	[-0.08, 0.01]	0.03	-1.53	.13

Notes: Sequence type: Nat. = naturalistic sequences, Non-nat. = non-naturalistic sequences, Foils = unstructured foil sequences. Exposure phase: Early, Intermediate, Late. Model fit: AIC = 10,924; BIC = 11,103; $R^2_m = 0.225$; $R^2_c = 0.395$; ICC = 0.219; RMSE = 1.246; $\sigma = 1$.

Given that linguistic proficiency was related to performance across the experiment, we scrutinized this further by replacing the variable of block with the more interpretable variable of phase (*early*, *intermediate*, *late*). We fitted the maximal model supported by the data (Barr et al., 2013) with sequence type (sliding contrast: *naturalistic*: 0.5 vs. *non-naturalistic*: -0.5, and *non-naturalistic*: 0.5 vs. *foils*: -0.5), exposure phase (sliding contrast: *early exposure*: -0.5 vs. *intermediate exposure*: 0.5, and *intermediate exposure*: -0.5 vs. *late exposure*: 0.5), and SRT performance (centered continuous variable), as well as the interactions between the three factors, as fixed effects, and random intercepts and slopes for participants and items, where appropriate. The outcomes of the model analyzing the data at the syllable level are given in Table 5, and the outcomes of the model analyzing the data at the bigram level are given in Table 6. As in the previous models, we found significant main effects of sequence type (with syllable recall only differing between naturalistic and non-naturalistic sequences, and bigram recall additionally differing between non-naturalistic and foil sequences) and SRT performance. The main effect of

Table 6. Summary of the linear mixed-effects model investigating the influence of sequence type, exposure phase, and SRT score on the children's bigram recall

	Estimate	95% CI	SE	<i>t</i>	<i>p</i>
(Intercept)	-1.13	[-1.30, -0.94]	0.09	-12.97	< .001
Nat. vs. Non-nat.	0.19	[0.08, 0.29]	0.06	3.49	< .001
Non-nat. vs. Foils	0.14	[0.03, 0.25]	0.05	2.75	.006
Early vs. Intermediate	0.02	[-0.06, 0.10]	0.04	0.43	.67
Intermediate vs. Late	0.02	[-0.06, 0.10]	0.04	0.48	.63
SRT	0.57	[0.39, 0.73]	0.09	6.71	< .001
Nat. vs. Non-nat. * Early vs. Intermediate	0.02	[-0.07, 0.11]	0.04	0.48	.63
Nat. vs. Non-nat. * Intermediate vs. Late	0.03	[-0.06, 0.11]	0.04	0.67	.50
Non-nat. vs. Foils * Early vs. Intermediate	0.05	[-0.04, 0.16]	0.05	1.13	.26
Non-nat. vs. Foils * Intermediate vs. Late	-0.01	[-0.11, 0.10]	0.05	-0.13	.89
Nat. vs. Non-nat. * SRT	-0.08	[-0.17, 0.01]	0.04	-1.84	.07
Non-nat. vs. Foils * SRT	0.00	[-0.08, 0.09]	0.04	0.08	.93
Early vs. Intermediate * SRT	0.09	[0.00, 0.17]	0.04	1.96	.05
Intermediate vs. Late * SRT	-0.04	[-0.11, 0.04]	0.04	-0.89	.37
Nat. vs. Non-nat. * Early vs. Intermediate * SRT	0.03	[-0.06, 0.11]	0.04	0.60	.55
Nat. vs. Non-nat. * Intermediate vs. Late * SRT	-0.06	[-0.14, 0.02]	0.04	-1.37	.17
Non-nat. vs. Foils * Early vs. Intermediate * SRT	-0.06	[-0.16, 0.03]	0.05	-1.28	.20
Non-nat. vs. Foils * Intermediate vs. Late * SRT	0.01	[-0.09, 0.11]	0.05	0.20	.84

Notes: Sequence type: Nat. = naturalistic sequences, Non-nat. = non-naturalistic sequences, Foils = unstructured foil sequences. Exposure phase: Early, Intermediate, Late. Model fit: AIC = 5647; BIC = 5789; $R^2_m = 0.222$; $R^2_c = 0.254$; ICC = 0.041; RMSE = 0.599; $\sigma = 1$.

exposure phase was not significant. At the syllable level, we found two significant interactions. The interaction between SRT performance and sequence type indicated that children with lower SRT scores showed a greater recall difference between naturalistic and non-naturalistic sequences than children with higher SRT scores. The interaction between SRT performance and exposure phase indicated that children with higher SRT scores showed stronger improvement between the early and intermediate exposure phase than children with lower SRT scores (see Figure 6, which plots performance for “High” and “Low” SRT performers, determined by median split).

Children's performance showed a dramatic decrease in the final block, most likely driven by fatigue, which could potentially obscure learning effects across the later part of the experiment. To rule out this possibility, we ran one final exploratory analysis without Block 12, reducing the late exposure phase to Blocks 9 to 11, and fitted the same four models predicting participants' syllable or bigram recall by sequence type, block or exposure phase, and SRT performance (as well as their

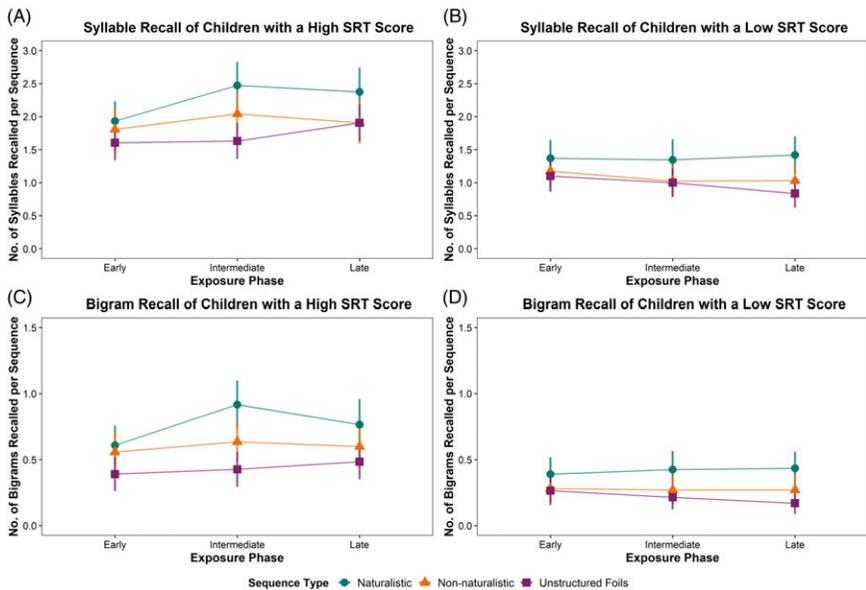


Figure 6. Mean syllable (top) and bigram (bottom) recall per sequence for the three sequence types (naturalistic, non-naturalistic, and unstructured foils) of the children with an SRT score higher than the median split (left) and lower than the median split (right), by exposure phase (early, intermediate, late). Error bars indicate standard error.

interactions). Importantly, all previously observed effects remained significant, with most effects becoming even slightly larger than in the original models including all 12 blocks. The two models including the factor block showed the same overall outcome whether Block 12 was included or not. In the model predicting bigram recall by sequence type, phase, and SRT performance, the two interactions already observed in the other models now also reached the traditional significance threshold. These were the interactions between sequence type and SRT performance, indicating that children with a lower SRT score showed a higher recall difference between naturalistic and non-naturalistic sequences than children with a higher SRT score ($\beta = -0.14$ [$-0.23, -0.03$], $t = -2.83$, $p = .005$), and the interaction between exposure phase and SRT performance, indicating that children with a higher SRT score improved more in the early stages of the experiment than children with a lower SRT score ($\beta = 0.09$ [$0.00, 0.18$], $t = 1.97$, $p = .049$). Finally, an interaction between sequence type, phase, and SRT performance became significant at the syllable level, indicating that the recall difference between naturalistic and non-naturalistic sequences decreased for children with higher SRT scores as compared to children with lower SRT scores toward the end of the experiment ($\beta = -0.05$ [$-0.10, 0.00$], $t = -2.02$, $p = .04$). Figure 7 in the Appendix illustrates high-SRT and low-SRT children's recall across the three exposure phases when Block 12 is removed.

Discussion

There is a growing body of evidence to suggest that SL is influenced by related prior knowledge – including phonotactic knowledge (e.g., Dal Ben et al., 2021; Finn & Hudson Kam, 2008; Mersad & Nazzi, 2011; Onnis et al., 2005), expectations about word length (Lew-Williams & Saffran, 2012), and knowledge of language structure (Thiessen et al., 2019). Since learners have been shown to form enduring memory representations for statistically defined input (see e.g., Batterink & Paller, 2017), it is conceivable that the influence of prior knowledge on SL also extends to prior knowledge of linguistic TPs (e.g., in the form of chunked syllables), obtained via exposure to natural language. There is promising early evidence for this possibility in adults (Elazar et al., 2022; Siegelman et al., 2018; Stärk et al., 2023); however, little is known about the way in which knowledge of syllable co-occurrences might influence SL in children. Here, we investigated whether seven- to nine-year-old German-speaking children used existing distributional knowledge to recall sequences of syllables and whether this was related to their language proficiency.

Children's recall was significantly better for naturalistic sequences (i.e., those which complemented the distributional properties of their native language) than non-naturalistic sequences, in line with our experimental hypothesis. This echoes the findings of related work with adults (Elazar et al., 2022; Stärk et al., 2023) and indicates that children too (in this case, seven- to nine-year-old native speakers of German) have well-entrenched memory traces of frequently co-occurring syllables, forged through experience with their native language and which shape their subsequent learning of related material (Siegelman et al., 2018). Past research has demonstrated that both infants and adults draw upon prior experience with language structure to guide their use of forward versus backward TPs when processing new input (Onnis & Thiessen, 2013; Thiessen et al., 2019). The present study goes further by demonstrating that children use the precise TPs previously encountered in their native language to both recall and hierarchically group (or chunk) syllable sequences. That is, we demonstrated that children had acquired the syllable co-occurrence patterns in natural German and implicitly brought these to bear to process new linguistic input.

Children also showed better recall for non-naturalistic sequences over unstructured foil sequences at the bigram level, demonstrating once more how children readily acquire new words from input regularities (Evans et al., 2009; Kidd et al., 2020; Saffran et al., 1997). The absence of the effect at the syllable level suggests that, while the children were sensitive to the in-experiment bigrams in the non-naturalistic condition, this advantage did not extend to a greater window as measured by the syllable recall measure. We suspect that the effect was not visible at the syllable level because of the difficulty of learning simultaneous sequences constructed from the same syllable inventory. Thus, with participants receiving sequences of all three sequence types simultaneously, the immediate advantage of the naturalistic sequences might have drawn participants' attention to these patterns first, leading to a disadvantage for any other learnable pattern in the input, such as in the non-naturalistic sequences (cf. Antovich & Graf Estes, 2018; Bulgarelli & Weiss, 2016; Gebhart et al., 2009).

Contrary to our expectations, at the group level, children's recall did not improve over the course of the task. Thus, we did not replicate the pattern of within-experiment learning observed in adults (cf. Stärk *et al.*, 2023). There are two potential explanations for this result. First, this could be explained by an implicit learning advantage for children relative to adults. That is, that the effect emerged early and did not change thereafter could be because children rapidly identified (parts of) the patterned sequences and did not learn anything more thereafter. There is some recent evidence in support of children's more rapid acquisition of statistical patterns in comparison to adults. Smalle *et al.* (2017) showed that nine- to ten-year-old children implicitly acquire phonotactic restrictions on novel words faster than adults do. Additional work by Smalle *et al.* (2018) using the Hebbian repetition paradigm suggests that eight- to nine-year-olds show greater retention of implicitly learnt syllable sequences than adults four hours and one week after they were first tested. Thus, it appears that, in repetition learning tasks at least, children can learn faster and retain verbal information better than adults, general differences in cognition like working-memory span notwithstanding. This finding is consistent with the generally held belief that there is a critical period for language learning (Hartshorne *et al.*, 2018; Newport, 1990). One tentative possibility is that the difference between our data and that of Stärk *et al.* (2023), where children seize upon the naturalistic sequences earlier, reflects children's greater ability to identify distributional patterns.

While possible, we think this explanation might be less likely than a second explanation that appeals both to the complex nature of the task and to individual differences in children's performance on the task. As outlined above, the fact that all three sequences came from the same syllable inventory introduced a complexity to the task that is not seen, for instance, in studies of Hebbian learning, which typically have patterned and foil sequences without syllable overlap, and which typically observe improvement in the repetition of the patterned sequence relative to foil repetition across time (e.g., Smalle *et al.*, 2018). This interference across sequence conditions may have prevented us from observing any interaction with block or phase, with the end result being that we only observed overall differences across the conditions.

Individual differences seem to have also impacted on learning across the experiment, with the children's language proficiency, as measured by the SRT task, interacting with learning over the course of the experiment. We observed a significant correlation between the SRT and performance in the naturalistic condition, even after partialling out variance attributable to foil repetition, suggesting that higher language proficiency was associated with better performance in that condition. Intriguingly, in our statistical models, we found that language proficiency interacted with the overall magnitude of the difference between the naturalistic and non-naturalistic conditions: children with lower SRT scores had a greater difference in sequence recall than children with higher SRT scores. One possible explanation for this effect is that the magnitude of the difference was smaller in the more proficient speakers because, in addition to learning the transitions in the naturalistic condition, they were also building more robust knowledge of the non-naturalistic sequences than were the children with lower proficiency. This interpretation is consistent with the fact that the high-SRT children showed a decreasing trend in their recall of the naturalistic sequences relative to the non-naturalistic sequences in the

latter part of the experiment, which emerged as an effect when we removed Block 12 from the analyses. If the children were also acquiring more robust knowledge of the non-naturalistic sequences during the latter half of the experiment, this newly acquired knowledge may have interfered with their performance in the naturalistic condition, given that the sequences contained the same syllables but different transitions. That is, the simultaneous learning of naturalistic and non-naturalistic sequences may have come at a price, with the acquisition of non-naturalistic sequences interfering with the naturalistic ones as their exposure to both sequences increased across time.

Our individual differences analyses revealed some suggestive patterns, but we stress that they should be interpreted with caution and treated as preliminary. Although our measures had good internal consistency (cf. Arnon, 2020), our sample size was not large and was further restricted due to circumstances outside of our control. Future studies could improve upon ours by recruiting a larger sample and by recruiting a larger age range. Language proficiency is correlated imperfectly with age, which may be due to the fact that the mechanisms underlying language acquisition vary across individuals in ways that are not completely age-dependent (Kidd & Donnelly, 2020; Kidd et al., 2018). Thus, more high-powered studies with wider age ranges may better elucidate the role of language proficiency in auditory SL of syllable sequences. At the moment, our data suggest that these individual differences exist and may be related to proficiency in nontrivial ways.

One final discussion point concerns the mechanism by which children are learning the syllable transitions in the task (and presumably, how they are learning the knowledge of syllable transitions they bring to the task). While the statistical structure of our sequences was defined in terms of TPs, we suspect that the process underlying learning is more likely to be the chunking of frequently co-occurring syllables into higher units – in our case, bisyllabic words – following models of SL and language that identify chunking as a basic learning mechanism (Christiansen & Chater, 2016; Frank et al., 2010; Perruchet, 2019; Perruchet & Vinter, 1998). One benefit of conceiving of SL as chunking is that it builds natural bridges to related areas of the literature on verbal learning (e.g., Isbilen et al., 2020; Jones, 2012; Jones et al., 2007), thus grounding the field within the broader cognitive domain of learning and memory. Tasks like serial recall are well-suited to exploring the overlap in these literatures since recall is sensitive to long-term knowledge and learning effects (e.g., Kidd et al., 2020; Smalle et al., 2018; Szewczyk et al., 2018).

Limitations and future directions

In the current study, we showed that seven- to nine-year-old children draw on their prior knowledge of syllable co-occurrences in their native language when processing new linguistic input, which has previously been found in adults (Elazar et al., 2022; Stärk et al., 2023). We also found that children's learning of attested and unattested syllable transitions was related to their language proficiency, as measured by sentence recall. The general conclusion from these data is that children form long-term representations for distributional information acquired over language development and use this knowledge to process new input. This skill may vary across individuals. We see these results as consistent with the general observation that language

acquisition involves attending to and inducing abstract knowledge from regularly occurring sequences of linguistic units (e.g., Arnon, 2021; Bannard & Matthews, 2008; Saffran *et al.*, 1996). However, we stress that we are not reducing acquisition purely to SL. From a very young age, children begin to build abstract knowledge at multiple levels of description. Thus, our demonstration that seven- to nine-year-olds are sensitive to distributional information does not entail that only this information is represented, but rather shows that this is one source of information that likely matters for learning. We note that almost every theory of language acquisition incorporates SL to some degree (e.g., Chang *et al.*, 2006; Lidz & Gagliardi, 2015; Pearl, 2021; Tomasello, 2003), but exactly how SL contributes in these theories differs substantially. An important future direction is accurately placing SL within the broader enterprise of language acquisition.

Conclusion

To conclude, we investigated the influence of children's prior distributional knowledge on their auditory SL performance and examined whether this effect was related to the children's language proficiency. Children drew upon their prior knowledge of TPs of their native language (i.e., highly frequent syllable co-occurrences) to process and learn new linguistic input, demonstrating that, like adults (Stärk *et al.*, 2023), German-speaking children had indeed developed entrenched knowledge of German syllable co-occurrences, which permeated through into the experiment to shape subsequent learning. In exploratory analyses, we found that children's performance on the SL task interacted with their language proficiency, with the results suggesting that children with higher proficiency were more sensitive to both the naturalistic and non-naturalistic patterned sequences. This is consistent with the idea that there are meaningful individual differences in SL that are related to language acquisition (Kidd *et al.*, 2018; Siegelman *et al.*, 2017), although significant additional work in this space is required to determine the exact nature of the effect.

Note

1 We included the following corpora from the CHILDES database (MacWhinney, 2000) in our analysis: Caroline (Von Stutterheim, 2010), Grimm (Grimm, 2006, 2007), Leo (Behrens, 2006), Manuela (Wagner, 2006), Miller (Miller, 1979), Rigol (Rigol, 2007), Stuttgart (Lintfert, 2010), TAKI (Lintfert, 2010), and Wagner (Wagner, 1974, 1985).

References

- Abed Ibrahim, L., Hamann, C., & Öwerdieck, D. (2018). *Identifying specific language impairment (SLI) across different bilingual populations: German sentence repetition task (SRT): Vol. Proceedings of the 42nd annual Boston University Conference on Language Development (BUCLD)* (A. B. Bertolini & M. J. Kaplan, Eds.; pp. 1–14). Cascadilla Press.
- Antovich, D. M., & Graf Estes, K. (2018). Learning across languages: Bilingual experience supports dual language statistical word segmentation. *Developmental Science*, *21*(2). <https://doi.org/10.1111/desc.12548>
- Arciuli, J. (2017). The multi-component nature of statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160058. <https://doi.org/10.1098/rstb.2016.0058>

- Arnon, I. (2020). Do current statistical learning tasks capture stable individual differences in children? An investigation of task reliability across modality. *Behavior Research Methods*, *52*(1), 68–81. <https://doi.org/10.3758/s13428-019-01205-5>
- Arnon, I. (2021). The Starting Big approach to language learning. *Journal of Child Language*, *48*(5), 937–958. <https://doi.org/10.1017/S0305000921000386>
- Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, *9*(4), 321–324. <https://doi.org/10.1111/1467-9280.00063>
- Audacity Team. (2018). *Audacity(R): Free audio editor and recorder (2.4.2)*. <https://audacityteam.org/>
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–248. <https://doi.org/10.1111/j.1467-9280.2008.02075.x>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Batterink, L. J., & Paller, K. A. (2017). Online neural monitoring of statistical learning. *Cortex*, *90*, 31–45. <https://doi.org/10.1016/j.cortex.2017.02.004>
- Behrens, H. (2006). The input–output relationship in first language acquisition. *Language and Cognitive Processes*, *21*(1–3), 2–24. <https://doi.org/10.1080/01690960400001721>
- Bonatti, L. L., Peña, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations: The role of consonants and vowels in continuous speech processing. *Psychological Science*, *16*(6), 451–459. <https://doi.org/10.1111/j.0956-7976.2005.01556.x>
- Bulgarelli, F., & Weiss, D. J. (2016). Anchors aweigh: The impact of overlearning on entrenchment effects in statistical learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1621–1631. <https://doi.org/10.1037/xlm0000263>
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234–272. <https://doi.org/10.1037/0033-295X.113.2.234>
- Christiansen, M. H., & Chater, N. (2016). The Now-or-Never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. <https://doi.org/10.1017/S0140525X1500031X>
- Cowan, N. (2016). Working memory maturation: Can we get at the essence of cognitive growth? *Perspectives on Psychological Science*, *11*(2), 239–264. <https://doi.org/10.1177/1745691615621279>
- Dal Ben, R., Souza, D. de H., & Hay, J. F. (2021). When statistics collide: The use of transitional and phonotactic probability cues to word boundaries. *Memory & Cognition*, *49*, 1300–1310. <https://doi.org/10.3758/s13421-021-01163-4>
- Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the “Tabula” is anything but “Rasa.” What determines performance in the auditory statistical learning task? *Cognitive Science*, *46*(2), e13102. <https://doi.org/10.1111/cogs.13102>
- 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), 321–335. [https://doi.org/10.1044/1092-4388\(2009/07-0189\)](https://doi.org/10.1044/1092-4388(2009/07-0189))
- Finn, A. S., & Hudson Kam, C. L. (2008). The curse of knowledge: First language knowledge impairs adult learners' use of novel statistics for word segmentation. *Cognition*, *108*(2), 477–499. <https://doi.org/10.1016/j.cognition.2008.04.002>
- Frank, M. C., Goldwater, S., Griffiths, T. L., & Tenenbaum, J. B. (2010). Modeling human performance in statistical word segmentation. *Cognition*, *117*(2), 107–125. <https://doi.org/10.1016/j.cognition.2010.07.005>
- Frost, R. L. A., Jessop, A., Durrant, S., Peter, M. S., Bidgood, A., Pine, J. M., Rowland, C. F., & Monaghan, P. (2020). Non-adjacent dependency learning in infancy, and its link to language development. *Cognitive Psychology*, *120*, 101291. <https://doi.org/10.1016/j.cogpsych.2020.101291>
- Gebhart, A. L., Aslin, R. N., & Newport, E. L. (2009). Changing structures in midstream: Learning along the statistical garden path. *Cognitive Science*, *33*(6), 1087–1116. <https://doi.org/10.1111/j.1551-6709.2009.01041.x>

- Green, P., & MacLeod, C. J.** (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Grimm, A.** (2006). *Intonational patterns and word structure in early child German*. (D. Bamman, T. Magnitskaia, & C. Zaller, Eds.; pp. 237–248).
- Grimm, A.** (2007). *The development of early prosodic word structure in child German: Simplex words and compounds*. [PhD Thesis]. University of Potsdam.
- Hamann, C., & Abed Ibrahim, L.** (2017). Methods for identifying specific language impairment in bilingual populations in Germany. *Frontiers in Communication*, *2*, 16. <https://doi.org/10.3389/fcomm.2017.00016>
- Hamann, C., Chilla, S., Ruigendijk, E., & Abed Ibrahim, L.** (2013). *A German sentence repetition task: Testing bilingual Russian/German children* [Poster].
- Hartshorne, J. K., Tenenbaum, J. B., & Pinker, S.** (2018). A critical period for second language acquisition: Evidence from 2/3 million English speakers. *Cognition*, *177*, 263–277. <https://doi.org/10.1016/j.cognition.2018.04.007>
- Hay, J. F., Pelucchi, B., Graf Estes, K., & Saffran, J. R.** (2011). Linking sounds to meanings: Infant statistical learning in a natural language. *Cognitive Psychology*, *63*(2), 93–106. <https://doi.org/10.1016/j.cogpsych.2011.06.002>
- Hebb, D. O.** (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Blackwell.
- Isbilen, E. S., McCauley, S. M., & Christiansen, M. H.** (2022). Individual differences in artificial and natural language statistical learning. *Cognition*, *225*, 105123. <https://doi.org/10.1016/j.cognition.2022.105123>
- Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H.** (2020). Statistically induced chunking recall: A memory-based approach to statistical learning. *Cognitive Science*, *44*(7), e12848. <https://doi.org/10.1111/cogs.12848>
- Johnson, P. C. D.** (2014). Extension of Nakagawa & Schielzeth's R2 GLMM to random slopes models. *Methods in Ecology and Evolution*, *5*(9), 944–946. <https://doi.org/10.1111/2041-210X.12225>
- Jones, G.** (2012). Why chunking should be considered as an explanation for developmental change before short-term memory capacity and processing speed. *Frontiers in Psychology*, *3*. <https://doi.org/10.3389/fpsyg.2012.00167>
- Jones, G., Gobet, F., & Pine, J. M.** (2007). Linking working memory and long-term memory: A computational model of the learning of new words. *Developmental Science*, *10*(6), 853–873. <https://doi.org/10.1111/j.1467-7687.2007.00638.x>
- Kail, R., & Salthouse, T. A.** (1994). Processing speed as a mental capacity. *Acta Psychologica*, *86*(2–3), 199–225. [https://doi.org/10.1016/0001-6918\(94\)90003-5](https://doi.org/10.1016/0001-6918(94)90003-5)
- Kidd, E., & Arciuli, J.** (2016). Individual differences in statistical learning predict children's comprehension of syntax. *Child Development*, *87*(1), 184–193. <https://doi.org/10.1111/cdev.12461>
- Kidd, E., & Donnelly, S.** (2020). Individual differences in first language acquisition. *Annual Review of Linguistics*, *6*(1), 319–340. <https://doi.org/10.1146/annurev-linguistics-011619-030326>
- Kidd, E., Arciuli, J., Christiansen, M. H., Isbilen, E. S., Revius, K., & Smithson, M.** (2020). Measuring children's auditory statistical learning via serial recall. *Journal of Experimental Child Psychology*, *200*, 104964. <https://doi.org/10.1016/j.jecp.2020.104964>
- Kidd, E., Donnelly, S., & Christiansen, M. H.** (2018). Individual differences in language acquisition and processing. *Trends in Cognitive Sciences*, *22*(2), 154–169. <https://doi.org/10.1016/j.tics.2017.11.006>
- Kuhl, P. K.** (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, *5*(11), 831–843. <https://doi.org/10.1038/nrn1533>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B.** (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lany, J.** (2014). Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child Development*, *85*(4), 1727–1739. <https://doi.org/10.1111/cdev.12199>
- Lany, J., & Saffran, J. R.** (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science*, *21*(2), 284–291. <https://doi.org/10.1177/0956797609358570>
- Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S.** (2015). *Peabody picture vocabulary test: Revision IV (German adaptation)*.

- Lew-Williams, C., & Saffran, J. R. (2012). All words are not created equal: Expectations about word length guide infant statistical learning. *Cognition*, *122*(2), 241–246. <https://doi.org/10.1016/j.cognition.2011.10.007>
- Lidz, J., & Gagliardi, A. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annual Review of Linguistics*, *1*(1), 333–353. <https://doi.org/10.1146/annurev-linguist-030514-125236>
- Lintfert, B. (2010). *Phonetic and phonological development of stress in German* [PhD Thesis]. University of Stuttgart.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Lawrence Erlbaum Associates.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, *22*(3), 276.
- Mersad, K., & Nazzi, T. (2011). Transitional probabilities and positional frequency phonotactics in a hierarchical model of speech segmentation. *Memory & Cognition*, *39*(6), 1085–1093. <https://doi.org/10.3758/s13421-011-0074-3>
- Miller, M. (1979). *The logic of language development in early childhood*. Springer-Verlag.
- Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. <https://doi.org/10.1111/j.2041-210x.2012.00261.x>
- Neurobehavioral Systems, Inc. (2014). *Presentation* (18.0). www.neurobs.com
- Newport, E. L. (1990). Maturation constraints on language learning. *Cognitive Science*, *14*(1), 11–28. https://doi.org/10.1207/s15516709cog1401_2
- Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268–284. <https://doi.org/10.1016/j.cognition.2012.10.008>
- Onnis, L., Monaghan, P., Richmond, K., & Chater, N. (2005). Phonology impacts segmentation in online speech processing. *Journal of Memory and Language*, *53*(2), 225–237. <https://doi.org/10.1016/j.jml.2005.02.011>
- Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737–3753. <https://doi.org/10.1098/rstb.2009.0173>
- Pearl, L. (2021). Theory and predictions for the development of morphology and syntax: A Universal Grammar + statistics approach. *Journal of Child Language*, *48*(5), 907–936. <https://doi.org/10.1017/S0305000920000665>
- Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535. <https://doi.org/10.1111/tops.12403>
- Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263. <https://doi.org/10.1006/jmla.1998.2576>
- R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Raviv, L., & Arnon, I. (2018). The developmental trajectory of children's auditory and visual statistical learning abilities: Modality-based differences in the effect of age. *Developmental Science*, *21*(4), e12593. <https://doi.org/10.1111/desc.12593>
- Rigol, R. (2007). German Rigol corpus [Data set]. *TalkBank*. <https://doi.org/10.21415/t50s34>
- Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *WIREs Cognitive Science*, *1*(6), 906–914. <https://doi.org/10.1002/wcs.78>
- RStudio Team. (2022). *RStudio: Integrated Development Environment for R*. RStudio, PBC. <http://www.rstudio.com/>.
- Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. <https://doi.org/10.1126/science.274.5294.1926>
- Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A., & Barrueco, S. (1997). Incidental language learning: Listening (and learning) out of the corner of your ear. *Psychological Science*, *8*(2), 101–105. <https://doi.org/10.1111/j.1467-9280.1997.tb00690.x>
- Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science*, *42*(8), 3100–3115. <https://doi.org/10.1111/cogs.12692>

- Siegelman, N., Bogaerts, L., Christiansen, M. H., & Frost, R. (2017). Towards a theory of individual differences in statistical learning. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1711), 20160059. <https://doi.org/10.1098/rstb.2016.0059>
- Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, 177, 198–213. <https://doi.org/10.1016/j.cognition.2018.04.011>
- Smalle, E. H. M., Muylle, M., Szmalec, A., & Duyck, W. (2017). The different time course of phonotactic constraint learning in children and adults: Evidence from speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1821–1827. <https://doi.org/10.1037/xlm0000405>
- Smalle, E. H. M., Page, M. P. A., Duyck, W., Edwards, M., & Szmalec, A. (2018). Children retain implicitly learned phonological sequences better than adults: A longitudinal study. *Developmental Science*, 21(5), e12634. <https://doi.org/10.1111/desc.12634>
- Stärk, K., Kidd, E., & Frost, R. L. A. (2022). Word segmentation cues in German child-directed speech: A corpus analysis. *Language and Speech*, 65(1), 3–27. <https://doi.org/10.1177/0023830920979016>
- Stärk, K., Kidd, E., & Frost, R. L. A. (2023). Close encounters of the word kind: Attested distributional information boosts statistical learning. *Language Learning*, 71(2). <https://doi.org/10.1111/lang.12523>
- Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, 179, 23–36. <https://doi.org/10.1016/j.cognition.2018.06.002>
- Thiessen, E. D., Onnis, L., Hong, S.-J., & Lee, K.-S. (2019). Early developing syntactic knowledge influences sequential statistical learning in infancy. *Journal of Experimental Child Psychology*, 177, 211–221. <https://doi.org/10.1016/j.jecp.2018.04.009>
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.
- Trecca, F., McCauley, S. M., Andersen, S. R., Bleses, D., Basbøll, H., Højen, A., Madsen, T. O., Ribu, I. S. B., & Christiansen, M. H. (2019). Segmentation of highly vocalic speech via statistical learning: Initial results from Danish, Norwegian, and English. *Language Learning*, 69(1), 143–176. <https://doi.org/10.1111/lang.12325>
- Von Stutterheim, C. (2010). German Caroline corpus [Data set]. *TalkBank*. <https://doi.org/10.21415/t5ns5s>
- Wagner, K. R. (1974). *Die Sprechsprache des Kindes. Teil 1: Theorie und Analyse*. Präger.
- Wagner, K. R. (1985). How much do children say in a day? *Journal of Child Language*, 12(2), 475–487. <https://doi.org/10.1017/S0305000900006565>
- Wagner, M. (2006). *First steps to communication: A pragmatic analysis*. Narr Verlag.
- West, G., Melby-Lervåg, M., & Hulme, C. (2021). Is a procedural learning deficit a causal risk factor for developmental language disorder or dyslexia? A meta-analytic review. *Developmental Psychology*, 57(5), 749–770. <https://doi.org/10.1037/dev0001172>
- Wickham, H. (2017). *tidyverse: Easily install and load the 'Tidyverse'*. R package version 1.2.1. <https://CRAN.R-project.org/package=tidyverse>.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

Appendix

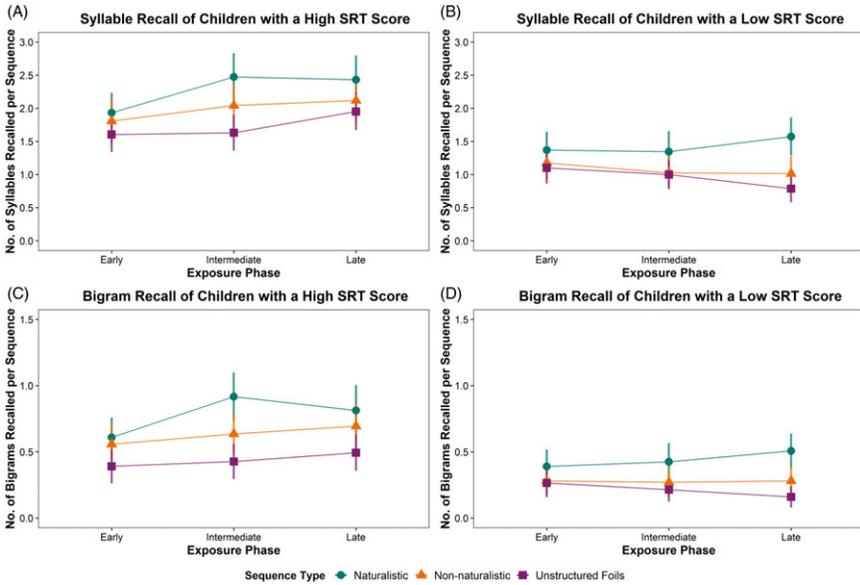


Figure 7. Mean syllable (top) and bigram (bottom) recall per sequence for the three sequence types (naturalistic, non-naturalistic, and unstructured foils) of the children with an SRT score higher than the median split (left) and lower than the median split (right), by exposure phase (early, intermediate, late) when removing the final block (late exposure phase = Blocks 9-11). Error bars indicate standard error.

Cite this article: Stärk, K., Kidd, E., and Frost, RLA. (2022). The effect of children’s prior knowledge and language abilities on their statistical learning. *Applied Psycholinguistics* 43, 1045–1071. <https://doi.org/10.1017/S0142716422000273>