

BRIEF RESEARCH REPORT

Children's simultaneous or successive acquisition of vocabulary and grammar: Evidence from cross-situational learning

Wensi Zhang¹ , Padraic Monaghan² , Sophie Bennett¹  and Patrick Rebuschat^{1,3} 

¹Department of Linguistics and English Language, Lancaster University, Lancaster, UK, ²Department of Psychology, Lancaster University, Lancaster, UK and ³Faculty of Science, University of Tübingen, Tübingen, Germany

Corresponding author: Wensi Zhang; Email: w.zhang31@lancaster.ac.uk

(Received 11 October 2024; revised 03 June 2025; accepted 10 June 2025)

Abstract

Recent evidence from cross-situational learning (CSL) studies have shown that adult learners can acquire words and grammar simultaneously when sentences of the novel language co-occur with dynamic scenes to which they refer. Syntactic bootstrapping accounts suggest that grammatical knowledge may help scaffold vocabulary acquisition by constraining possible meanings, thus, for children, words and grammar may be acquired at different rates. Twenty children (ages 8 to 9) were exposed in a CSSL study to an artificial language comprising nouns, verbs, and case markers occurring within a verb-final grammatical structure. Children acquired syntax (i.e., word order) effectively, but we found no evidence of vocabulary learning, whereas previous adult studies showed learning of both from similar input. Grammatical information may thus be available early for children, to help constrain and support later vocabulary learning. We propose that gradual maturation of declarative memory systems may result in more effective vocabulary learning in adults.

Keywords: child language acquisition; cross-situational word learning; statistical learning; syntactic bootstrapping

1. Introduction

In early language acquisition, a key challenge is to determine the exact referent from the infinite number of possible ones when a word is heard in speech. This is often referred to as the “Gavagai” problem, following Quine (1960). Imagine an infant or child hearing the utterance “Gavagai!” while observing a landscape in which a rabbit is dashing across a field. In this case, the utterance might refer to multiple referents, including the whole rabbit, the rabbit’s ear, the texture of its fur, or its movement. How do infants and children know what “Gavagai” refers to? One underlying process that researchers (e.g., Fazly et al., 2010; Monaghan et al., 2021; Yu & Smith, 2007) have long assumed to support learning in such ambiguous situations is statistical learning (SL), by which learners acquire regularities in the language patterns through exposure. Specifically, for word learning under

© The Author(s), 2025. 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.

conditions with referential ambiguity, learners can track the co-occurrence of the word with its referent in the environment over multiple situations (Rebuschat *et al.*, 2021; Yu & Smith, 2007), in order to establish the intended word–referent mapping.

Learning word–object mappings through a cross-situational learning (CSL) paradigm has been found to be rapid and successful in infants (e.g., Smith & Yu, 2008), children (e.g., 4- to 7-year-olds, Benitez *et al.*, 2020; 4- and 10-year-olds, Fitneva & Christiansen, 2017; 5- to 7-year-olds, Suanda *et al.*, 2014; 2- to 7-year-olds, Venker, 2019; 2- to 5-year-olds, Vlach & DeBrock, 2017), and adults (e.g., Yu & Smith, 2007). In these previous experiments, participants were exposed to sets of pseudowords while observing multiple objects. Within a single trial, it was not possible to correctly map a noun to its referent due to the ambiguity of possible correspondences between words and objects. However, the appropriate word–object mapping can be determined through tracking cross-situational statistics, as each word consistently appeared with its referent while the other words and objects varied over trials. Learners can use CSL to acquire words from other grammatical categories (for verb learning, see Childers *et al.*, 2023; Scott & Fisher, 2012; for adjective learning, see Akhtar & Montague, 1999), and in the case of adults, they can rely on CSL to acquire multiple grammatical categories simultaneously (e.g., Monaghan *et al.*, 2015; Rebuschat *et al.*, 2021).

However, the typical CSL paradigm, which focuses on words from a single grammatical category, all occurring with their referents, is a simplification that does not apply to naturalistic language, where scenes and utterances are more complex (e.g., Yu & Ballard, 2007). Furthermore, there is a close interdependence in the meaning and grammatical roles of words within an utterance (Fisher *et al.*, 2010; Gleitman, 1990; Monaghan *et al.*, 2023). This bootstrapping of word learning and syntax has been a topic of theoretical interest in language acquisition research (Abend *et al.*, 2017; Höhle & Weissenborn, 2001), but a clear demonstration of how words and syntax are acquired simultaneously has been rarely observed. One reason for this is practical: learning a language with sufficient complexity to incorporate both vocabulary from different grammatical categories and syntactic structure is a substantial challenge. The rare exceptions, however, tend to pretrain learners on vocabulary, or using already known words, before exposing them to multi-word sentences (e.g., Amato & MacDonald, 2010; Friederici *et al.*, 2002; Hu, 2017; Morgan-Short *et al.*, 2014; Spit *et al.*, 2022).

The possibility of simultaneous acquisition of vocabulary and grammar without prior vocabulary training was recently demonstrated in studies with adults (e.g., Monaghan *et al.*, 2019, 2021; Rebuschat *et al.*, 2021; Walker *et al.*, 2020), where adult learners were exposed to a complex artificial language consisting of transitive sentences presented alongside dynamic scenes relating to the sentences. However, whether this CSL is accessible to children learning a language remains unclear.

Vlach and DeBrock (2017) investigated multiple factors that underlie effective CSL of nouns in children aged 2 to 5. They found that children’s declarative memory ability (i.e., visual and auditory recognition memory) and language skills (i.e., receptive vocabulary) were strong predictors of learning. Vocabulary learning has been related to declarative memory ability (Ruiz *et al.*, 2018; Ullman, 2004; Walker *et al.*, 2020), consistent with Vlach and DeBrock’s (2017) finding that it relates to CSL noun learning. However, learning syntax has been related to procedural memory ability (Morgan-Short *et al.*, 2014). The developmental trajectories of these memory systems are very different (Pili-Moss, 2021). Procedural memory tends to reach maturity at an earlier stage of the life span (i.e., infancy and early childhood), while declarative memory matures more slowly, starting to develop in childhood and not becoming fully functional until early adulthood (Bauer, 2008; Hulstijn, 2015; Ullman, 2004). Lum *et al.* (2010) found that procedural

memory skills were relatively stable from ages 5 to 6 years, though declarative memory performance was changing substantially, and a similar pattern was found for ages 6 to 10 by Finn et al. (2016), who found procedural memory skills were similar to adults by these ages, though declarative memory skills were significantly lower. Thus, when children younger than 10 years old are faced with learning both vocabulary and syntax from a novel language, we might expect syntax, served by the procedural memory system, to be acquired more effectively, with greater variation in vocabulary learning possible due to expression of individual differences in the development of declarative memory.

In the current study, we investigate the effect that the complex environment that children experience, where there are multiple words in sentences and many possible referents around them in the environment, has upon children's simultaneous learning of vocabulary and syntax. We exposed children aged 8 to 9 years old to the complex utterances and complex scenes adapted from Rebuschat et al. (2021). We focused our investigation on children aged 8 to 9 years to meet the gap in statistical language learning studies of children of this age (Isbilen & Christiansen, 2022), and also because of this being a point in development where divergence in procedural and declarative memory skills development is observed (Ferman & Karni, 2010; Finn et al., 2016; Lum et al., 2010; Meulemans et al., 1998).

We predicted that children would be able to learn the sentence–scene correspondences, due to their ability to track cross-situational statistics for simpler word–referent mappings (e.g., Childers et al., 2023; Scott & Fisher, 2012; Vlach & DeBrock, 2017). We also predicted that they would be able to acquire syntax from cross-situational statistics, as this was readily acquired by adults (e.g., Rebuschat et al., 2021) and supported by the earlier maturing procedural memory system, providing insight into the bootstrapping process of acquisition of vocabulary and syntax.

2. Method

2.1. Participants

Twenty participants (mean age = 9.1 years, $SD = 4$ months, 13 female) aged from 8;11 to 9;10 at a primary school in Greater Manchester, UK, participated in this study. The school is in a moderate socio-economic area (within the 40% least deprived areas in England, English Indices of Deprivation, 2019) with a relatively high education level: 65% of the parents held qualifications above university undergraduate degree, 25% had completed education up to school or college (up to age 18), and 10% held a FE college diploma. Seventeen participants were monolingual native speakers of English, two spoke English and Urdu, and one spoke English, Russian, and Portuguese.¹ All had normal vision and hearing.

2.2. Materials

Artificial language: The artificial language was adapted from Rebuschat et al. (2021), with some simplifications to make the language potentially easier to acquire. Specifically, we excluded the adjectives in the artificial language in Rebuschat et al. (2021).

Vocabulary: There were 12 pseudowords, 10 of which were content words (6 nouns and 4 verbs) and 2 of which were case markers indicating the grammatical role (i.e., either subject

¹Note that Urdu is a verb-final language just like the artificial language used in this study. For this reason, we conducted separate analyses for the two Urdu–English bilingual children in our sample. These are reported in the [Supplementary Materials](#).

or object) of the preceding noun. These words were read and recorded separately by a female native English speaker in monotone and presented using E-prime with a 250 ms pause between each word. The pseudowords can be found in the [Supplementary Materials](#).

Syntax: The syntax of the artificial language was based on Japanese, with a fixed position of the verb phrase (VP), always appearing at the end of the sentence, while the subject noun phrase (NP) and object NP could alternate between the initial and second positions. The VP contained only the verb, whereas the NPs always comprised a noun followed by a post-nominal case marker, which reliably indicated whether the preceding noun was the agent or the patient in the sentence. Half of the sentences were in SOV order, and the other half were OSV. A total of 112 unique sentences were generated by E-prime by concatenating pseudowords, with a 250 ms pause between each word, as demonstrated in the following example.

- (1) Cheelow tha bimdah noo dingep.
 Animal₁ SUBJECT Animal₂ OBJECT pushes.
 ‘Animal₁ pushes animal₂’.

We balanced the frequency of vocabulary, subject and object assignment, and word order across blocks.

Visual stimuli: The visual stimuli used in the current study consisted of a series of animated scenes generated by E-prime (2.0 Psychology Software Tools, Pittsburgh, PA). In these scenes, six cartoon animals (elephant, cow, chicken, turtle, zebra, and owl) were selected as the referents of the six nouns in the artificial language, performing one of the four actions (hiding, jumping, lifting, and pushing) as the referents of the four verbs. We randomly allocated the mapping of the words to the animal characters and actions for each participant to avoid the association of certain sounds to objects and actions (Rebuschat *et al.*, 2021). The animal characters are presented in the [Supplementary Materials](#).

2.3. Procedure

Parental questionnaires and consent forms were distributed and collected by the deputy headteacher before the experiment. Children were then trained and tested on the artificial language on two consecutive days, with each session lasting about 30 minutes. The procedure was identical on each day.

Cross-situational learning task: Participants were told they would learn an alien language spoken by “friends from a distant planet.” Two practice trials were then presented in which participants had a chance to familiarize themselves with the task, observing two animated scenes (see [Figure 1](#) for example), listening to artificial language

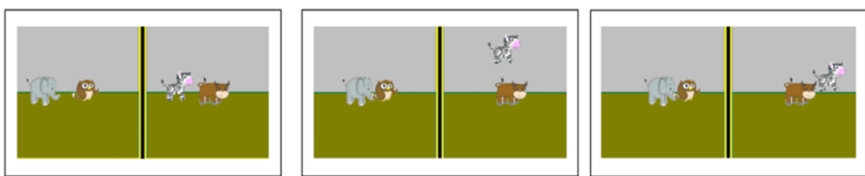


Figure 1. Example of a training trial of the CSL task, illustrating screenshots of the animated scenes. The left scene in this example trial depicts an elephant (agent) pushing an owl (patient), and the right scene shows a zebra (agent) jumping over a cow (patient).

sentences (e.g., *Cheelow tha bimdah noo dingep.*), and responding by pressing the keys on the keyboard associated with which scene they felt matched the sentence. An “L” sticker for the left scene and an “R” sticker for the right scene were placed over the keys “1” and “2,” respectively, on a computer keyboard. Animal characters and pseudowords contained in practice trials were not included in the main part of the study.

The CSL task on each day comprised six blocks. Each block contained 12 training trials, different for each block, but balanced such that the occurrence of each word and visual stimulus occurred an equal number of times. A total of 72 sentences were used across the six training blocks, with 12 sentences in each block.

In each trial of this task, children observed two animated scenes, each depicting two of the six animal characters performing one of four actions. The two scenes differed in terms of the animals, the actions, and the agent and patient roles of the animals performing the action present in each scene. After the action was displayed once, the corresponding sentence was played. For example, children might see, on the left screen, an elephant jumping over an owl, and on the right a zebra hiding behind a cow, while hearing the artificial sentence “*Cheelow tha bimdah noo dingep*” (see Figure 1 for example). The actions were then repeated in a loop until participants entered their response, at which point the experiment advanced to the next trial without delay. After responding, participants received no feedback, with the task proceeding to the next trial immediately after a response was made.

In blocks 3 and 6, the 12 training trials were intermingled with 28 vocabulary testing trials, which were again balanced in terms of the occurrence of words and visual stimuli. In each block, there were six test trials for nouns, four for verbs, and four for marker words. The test trials were identical to the training trials, except that the two scenes were identical to one another except for one feature. To test nouns, the scenes varied by one animal; to test verbs, the scenes varied by the action only; and to test the marker words, the scenes varied in terms of which animal was the agent and which the patient of the action. Each noun was tested once per block, each verb once, and the marker words were tested four times.

Grammaticality judgement task: After the final block for the CSL task, there was a Grammaticality Judgement Task (GJT) block, which tested the acquisition of syntax (in terms of knowledge of word order). This consisted of 12 trials, each of which comprised one sentence occurring with one scene. All words corresponded with their referents in the scene, but in half the sentences, the syntax of the language was followed (either OSV or SOV), while the remainder contained syntactic violations, with word order either VSO, VOS, SVO, or OVS. There were three verb-initial trials, and three verb-medial trials. None of the testing sentences were used in training trials.

Participants were informed that sentences would now be spoken by an alien from another planet who was also learning this “alien language,” and they had to determine whether the sentence sounded “good” or “funny” based on the sentences they had heard earlier. One of the researchers clarified “good” and “funny” to each participant one by one to make sure children understood that this referred to whether it sounded like or unlike the previous sentences. A label with “good” covered the 9 and a label “funny” covered the 0 key on a computer keyboard.

After completing the study on Day 1, children returned the following day for the second training and test session at approximately the same time of day. Note that the vocabulary tests were included twice per session, and the GJT at the end of the training session. This design was so that potential subtle differences in the order of acquisition of vocabulary items (e.g., nouns learned before or after verbs) could be tracked through the training, with the vocabulary test trials not interrupting exposure to the language because

these test trials were identical in form to the training trials. The GJT required a different presentation and response, and so was positioned after the end of training on each day so as not to disrupt the learning.

3. Results

As noted above, two of our participants were Urdu–English bilingual children. Since Urdu is an SOV language, we conducted separate analyses to find out if this affected the results. The analyses can be found in the [Supplementary Materials](#), and the results are similar to those presented here for the whole group of participants.

3.1. Performance on the cross-situational learning task

Accuracy on the training trials: The descriptive statistics of performance across training trials on two consecutive days (6 blocks each) are displayed in [Table 1](#) and [Figure 2](#). First, in descriptive statistics, whether performance across training trials was greater than chance (0.5) was determined by one-sample *t* tests (see [Table 1](#)). For training on Day 1, we observed significant learning effects over blocks 1, 4, and 5, and marginally significant learning effect over blocks 3 and 6. However, for Day 2, a learning effect was only found in block 2.

To further investigate whether learning was affected by training block and day, generalized linear mixed-effects modelling was used with accuracy (0 or 1) as the dependent variable, employing the binomial logistic function. We tested potential non-linear effects of learning using orthogonal polynomials for the block predictor. We constructed mixed-effects models starting with a null model predicting accuracy with the binomial logit link function. Models included random intercepts for participant and item. However, due to singularity, the random slopes for the within-participant predictors were not included. We then compared this model against models that incrementally added fixed effects: linear (ot1), quadratic (ot2), and cubic (ot3) polynomial terms for Block, then adding day, and their interaction. Log-likelihood comparisons were used to test whether each of the fixed effects improved model fit. Results showed that the intercept

Table 1. Descriptive statistics for training trials in CSL task in the six blocks on two days. Showing *t*-test values compared against chance performance

	Block											
	Day 1						Day 2					
	1	2	3	4	5	6	1	2	3	4	5	6
M	0.57	0.54	0.55	0.56	0.59	0.56	0.51	0.58	0.53	0.46	0.55	0.55
SD	0.14	0.14	0.13	0.11	0.15	0.14	0.18	0.15	0.19	0.22	0.17	0.16
<i>t</i>	2.20	1.21	1.79	2.52	2.50	2.04	0.21	2.31	0.70	−0.79	1.29	1.29
<i>p</i>	0.040	0.243	0.090	0.021	0.022	0.056	0.834	0.032	0.491	0.439	0.212	0.212
<i>d</i>	0.49	0.27	0.40	0.56	0.56	0.45	0.05	0.52	0.16	−0.18	0.29	0.29

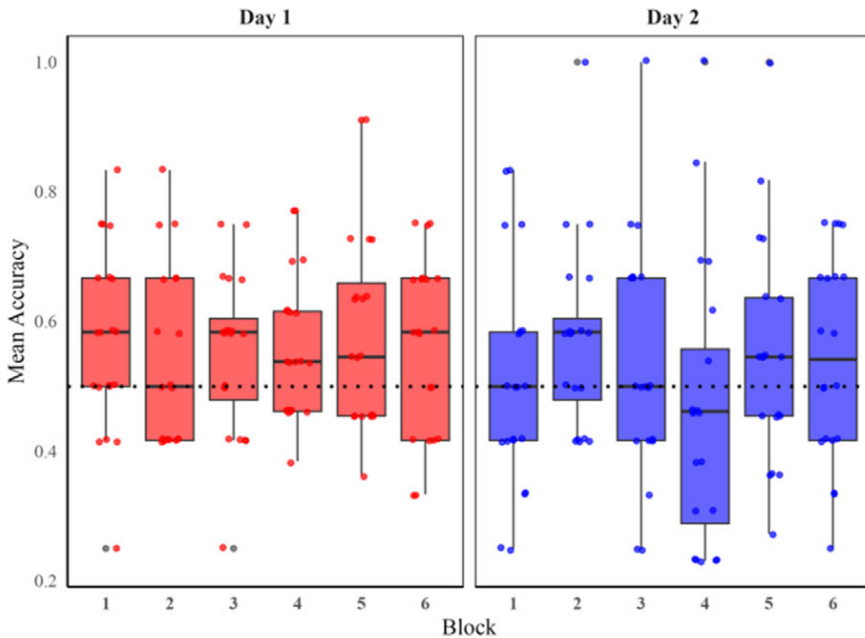


Figure 2. Accuracy for training trials in CSL task for days 1 and 2. The box indicates the median (horizontal line) and interquartile range, with dots indicating individuals' accuracy. Dotted line indicates chance level (0.5).

was significant, such that accuracy was slightly above chance, averaged across all training trials, estimate = 0.184, SE = 0.076, $z = 2.42$, $p = 0.016$, odds ratio = 1.20. Adding the fixed effects of ot1 ($\chi^2(1) = 0.12$, $p = 0.727$), ot2 ($\chi^2(1) = 0.51$, $p = 0.476$) and ot3 ($\chi^2(1) = 0.17$, $p = 0.683$) did not significantly improve model fit. However, adding the fixed effect of day ($\chi^2(1) = 3.43$, $p = 0.064$) showed a marginally significantly improved model fit, suggesting slightly lower accuracy on Day 2 compared with Day 1. The interaction between block and day was not significant ($\chi^2(1) = 0.67$, $p = 0.414$), indicating no evidence for changing pace of learning on Day 2 compared to Day 1.

For the performance across test trials (i.e., vocabulary and syntax tests), in our descriptive statistics we conducted one-sample t tests to determine whether the accuracy of vocabulary (i.e., nouns, verbs, and markers) and syntax (i.e., word order) test trials in each block was significantly above chance level. The descriptive statistics and results are presented in Tables 2 and 3 and Figure 3.

Accuracy on the test trials: For vocabulary acquisition, results from the t tests showed no evidence of a learning effect of nouns, verbs, and markers in all vocabulary tests (see Table 2 for test results). To determine whether block and day had a significant effect on vocabulary acquisition for nouns, verbs, and markers, we used generalized linear mixed-effects models similar to the analysis conducted on training trials. The models included block and day as fixed effects, and the maximal random-effects structure that allowed models to converge without singularity warnings. Specifically, for noun test trials, a random intercept for item was included; for verb test trials, a random intercept for participant was included; for models testing marker test trials, random intercepts for participant and item were included. Random slopes for block and day were not included due to the singularity warnings. Results indicated that neither the fixed effect of block nor

Table 2. Descriptive statistics for vocabulary test trials in CSL task in block 3 and 6 on days 1 and 2. T-test values are compared against chance

		Blocks			
		Day 1		Day 2	
		3	6	3	6
<i>Nouns</i>	<i>M</i>	0.54	0.52	0.48	0.52
	<i>SD</i>	0.22	0.22	0.22	0.22
	<i>t</i>	0.86	0.32	−0.32	0.32
	<i>p</i>	0.398	0.741	0.733	0.741
	<i>d</i>	0.19	0.07	−0.08	0.07
<i>Verbs</i>	<i>M</i>	0.54	0.54	0.56	0.53
	<i>SD</i>	0.31	0.25	0.25	0.26
	<i>t</i>	0.55	0.68	1.10	0.44
	<i>p</i>	0.591	0.505	0.287	0.666
	<i>d</i>	0.12	0.15	0.25	0.10
<i>Markers</i>	<i>M</i>	0.51	0.54	0.49	0.54
	<i>SD</i>	0.24	0.31	0.26	0.27
	<i>t</i>	0.24	0.55	−0.21	0.62
	<i>p</i>	0.815	0.591	0.834	0.545
	<i>d</i>	0.05	0.12	−0.05	0.14

Table 3. Descriptive statistics for syntax test trials in CSL task on days 1 and 2. *t*-test values are compared against chance

	Day 1	Day 2
<i>M</i>	0.58	0.62
<i>SD</i>	0.15	0.22
<i>t</i>	2.30	2.42
<i>p</i>	0.033	0.026
<i>d</i>	0.51	0.54

day improved the model fit for any of the vocabulary types (model results can be found in [Supplementary Materials](#)).

For syntax (word order) acquisition, *t* test results showed significant learning effects on both Day 1 and Day 2 (see [Table 3](#) for test results). To determine whether day had a significant effect on syntax acquisition, we again tested generalized linear mixed-effects models in the same way as for the vocabulary test trials, except only with day as a predictor for accuracy (note that syntax was only tested once per day so block was not relevant to this analysis). The

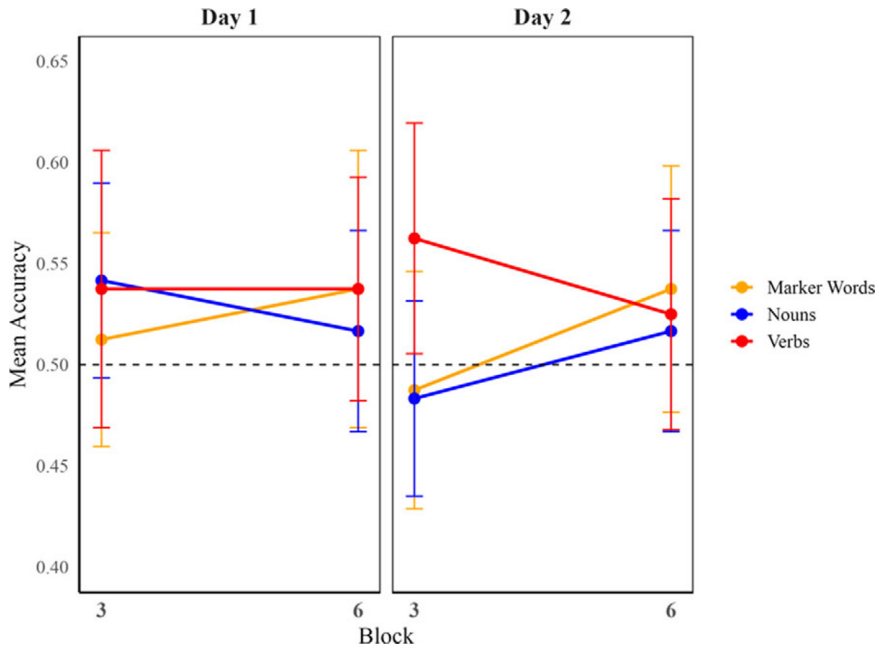


Figure 3. Mean accuracy performance by vocabulary test trials in CSL task, for Day 1 (left) and Day 2 (right). Error bars represent the standard error of the mean for each block. The dotted horizontal line at 0.5 indicates chance.

intercept of the model with no fixed effect was significant (estimate = 0.427, SE = 0.156, $z = 2.74$, $p = 0.006$, odds ratio = 1.53), indicating that the overall accuracy in the syntax test was greater than chance. However, we did not find that adding day significantly improved fit ($\chi^2(1) = 0.75$, $p = 0.387$).

4. Discussion

The current study explored what features of an artificial language children aged 8 to 9 years old can acquire through CSL. Our results indicated, for the first time, that when acquiring a novel language comprising previously unknown syntax and vocabulary, the conjunction of sentences and scenes to which they refer was navigable by children in order to effectively acquire the syntax (i.e., word order knowledge). However, in contrast, there was evidence that children had only just begun to make a start on learning individual vocabulary items. Whereas overall performance on training trials (which required the coordination of words in the sentences and referents in the scenes) was above chance, accuracies for the individual vocabulary types (i.e., nouns, verbs, and case markers) were not found to be significantly above chance.

When referential ambiguity is minimized, i.e., restricted to only noun-object and verb-action mappings, children are able to learn vocabulary with relative ease (Childers et al., 2023; Scott & Fisher, 2012; Smith & Yu, 2008; Vlach & DeBrock, 2017). Furthermore, when vocabulary is pre-trained, children can also respond to the syntax of the language (Spit et al., 2022). However, in our study, when the learning environment mimicked the natural language situation by incorporating greater syntactic complexity and referential

ambiguity, we found that acquiring the vocabulary through cross-situational statistics was a challenge for children. In previous studies with adults, acquiring both the syntax (word order) and the vocabulary simultaneously during learning was found to be possible (Monaghan *et al.*, 2021; Rebuschat *et al.*, 2021; Walker *et al.*, 2020). In order to ascertain whether there were quantitative differences in learning between adults and children, we compared the children's performance in the current study to the adult "implicit" condition from Monaghan *et al.* (2021). This condition was an adult study that was qualitatively similar to the current study, though note that the language was somewhat more complex because it also contained adjectives and more vocabulary items.

Using equivalence tests, we found that there was a significant difference in learning for verbs (adults mean accuracy = 0.82 (SD = 0.22), $t(36.4) = 4.16$, $p < 0.001$), but not nouns (adults mean accuracy = 0.60 (SD = 0.16), $t(32.54) = 1.45$, $p = 0.157$) and marker words (adults mean accuracy = 0.46 (SD = 0.12), $t(23.73) = -1.23$, $p = 0.232$). There was also a significant difference in learning for syntax (adults mean accuracy = 0.85 (SD = 0.16), $t(33.02) = 4.07$, $p < 0.001$). Thus, adults were able to learn both vocabulary and syntax more readily than children, with evidence of learning both simultaneously. Children, however, showed evidence for learning syntax, but no evidence for learning vocabulary. It must be noted that this null effect for vocabulary is not the same as providing evidence that children did *not* learn vocabulary learning; however, there is a difference between adult and child learning in that adults could learn verbs to a level equivalent to that of syntax, whereas there was considerable disparity for children. What might result in this possible child–adult distinction in cross-situational learning?

One possible explanation might lie with the different memory system required for language learning to proceed. Neurobiological evidence showed that procedural memory, which relates more closely to the processing of syntactic regularities, tends to mature in early childhood (Pili-Moss, 2021). Declarative memory, which supports vocabulary and grammar learning, tends to reach maturity later in adolescence (Gómez & Edgin, 2016; Morgan-Short *et al.*, 2014; Ullman, 2004). This is consistent with the current study as the syntactic regularity (i.e., word order) was the first, and only, language property learned to a significant level by the children.

The results suggest that, when neither syntax nor vocabulary is known to the child, knowledge about syntax, in terms of word order constraints, quickly becomes available to the learner. This provides an indication of how syntactic information might help scaffold and support vocabulary acquisition. Syntactic bootstrapping is thus available to children to acquire vocabulary items whose meaning can be in part dependent upon their syntactic role (Babineau *et al.*, 2024; Gleitman, 1990; Höhle & Weissenborn, 2001; Monaghan *et al.*, 2023).

4.1. Limitations and further directions

The current study is the first to investigate children's simultaneous acquisition of vocabulary and grammar through CSL. However, the sample size in this study is relatively small ($n = 20$), and the age range is relatively constrained (8–9 years), which might limit the generalizability of our findings as it may not fully capture the variability in CSL performance across a broader population of children. Extending this study to a larger age range may provide us with fuller insight into how learning vocabulary and learning syntax interrelate in children's language development.

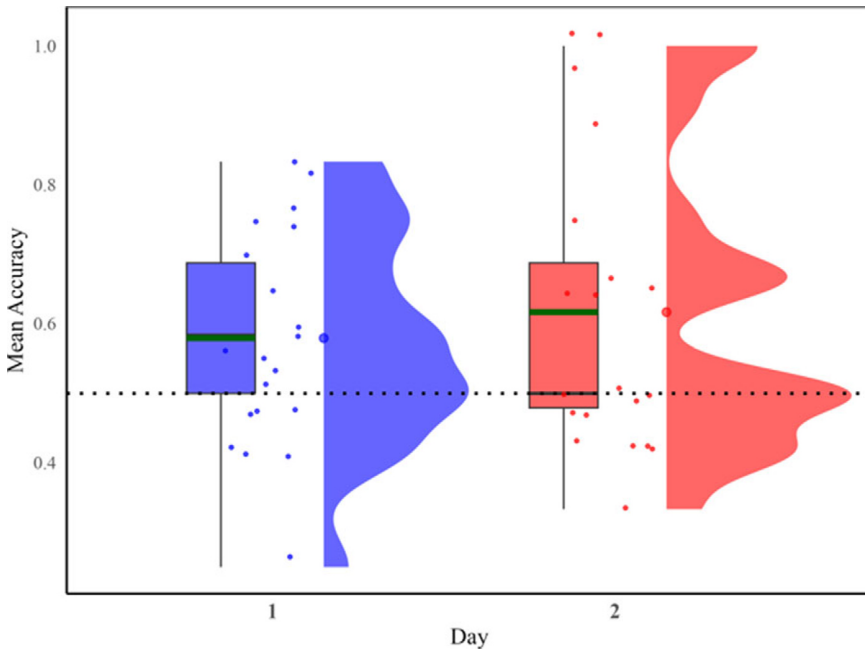


Figure 4. Mean accuracy performance by syntax test trials in CSL task for Days 1 and 2. The box indicates the median (horizontal line) and interquartile range, with dots indicating individuals' accuracy.

Our study also did not fully encapsulate the individual differences that were driving children's performance. Note that in [Figures 2 and 4](#), there is a large range in accuracy for individual children, indicating that whereas some children were able to acquire both vocabulary and syntax, other children failed to gain a foothold in learning the language at all. Future research that includes cognitive skill measurements (in particular, procedural memory and declarative memory) will be a useful extension to determine how memory systems relate to different aspects of language learning when learning a language immersively. Existing adult data using a similar paradigm ([Walker et al., 2020](#)) indicated that learning vocabulary and syntax simultaneously may not relate neatly to abilities in declarative and procedural memory systems. [Walker et al. \(2020\)](#) found that learning all aspects of a language from cross-situational statistics (both vocabulary and syntax) related to procedural memory ability early in training, and, as learning advanced, declarative memory became more important as a predictor of syntax and verb learning accuracy. [Ferman and Karni \(2010\)](#) found that language learning associated with procedural skills also improved slightly with age from 8 to 12 years, to adulthood. Hence, there may be crucial stages of learning as the interactions between syntax and vocabulary emerge through exposure, each relating to different memory systems at different stages of learning, and children's mastery of vocabulary – and syntax – may well require the later-developing declarative memory system.

As revealed in [Isbilen & Christiansen \(2022\)](#), the effect size of SL in the child population is significantly affected by test types, with processing-based tests yielding a larger effect size than reflection-based tests, and production and recall tests yielding larger effect sizes than forced-choice tests. Furthermore, testing implicit knowledge, such as

syntax, is more appropriately tested through implicit, online measures, whereas testing explicit knowledge, such as vocabulary, can be measured effectively using explicit tests (Isbilen *et al.*, 2022). The current study used such an explicit measure – requiring a forced choice. A blend of online and offline measures, then, would be a valuable extension of the current study in order to explore in greater detail the quality and quantity of children's language learning.

A further advantage of online tasks, such as eye-tracking, would enable us to determine children's attention to different elements of the scenes during learning. One possibility for children's greater learning of syntax than vocabulary may have been due to children reducing attention to the visual scene and focusing more processing on the auditory stimuli. It is the case that the syntax test could be solved without requiring processing of the scene at all, as it tested sensitivity to the word order within sentences. However, our task was designed to keep children's focus on the screen by providing a visual reward of a coin for correct answers during training and requiring a response for each trial, which could only be accomplished by processing the relations among auditory sentences and visual scenes. Eye tracking would enable us to confirm how children use the visual information in conjunction with the auditory input. Note, however, that performance did not change from Day 1 to Day 2 for either the overall training trials or for the syntax tests, and so there were no evident quantitative changes in children's performance over the task. Furthermore, there was above-chance performance on the training trials, though at a level substantially below that of the syntax trials. Thus, word order information was available and processed more readily by children than the vocabulary information, and this highlights that potential information for syntactic bootstrapping is present to help support subsequent vocabulary acquisition. The benefit of the current paradigm is that it can offer opportunities for investigating all these issues, including effects of the type of task, as well as environmental exposure, cognitive effects, and language background effects on children's early language development.

Previous studies of cross-situational word learning have established that children can use these statistics to acquire new vocabulary (e.g., Vlach & DeBrock, 2017). We are not claiming that children are unable to do this, only that, from input that involves novel syntax and vocabulary, children seem to pick up on the syntax more readily than the vocabulary. The lower levels of learning of vocabulary observed in our study compared to previous cross-situational studies with children likely rest with the greater complexity entailed by multiple words relating to multicomponent scenes. For instance, in previous simpler cross-situational studies, all objects and all words tend to be present in each learning situation. Then, the conditional probability of an object appearing with a word is 1, and the conditional probability of another object appearing with a word is $1/(n - 1)$, where n = number of word/object pairings. So, for learning six words, the difference in frequency of cross-situational mappings is 1 versus 1/5 for intended versus unintended pairings. In our paradigm, however, as there are multiple words and two possible scenes occurring in each trial, the difference in frequency of mappings between intended and unintended pairings is smaller.

Table 4 shows the conditional probabilities for mappings, for nouns, and for verbs. Adapting our paradigm to reduce the scene and sentence complexity would likely result in more successful word learning. We contend that with this simpler sentence structure, children would still be more successful at acquiring the word order because of the earlier maturation of their procedural memory systems. This, however, is a matter for future investigation.

Table 4. Conditional probabilities of noun-object and verb-action pairings in our study, with probabilities also shown for a standard cross-situational learning study for 6 noun-object pairings

Conditional	Our study's probability	e.g., Vlach & DeBrock (2017) 2 words x 2 objects, 12 pairings probability
$p(\text{noun} \mid \text{target object})$	0.5	1
$p(\text{noun} \mid \text{foil object})$	0.3	1/12
$p(\text{target object} \mid \text{noun})$	1	1
$p(\text{foil object} \mid \text{noun})$	0.3	1/11
$P(\text{verb} \mid \text{target action})$	0.5	
$P(\text{verb} \mid \text{foil action})$	0.17	
$P(\text{target action} \mid \text{verb})$	1	
$P(\text{foil action} \mid \text{verb})$	0.33	

In conclusion, we showed that children aged 8 to 9 were able to learn syntax (i.e., word order) through tracking the co-occurrence of target sentences and referential scenes with no need for prior vocabulary knowledge. However, there was no evidence that cross-situational statistics alone were sufficient to support robust simultaneous early-stage acquisition of both vocabulary and syntax in children at this age, as it has been shown to do in adults.

Supplementary material. The supplementary material for this article can be found at <http://doi.org/10.1017/S0305000925100135>.

Data availability statement. Our materials, anonymized data, and data analysis scripts are available on our project site² (https://osf.io/y3wp4/?view_only=af66c1da694a4cbdaa39742d50872c3a) on the Open Science Framework (OSF) platform.

Funding statement. We gratefully acknowledge the financial support provided by Lancaster University's Camões Institute Cátedra for Multilingualism and Diversity and by Research Catalyst Fund of the Faculty of Arts and Social Sciences, Lancaster University. Padraic Monaghan was supported by the International Centre for Language and Communicative Development (LuCiD) at Lancaster University and the University of Liverpool, funded by the Economic and Social Research Council (UK) (ES/S007113/1). Padraic Monaghan and Patrick Rebuschat jointly supervised this project and thus share senior authorship.

Competing interests. The authors declare none.

Disclosure of use of AI tools. No AI tools were used in the preparation of this manuscript in any aspect, including generating images, generating text, or analysing and extracting insights.

²Our initial pre-registration involved testing an additional group of children who received feedback during the same experimental paradigm. However, a software fault meant that these children could respond before the end of the sentence had played, and in most cases they did so. We therefore did not proceed to analyse the data for this group, and instead focused on the hypotheses of our pre-registration that related to cross-situational learning of vocabulary and grammar.

References

- Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., & Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*, **164**, 116–143.
- Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, **19**(57), 347–358.
- Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, **116**(1), 143–148.
- Babineau, M., Barbir, M., de Carvalho, A., Havron, N., Dautriche, I., & Christophe, A. (2024). Syntactic bootstrapping as a mechanism for language learning. *Nature Reviews Psychology*, **3**(7), 463–474.
- Bauer, P. J. (2008). Toward a neuro-developmental account of the development of declarative memory. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, **50**(1), 19–31.
- Benitez, V. L., Zettersten, M., & Wojcik, E. (2020). The temporal structure of naming events differentially affects children's and adults' cross-situational word learning. *Journal of Experimental Child Psychology*, **200**, 104961.
- Childers, J. B., Cutilletta, B., Capps, K., Tovar-Perez, P., & Smith, G. (2023). Can children learn verbs from events separated in time? Examining how variability and memory contribute to verb learning. *Journal of Experimental Child Psychology*, **227**, 105583.
- Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, **34**(6), 1017–1063.
- Ferman, S., & Karni, A. (2010). No childhood advantage in the acquisition of skill in using an artificial language rule. *PLoS One*, **5**(10), e13648.
- Finn, A. S., Kalra, P. B., Goetz, C., Leonard, J. A., Sheridan, M. A., & Gabrieli, J. D. (2016). Developmental dissociation between the maturation of procedural memory and declarative memory. *Journal of Experimental Child Psychology*, **142**, 212–220.
- Fisher, C., Gertner, Y., Scott, R. M., & Yuan, S. (2010). Syntactic bootstrapping. *Wiley Interdisciplinary Reviews: Cognitive Science*, **1**(2), 143–149.
- Fitneva, S. A., & Christiansen, M. H. (2017). Developmental changes in cross-situational word learning: The inverse effect of initial accuracy. *Cognitive Science*, **41**, 141–161.
- Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences*, **99**(1), 529–534.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, **1**(1), 3–55.
- Gómez, R. L., & Edgin, J. O. (2016). The extended trajectory of hippocampal development: Implications for early memory development and disorder. *Developmental Cognitive Neuroscience*, **18**, 57–69.
- Höhle, B., & Weissenborn, J. (2001). *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition* (Vol. 1). John Benjamins.
- Hu, C. F. (2017). Resolving referential ambiguity across ambiguous situations in young foreign language learners. *Applied PsychoLinguistics*, **38**(3), 633–656.
- Hulstijn, J. H. (2015). Explaining phenomena of first and second language acquisition with the constructs of implicit and explicit learning: The virtues and pitfalls of a two-system view. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 25–46). Benjamins.
- Isbilen, E. S., & Christiansen, M. H. (2022). Statistical learning of language: A meta-analysis into 25 years of research. *Cognitive Science*, **46**(9), e13198.
- Isbilen, E. S., Frost, R. L., Monaghan, P., & Christiansen, M. H. (2022). Statistically based chunking of nonadjacent dependencies. *Journal of Experimental Psychology: General*, **151**(11), 2623–2640.
- Lum, J., Kidd, E., Davis, S., & Conti-Ramsden, G. (2010). Longitudinal study of declarative and procedural memory in primary school-aged children. *Australian Journal of Psychology*, **62**(3), 139–148.
- Meulemans, T., Van der Linden, M., & Perruchet, P. (1998). Implicit sequence learning in children. *Journal of Experimental Child Psychology*, **69**(3), 199–221.
- Ministry of Housing, Communities & Local Government. (2019). The English indices of deprivation 2019 (Statistical release). <https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019>

- Monaghan, P., Donnelly, S., Alcock, K., Bidgood, A., Cain, K., Durrant, S., ... Rowland, C. F. (2023). Learning to generalise but not segment an artificial language at 17 months predicts children's language skills 3 years later. *Cognitive Psychology*, **147**, 101607.
- Monaghan, P., Mattock, K., Davies, R. A., & Smith, A. C. (2015). Gavagai is as Gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science*, **39**(5), 1099–1112.
- Monaghan, P., Ruiz, S., & Rebuschat, P. (2021). The role of feedback and instruction on the cross-situational learning of vocabulary and morphosyntax: Mixed effects models reveal local and global effects on acquisition. *Second Language Research*, **37**(2), 261–289.
- Monaghan, P., Schoetensack, C., & Rebuschat, P. (2019). A single paradigm for implicit and statistical learning. *Topics in Cognitive Science*, **11**(3), 536–554.
- Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, **17**(1), 56–72.
- Pili-Moss, D. (2021). Cognitive predictors of child second language comprehension and syntactic learning. *Language Learning*, **71**(3), 907–945.
- Quine, W. V. O. (1960). *Word and object*. MIT Press.
- Rebuschat, P., Monaghan, P., & Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition*, **206**, 104475.
- Ruiz, S., Tagarelli, K. M., & Rebuschat, P. (2018). Simultaneous acquisition of words and syntax: Effects of exposure condition and declarative memory. *Frontiers in Psychology*, **9**, 1168.
- Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, **122**(2), 163–180.
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, **106**(3), 1558–1568.
- Spit, S., Andringa, S., Rispens, J., & Aboh, E. O. (2022). Kindergarteners use cross-situational statistics to infer the meaning of grammatical elements. *Journal of Psycholinguistic Research*, **51**(6), 1311–1333.
- Suanda, S. H., Mugwanya, N., & Namy, L. L. (2014). Cross-situational statistical word learning in young children. *Journal of Experimental Child Psychology*, **126**, 395–411.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, **92**(1–2), 231–270.
- Venker, C. E. (2019). Cross-situational and ostensive word learning in children with and without autism spectrum disorder. *Cognition*, **183**, 181–191.
- Vlach, H. A., & DeBrock, C. A. (2017). Remember Dax? Relations between children's cross-situational word learning, memory, and language abilities. *Journal of Memory and Language*, **93**, 217–230.
- Walker, N., Monaghan, P., Schoetensack, C., & Rebuschat, P. (2020). Distinctions in the acquisition of vocabulary and grammar: An individual differences approach. *Language Learning*, **70**(S2), 221–254.
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, **70**(13–15), 2149–2165.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, **18**(5), 414–420.