

## Learning vocabulary and grammar from cross-situational statistics

Patrick Rebuschat<sup>a,b,\*</sup>, Padraic Monaghan<sup>a,c,d</sup>, Christine Schoetensack<sup>a</sup>

<sup>a</sup> Lancaster University, UK

<sup>b</sup> University of Tübingen, Germany

<sup>c</sup> University of Amsterdam, the Netherlands

<sup>d</sup> Max Planck Institute for Psycholinguistics, the Netherlands

### ARTICLE INFO

#### Keywords:

Statistical learning  
Cross-situational learning  
Language acquisition  
Implicit learning

### ABSTRACT

Across multiple situations, child and adult learners are sensitive to co-occurrences between individual words and their referents in the environment, which provide a means by which the ambiguity of word-world mappings may be resolved (Monaghan & Mattock, 2012; Scott & Fisher, 2012; Smith & Yu, 2008; Yu & Smith, 2007). In three studies, we tested whether cross-situational learning is sufficiently powerful to support simultaneous learning the referents for words from multiple grammatical categories, a more realistic reflection of more complex natural language learning situations. In Experiment 1, adult learners heard sentences comprising nouns, verbs, adjectives, and grammatical markers indicating subject and object roles, and viewed a dynamic scene to which the sentence referred. In Experiments 2 and 3, we further increased the uncertainty of the referents by presenting two scenes alongside each sentence. In all studies, we found that cross-situational statistical learning was sufficiently powerful to facilitate acquisition of both vocabulary and grammar from complex sentence-to-scene correspondences, simulating the situations that more closely resemble the challenge facing the language learner.

In order to understand an utterance, the language learner has to develop an understanding of the meaning of words within the utterance by acquiring the vocabulary, and also determine the grammatical roles of those words from the syntactic structure of the sentence. There are alternative perspectives on how these two aspects of language are acquired. Learning of vocabulary and grammar may be driven by separable modular processes (Peña et al., 2002; Pinker, 1998), or through the operation of statistical learning mechanisms that can acquire individual vocabulary items as well as generalise over relations between elements in speech (Frost & Monaghan, 2016; Plunkett & Marchman, 1993; Reeder et al., 2013). Relatedly, there are distinctions in theory over whether vocabulary and grammar are acquired simultaneously or in sequence (Marchman & Bates, 1994; Pentimonti et al., 2015; Tomblin & Zhang, 2006), with the prevailing view that grammatical knowledge is bootstrapped from acquisition of a small vocabulary (see, for example, the contributions in Morgan & Demuth, 1996).

Cross-situational learning has been proposed as a key mechanism by which information from the environment can be aligned with information in the utterance to permit language learning. From a single situation, the meaning of a word is highly ambiguous with respect to many possible referents in the environment (Quine, 1960). But across multiple learning situations, the co-occurrence of a particular word with one of several referents in the environment can be tracked by the

learner to determine the intended referent (Siskind, 1996). In a seminal paper, Yu and Smith (2007) tested whether adult learners can use cross-situational statistics to acquire novel word-object mappings. On each trial, learners were presented with multiple objects on a computer screen (e.g., canister, facial sauna) while listening to sequences of pseudowords, with no information as to which label went with which referent. Critically, the presentation sequence of the pseudowords was unrelated to the positioning of the objects on the screen. To acquire the word-object mappings, learners thus had to keep track of potential referents and labels over multiple trials. Yu and Smith (2007) showed that learners can use cross-situational information to acquire novel nouns after a very brief exposure period (6 min). Impressively, participants were able to do so even when the within-trial ambiguity was further increased, with rapid learning still observed when each trial involved four referents and four potential labels. Smith and Yu (2008) further demonstrated that 12-month and 14-month-old infants are also able to learn novel word-object mappings via cross-situational learning.

Cross-situational statistics have been shown to be effective in supporting speech segmentation (Cunillera et al., 2010) and in learning nouns (e.g., Hamrick & Rebuschat, 2012; Smith & Yu, 2008; Suanda & Namy, 2012; Vlach & Johnson, 2013; Yu & Smith, 2007), verbs (e.g., Childers & Paik, 2009; Scott & Fisher, 2012), adjectives (Akhtar & Montague, 1999), and noun categories (Chen et al., 2018). Monaghan

\* Corresponding author at: Department of Linguistics and English Language, Lancaster University, Lancaster LA1 4YL, UK.

E-mail address: [p.rebuschat@lancaster.ac.uk](mailto:p.rebuschat@lancaster.ac.uk) (P. Rebuschat).

<https://doi.org/10.1016/j.cognition.2020.104475>

Received 10 October 2018; Received in revised form 11 September 2020; Accepted 22 September 2020

0010-0277/ © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

et al. (2015) further demonstrated that adult learners can acquire nouns and verbs simultaneously, with nouns being learned more quickly than verbs. Importantly, they showed that cross-situational learning is robust even under conditions of increased ambiguity. Whereas in previous research (e.g., Yu & Smith, 2007), the appropriate referents and labels always co-occurred in a given trial, this was not the case in Monaghan et al. (2015). Here, participants observed two dynamic scenes while listening to artificial language sentences that described only one of the scenes, thus increasing within-trial ambiguity in both the target reference and referent.

Previous studies have provided words from only one or two grammatical categories, and presented these alongside highly constrained possible sets of referents in the learner's environment. In contrast, in natural language learning, the listener hears utterances comprising multi-word sequences composed of words from multiple grammatical categories, and has to learn to map each of these words to objects and actions, their properties and relations between objects and events in the environment. The question remains whether cross-situational learning is powerful enough as a mechanism to resolve the degree of ambiguity in both the utterance and the scene that faces the language learner. The complexity of natural language raises a further difficulty for the learner. Understanding the syntactic dependencies between words and the grammatical roles of those words are necessary before the word's meaning can be determined. For instance, in the case of the distinction between *give* and *receive*, the child needs to understand the roles of agent and patient in the syntax before the meaning of the verb can be properly linked to the event, as the event associated with "John gives the book to Mary" and "Mary receives the book from John" will be identical (Childers et al., 2012; Gleitman, 1990; Gleitman et al., 2005). However, grammar learning in turn appears to require knowledge of the meaning of words that constitute the grammatical categories. For instance, determining that English has subject-verb-object (SVO) word order requires identifying that one category of words (nouns) tends to occur before another category of words (verbs) which has a different set of properties. Abstracting over word-order regularities between grammatical categories would not be possible prior to the learning of such grammatical categories.

This "chicken and egg" puzzle has led to solutions proposing that the semantic features that grammatical categories relate to are innately specified, and that language learning requires acquisition of the links between those words and the innate semantic features (Pinker, 1998). Yet, even with these innate semantic features, learning the links between language and the features still seems to require the simultaneous acquisition of the vocabulary and the grammar – to cluster words according to their semantic features requires knowing that those words possess the semantic features in the first place. To avoid this difficulty of simultaneous acquisition, artificial language studies that investigated grammar learning have often pre-trained participants on the language's (pseudoword) vocabulary prior to exposure to the target grammar (e.g., Amato & MacDonald, 2010; Friederici et al., 2002; Morgan-Short et al., 2014).

In this paper, we test whether cross-situational learning is sufficiently powerful to support mapping from multi-word utterances to complex scenes. We also test whether vocabulary and grammar can both be acquired, at the same time, from cross-situational statistics, or whether training on vocabulary is a prerequisite for successful acquisition of grammatical properties of the language. It is known that cross-situational statistics can provide information both about individual nouns and noun categories (Chen et al., 2018), where participants could simultaneously learn labels for individual objects and for the superordinate category to which the individual object belonged. However, it is not yet known whether mappings between sentences comprising words from multiple grammatical categories and scenes depicting multiple potential referents is also learnable.

In Experiment 1, we exposed adult learners to an artificial language consisting of pseudowords from four different lexical categories, which

were arranged in accordance with Japanese syntax. Sentences always featured nouns, verbs and grammatical particles that provided information about subject-object assignment. Adjectives were optional and so did not occur in all sentences. Each sentence was presented along with a complex scene that was described by the sentence. We investigated whether words from within each category could be acquired from cross-situational statistics, without providing feedback. We also tested whether participants had acquired knowledge about the grammar of the language from this exposure. In Experiment 2, we further increased the ambiguity of the information to be learned by presenting two scenes along with each sentence, to determine whether cross-situational statistics could resolve the increased complexity of mapping from a multi-word utterance to an unspecified set of referents in the environment. In Experiment 3, we adjusted the training and testing regime to determine whether testing had an influence on learning vocabulary and grammar from cross-situational statistics. While in Experiments 1 and 2 participants were tested repeatedly to determine acquisition sequences of words and syntax, in Experiment 3 participants were only tested after the exposure phase. The data have been made available online on a third-party archive (osf.io/sbxm4).

## 1. Experiment 1: Learning vocabulary and grammar from cross-situational statistics

We employed an artificial language containing nouns, verbs, adjectives, and grammatical markers. The syntax of the artificial language was based on Japanese, with a flexible word order, either SOV or OSV. In this experiment, we tested whether adult learners can simultaneously acquire the vocabulary and the grammar of this more complex language through passive observation of a scene and listening to a sentence describing that scene. Successful completion of the task required participants to learn the correspondence between nouns, verbs, and adjectives with their referents across scenes without feedback and without any explicit information about the structure of the language. Furthermore, it required participants to learn two marker words that reliably identified subject and object in each sentence, as well as acquire the word order itself. If participants are able to resolve these tasks, it suggests that they can track the individual statistical correspondences between particular words and their objects, events or featural property referents across scenes, despite their uncertainty of both the word-referent mappings and of the grammar of the language.

### 1.1. Method

#### 1.1.1. Participants

Twenty university students (Mean age = 22.6 years,  $SD = 4.1$ , 15 women) volunteered to participate. All participants were native speakers of English, and none had a background in Japanese or any other verb-final language. Participants were remunerated for their time in accordance with the standard hourly rate of the Department of Psychology at Lancaster University (7 GBP per hour). The study was approved by the ethics review panel of the Faculty of Arts and Social Sciences at Lancaster University and conducted in accordance with the provisions of the World Medical Association Declaration of Helsinki. Sample size was inferred from Monaghan and Mattock's (2012) and Monaghan et al.'s (2015) studies of cross-situational learning of words from single grammatical categories resulting in effect sizes of 0.7, 1.5 and 1.7. If participants are able to learn sentences to scenes cross-situational correspondences, then with 20 participants, power for finding effects in a similar range would be 0.85 to 1.

#### 1.1.2. Materials

Eight alien cartoon characters served as referents to nouns in the artificial language (see Supplementary Materials for images). The aliens appeared in either red or blue and were depicted performing one of four actions (hiding, jumping, lifting, pushing) in animated scenes generated



Fig. 1. Screenshot of three time points for the visual scene in a training trial in Experiment 1. Participants were presented with coloured aliens performing an action and heard a sentence (e.g., “Haagle chelad tha goorshell sumbark noo fisslin”, meaning: red alien<sub>5</sub> jumps over blue alien<sub>7</sub>.) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

by E-Prime 2.0 (Schneider et al., 2002). Fig. 1 shows a sample scene.

The artificial language contained 16 pseudowords, taken from Monaghan and Mattock (2012) (see Supplementary Materials for list of stimuli). Fourteen bisyllabic pseudowords were content words: Eight nouns (one per alien), four verbs (one per action), and two adjectives (one per colour). Two monosyllabic pseudowords served as grammatical markers that reliably indicated if the preceding noun referred to the subject or the object of the sentence. We opted for this distribution so that the artificial language mirrored natural language properties more closely (see also Monaghan & Mattock, 2012). Word-referent mappings were randomly generated for each participant to control for preferences in associating certain sounds to objects, actions, or colours (see Monaghan & Fletcher, 2019, for discussion). The pseudowords were read and recorded individually by a female native speaker of English in a monotone. The artificial language sentences were assembled by E-Prime, with a 250 ms pause between each word.

The grammar of the artificial language was based on Japanese. Sentences could either be SOV or OSV, i.e. the verb phrase (VP) had to be placed in final position but the order of subject and object noun phrases (NP) was free. NPs contained an optional Adjective (A) pre-nominally, a noun (N), and a post-nominal grammatical marker that indicated if the preceding noun was the subject (SUBJECT) or the object (OBJECT) of the action. Adjectives occurred in half the NPs. Sentence length ranged between five and seven words. We generated 192 sentences which were divided into four training blocks each of 48 sentences. Within each block lexical frequencies, subject or object assignment, and word order were balanced. Table 1 summarizes the grammatical sentence patterns that occurred, with equal frequency, in the experiment. A further 96 test sentences were also generated and were controlled in a similar way to the training sentences.

**Table 1**  
Grammatical sentence patterns that occurred in the experiment.

Word order	Sequence		
	First	Second	Third
SOV	NP <sub>subj</sub> (A N <sub>subj</sub> )	NP <sub>obj</sub> (A N <sub>obj</sub> )	VP (V)
	NP <sub>subj</sub> (A N <sub>subj</sub> )	NP <sub>obj</sub> (N <sub>obj</sub> )	VP (V)
	NP <sub>subj</sub> (N <sub>subj</sub> )	NP <sub>obj</sub> (A N <sub>obj</sub> )	VP (V)
	NP <sub>subj</sub> (N <sub>subj</sub> )	NP <sub>obj</sub> (N <sub>obj</sub> )	VP (V)
OSV	NP <sub>obj</sub> (A N <sub>obj</sub> )	NP <sub>subj</sub> (A N <sub>subj</sub> )	VP (V)
	NP <sub>obj</sub> (A N <sub>obj</sub> )	NP <sub>subj</sub> (N <sub>subj</sub> )	VP (V)
	NP <sub>obj</sub> (N <sub>obj</sub> )	NP <sub>subj</sub> (A N <sub>subj</sub> )	VP (V)
	NP <sub>obj</sub> (N <sub>obj</sub> )	NP <sub>subj</sub> (N <sub>subj</sub> )	VP (V)

### 1.1.3. Procedure

Participants were trained and tested on the artificial language over four blocks of training, interspersed with four testing blocks. This allowed us to determine the acquisition order of the different linguistic elements of the system (nouns, verbs, adjectives, marker words, word order). The entire procedure took approximately 45 min.

**Training blocks.** After providing informed consent, participants were instructed that they would learn a new language, spoken by the “friendly inhabitants of a distant planet”. They were not asked to pay attention to any particular aspect of the alien language. Their task was to passively observe a dynamic scene on the screen and to listen to an artificial language sentence describing the scene.

In each trial, participants first viewed an animated scene in which two alien characters performed an action. They then heard the sentence describing the scene, e.g., for the scene shown in Fig. 1:

haagle chelad tha goorshell sumbark noo fisslin  
gloss: blue alien<sub>7</sub> OBJECT red alien<sub>5</sub> SUBJECT jumps  
“red alien<sub>5</sub> jumps over blue alien<sub>7</sub>”

This was then followed by another presentation of the action. The next trial then immediately began. No response was required. There were four training blocks ( $n = 48$  unique trials per block).

Prior to the first training block, participants observed two practice trials with aliens performing actions, accompanied by random pseudoword sequences. The aliens, their colours (green), the actions and the pseudowords used in the practice did not further occur in the experiment.

**Test blocks.** Each training block was followed by a test block. In each of these, vocabulary learning was assessed first. This was done by means of a two-alternative forced-choice task, in which participants were presented with two animated scenes and played a test sentence. Their task was to decide, as quickly and accurately as possible, which scene the sentence referred to. Words from each lexical category were assessed by varying the target and distractor scenes by one piece of information, such that knowledge of the vocabulary relating to the individual piece of information was required to determine which scene was described by the utterance. Thus, to test noun learning, participants saw two scenes that only differed with regards to one alien character. In the verb test trials, only the actions were different between the scenes. In the adjective test trials, the colours of the aliens were switched. Finally, in the grammatical marker test trials, the subject/object assignment was reversed, though note that understanding the grammatical role markers could be considered part of the grammar rather than the vocabulary.

There were with 24 trials in the vocabulary test. Of these 24 trials, sixteen were used to determine if participants had acquired the words from the different lexical categories (four trials per grammatical category), with a further eight trials used as fillers, in which the distractor scene was randomly assigned and so could vary by several aspects. Trials occurred in randomised order, and no feedback was provided. An example lexical test trial is shown in Fig. 2.

After completing the lexical test trials, grammar learning was assessed next in each of the test blocks. This was done by means of a grammaticality judgment task. Participants were told that they would see a scene and hear a sentence spoken by another alien from a very different planet who was also learning the new language. Their task was to decide, as quickly and accurately as possible, whether the new alien was speaking correctly. If the sentence sounded “good”, participants had to press a green button on a computer keyboard. If it sounded “funny”, they had to press a red button. There were 16 syntactic test trials in each block. Half the trials followed the grammar of the artificial language, with SOV and OSV sentence patterns carefully counter-balanced. The other half involved sentences with syntactic violations (\*SVO, \*OVS, \*VSO, \*VOS). Trials occurred in randomised order, and no feedback was provided on response accuracy.

Of the 160 test items (96 lexical tests, 64 syntactic tests) used in the four test blocks, seven test items had previously occurred in the study materials. One of these was a syntax test item, three were noun test items (distributed across test blocks), one was an adjective test item, and two were marker test items (again in different test blocks). No repetitions occurred in adjacent blocks. Due to the low level of repetition in the study, the lack of proximity between the rare repetitions of sentences, and the substantial number of unique sentences used across training and testing, we believe that memorizing sentences is a highly unlikely strategy to be used by participants.

After the final test block, participants completed a debriefing questionnaire that probed any strategies that participants' may have used for the task, and also completed a language background questionnaire.

## 1.2. Results

Accuracy for acquisition of nouns, verbs, adjectives, marker words, and word order for each of the four test blocks are shown in Fig. 3.

Whether performance was greater than chance was determined by uncorrected one-sample *t*-tests for each test. We report these results with uncorrected *p*-values, thus they should be considered in conjunction with the linear contrast effects of test block to illustrate for which language features learning improves. The results are shown in Table 2. By test 4, performance was significantly above chance for all lexical categories, except for marker words, which were significantly above chance only at test 3, though this was not a significant effect after controlling for multiple comparisons. For word order, performance was significantly above chance across all the tests.

In order to determine the rate of learning across the four test blocks, we conducted one-way ANOVAs with test block as within subjects factor on each lexical category and word order separately. For nouns, there was a significant effect of test block,  $F(3, 57) = 9.24, p < .001, \eta_p^2 = 0.33$ , indicating that learning increased with additional training. Polynomial contrasts confirmed that the best fit to the effect of block was linear,  $F = 20.24, p < .001, \eta_p^2 = 0.52$ . Quadratic and cubic fits were not significant,  $p > .09$ . For verbs, there was no significant effect of test block,  $F < 1$ , with performance already high by the first testing block. For verbs there were no significant linear, quadratic, or cubic contrasts,  $p > .31$ . A significant learning effect was found for adjectives,  $F(3, 57) = 3.43, p = .023, \eta_p^2 = 0.15$ , which was significant in the linear contrast,  $F = 6.77, p = .018, \eta_p^2 = 0.26$ , but this was not significant for higher-order contrasts,  $p > .35$ . There was also a significant effect for marker words,  $F(3, 57) = 3.53, p = .020, \eta_p^2 = 0.16$ , which was a significant linear effect,  $F = 5.38, p = .032, \eta_p^2 = 0.22$ ,

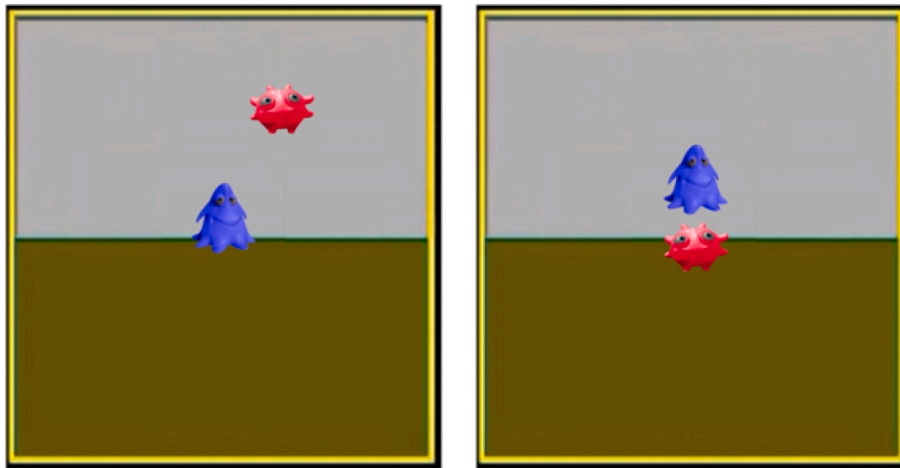
demonstrating that learning the meaning of these vocabulary items improved with cross-situational training exposure. The quadratic contrast was not significant,  $p = .617$ , but the cubic contrast was significant,  $F = 5.19, p = .034, \eta_p^2 = 0.22$ , but we favour the linear fit because it is a simpler model that accounts for slightly more variance in responses. For word order, more training did not result in significantly improved performance,  $F < 1$ , as performance was already accurate in the first block of training. No contrasts were significant,  $p > .25$ .

## 1.3. Discussion

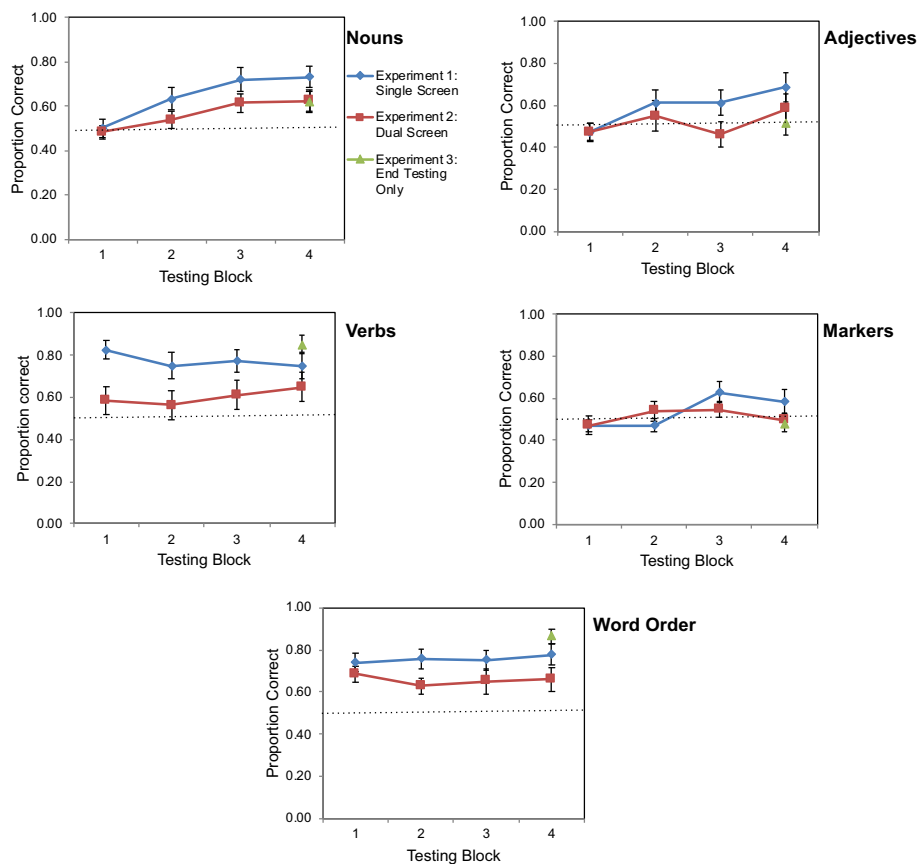
The results demonstrated that cross-situational statistics are sufficiently powerful to enable adult learners to acquire multiple lexical categories (nouns, verbs, adjectives) as well as word order patterns, without prior knowledge of either and without feedback. For marker words, though the comparisons against chance indicated that learning was not significant at any individual test point, the learning effect revealed in the ANOVA linear contrast showed that performance was improving with training. In terms of acquisition sequence, our results suggest that participants first acquired verbs and word order (above-chance performance on all four tests), then nouns (above chance from test 2), followed by adjectives. The size of the learning effects also follow this pattern: with large effects of accuracy above chance for verbs and word order, with medium sized effects of learning becoming large effect sizes by the end of training, and medium effects of learning for adjectives and markers by the end of the study. The key result is that all these language features - both vocabulary and word order are learnable as a consequence of information present in cross-situational statistics. However, the precise order of learning of each language feature may be partly a consequence of the particular language structure in the design, and so conclusions about prioritization of particular lexical categories cannot be deduced from the current study. We return to these points in the General Discussion. Given that participants were not instructed to consciously learn the pseudowords or discover the underlying syntactic structure, we assume that most learning was incidental, as a by-product of exposure.<sup>1</sup>

In Experiment 1, though the precise mappings between words in the sentence and aspects of the scene were unspecified, the set of referents to which the utterance related were still apparent to the learner. However, precisely which scene, or part of a scene, is being described by an utterance is rarely prespecified in natural language interactions (Cartmill et al., 2013; Clerkin et al., 2017). In the next study, we introduced additional complexity by presenting the utterance with two scenes, only one of which is referred to by the sentence. In Experiment 2, we tested whether adult participants can still learn vocabulary and grammar under these conditions of increased uncertainty of the cross-situational statistical correspondences.

<sup>1</sup> Participants were informed that they were going to learn a novel alien language but not instructed to pay attention to any particular aspects of the language or to engage in explicit hypothesis testing. Since it was very obvious from the task that participants were going to learn something new, it made little sense not to mention the alien language at the beginning. In fact, this is not very different from natural language learning in the real world, where adult learners are clearly aware that they are acquiring a new language, either inside or outside of the classroom. Of course, once participants knew they were learning a new language, they could have tried to deploy conscious strategies to boost learning. However, in this case, cross-situational learning should still take place (the process is automatic). Also, given the complexity of the artificial language, this type of strategy would probably not lead to a significantly better performance as there was too much information to consciously recall over multiple learning trials.



**Fig. 2.** Example of a lexical test trial, measuring knowledge of the verb (aliens, colours, and subject/object roles are the same in both scenes). Here, participants might hear, for example, “Dingep noo makkot tha fisslin.”, meaning: alien<sub>s</sub> jumps over alien<sub>s</sub>, which would refer to the left scene. The right scene displays a lifting motion.



**Fig. 3.** Performance on test blocks for lexical categories (nouns, verbs, adjectives, grammatical markers) and syntax (word order). Error bars show standard error of the mean. Dotted line at 0.5 shows chance performance.

## 2. Experiment 2: Learning vocabulary and grammar with increased uncertainty

### 2.1. Method

#### 2.1.1. Participants

Twenty university students (Mean age = 25.2,  $SD = 4.0$ , 14 women) participated and received payment for their time. All participants were native speakers of English. None of the participants had a

background in Japanese or any other verb-final language, and none took part in Experiment 1.

#### 2.1.2. Materials

The materials were identical to Experiment 1.

#### 2.1.3. Procedure

The procedure was identical to Experiment 1, except that in each trial participants observed two animated scenes (rather than one) while



**Table 2**  
One-sample *t*-tests and Cohen's *d* for performance against chance level (0.5) for each vocabulary type and word order, at each test block in Experiment 1.

	Test 1 <i>t</i> (19), <i>d</i>	Test 2 <i>t</i> (19), <i>d</i>	Test 3 <i>t</i> (19), <i>d</i>	Test 4 <i>t</i> (19), <i>d</i>
Nouns	0.04, 0.01	2.62*, 0.59	4.20***, 0.94	4.77***, 1.07
Verbs	7.26***, 1.62	3.98***, 0.89	5.08***, 1.14	4.16***, 0.93
Adjectives	-0.57, -0.13	1.83, 0.41	1.83, 0.41	2.78**, 0.62
Marker Words	-1.02, -0.23	-0.88, -0.20	2.59*, 0.58	1.48, 0.33
Word Order	5.43***, 1.22	5.25***, 1.17	5.26***, 1.18	5.39***, 1.21

\*  $p < .05$ .

\*\*  $p < .0125$ .

\*\*\*  $p < .001$ .

being presented with an artificial language sentence that matched one of the scenes, see Fig. 4. Locations of target and distractor scenes were counterbalanced and randomised. The distractor scene was randomly selected, with actions and aliens in the distractor scene always different from those in the target scene. This was done to ensure that the randomly generated distractor scene was not identical to the target scene. The next trial began after three seconds.

As with Experiment 1, debrief and language background questionnaires were given to participants.

## 2.2. Results and discussion

For each test of lexical category and word order, we conducted one-sample *t*-tests against chance level. The results are shown in Table 3.

One-way ANOVAs with test block as within subjects factor demonstrated: For nouns, there was a marginally significant effect of test block,  $F(3, 57) = 2.72, p = .053, \eta_p^2 = 0.16$ , which was significant in the linear contrast,  $F = 6.37, p = .021, \eta_p^2 = 0.25$ , but not significant for higher-order contrasts,  $p > .57$ . For verbs, there was no significant effect of training,  $F < 1$ , and contrasts were also not significant,  $p > .28$ . For adjectives, there was again no significant effect,  $F(3, 57) = 1.21, p = .316, \eta_p^2 = 0.06$ , not significant also in the polynomial contrasts,  $p > .13$ . There was also no significant effect for marker words,  $F < 1$ , polynomial contrasts all  $p > .11$ , or for word order,  $F < 1$ , polynomial contrasts  $p > .25$ . Thus, for nouns, verbs, and word order there was evidence of learning by the end of the training, but performance on adjectives and marker words did not demonstrate a clear improvement in performance with training time.

As shown in Table 3 and Fig. 3, overall accuracy for the two-screen version of the task seemed to reduce compared to the one-screen version of the task in Experiment 1, with smaller effects of learning evident

**Table 3**  
One-sample *t*-tests and Cohen's *d* values for performance against chance level (0.5) for each vocabulary type and word order, at each test block in Experiment 2.

	Test 1 <i>t</i> (19), <i>d</i>	Test 2 <i>t</i> (19), <i>d</i>	Test 3 <i>t</i> (19), <i>d</i>	Test 4 <i>t</i> (18/19) <sup>a</sup> , <i>d</i>
Nouns	-0.53, -0.12	0.98, 0.22	2.77**, 0.62	2.59*, 0.58
Verbs	1.32, 0.30	0.93, 0.21	1.69, 0.38	2.18*, 0.49
Adjectives	-0.62, -0.14	0.68, 0.15	-0.65, -0.14	1.23, 0.27
Marker Words	-0.68, -0.15	0.81, 0.18	1.25, 0.28	-0.13, -.03
Word Order	4.92***, 1.10	3.24**, 0.73	2.57*, 0.58	2.79**, 0.64

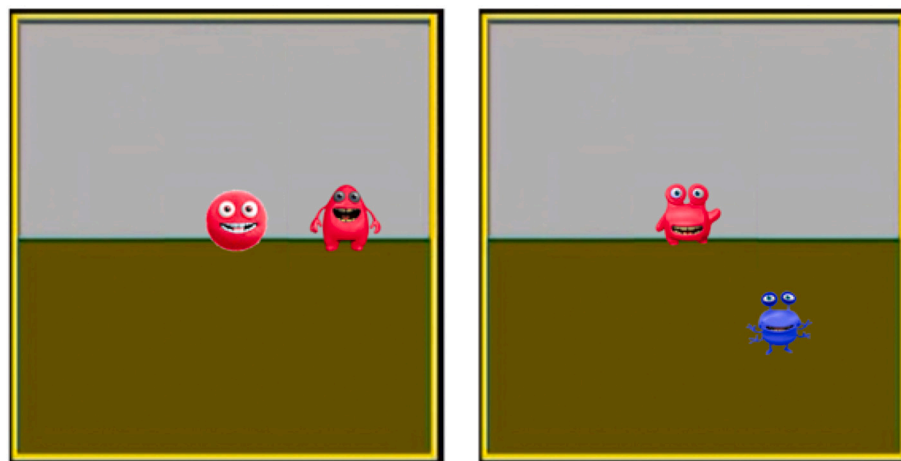
\*  $p < .05$ .

\*\*  $p < .0125$ .

\*\*\*  $p < .001$ .

<sup>a</sup> Note, one person's data was lost for Word order syntax test 4 due to computer error.

for all language features. An ANOVA comparing performance on Experiments 1 and 2, with experiment as between subjects factor, language feature (noun, verb, adjective, marker word, word order), and test (1 to 4) as within subjects factors resulted in a significant effect of experiment,  $F(1, 37) = 4.54, p = .040, \eta_p^2 = 0.11 (\eta^2 = 0.12)$ , indicating that the increased uncertainty of Experiment 2, though still demonstrating effective learning of nouns, verbs, and word order, resulted in lower overall performance ( $M = 0.52, SD = 0.11$ , compared to Experiment 1:  $M = 0.66, SD = 0.14$ ). There was a significant effect of language feature,  $F(4, 148) = 16.59, p < .001, \eta_p^2 = 0.31 (\eta^2 = 0.45)$ , with verbs and word order learned similarly ( $p = 1.0$ , all *p*-values Bonferroni corrected) and better than nouns ( $p \leq 0.012$ ), which were in turn learned better than marker words ( $p < .001$ ) but were similar to adjectives ( $p = .90$ ). Adjectives and marker words were not learned significantly differently ( $p = 1.0$ ). There was a significant effect of test time,  $F(3, 111) = 6.52, p < .001, \eta_p^2 = 0.15 (\eta^2 = 0.18)$ , which was also significant in the linear contrast,  $F(1, 37) = 13.03, p < .001, \eta_p^2 = 0.26 (\eta^2 = 0.35)$ . The interaction between experiment and language feature was not significant,  $F(4, 148) = 1.28, p = .28, \eta_p^2 = 0.03 (\eta^2 = 0.03)$ . The interaction between language feature and test time was significant,  $F(12, 444) = 2.35, p = .004, \eta_p^2 = 0.06 (\eta^2 = 0.07)$ . This interaction was due to the significant effect of time found for noun learning,  $F(3, 114) = 10.39, p < .001, \eta_p^2 = 0.22 (\eta^2 = 0.27)$ , adjective learning,  $F(3, 114) = 3.61, p = .016, \eta_p^2 = 0.09 (\eta^2 = 0.09)$ , and marker word learning,  $F(3, 114) = 3.00, p = .033, \eta_p^2 = 0.07 (\eta^2 = 0.08)$ , but no significant effect of time for verb learning or syntax learning, both  $F < 1$ , both of which were already learned effectively from the first block of testing. The three-way interaction was not



**Fig. 4.** Example of a training trial for Experiment 2. One of the scenes matched the sentence, and the other was a foil which could vary over the aliens, their colours, the action, and subject/object roles.

significant,  $F(12, 444) = 1.15, p = .318, \eta_p^2 = 0.03 (\eta^2 = 0.03)$ . Thus, across the two experiments, nouns, adjectives and marker words improved in accuracy with training but verbs, and word order tended to improve at a lower rate given their higher initial learning levels. These effects were regardless of whether there was uncertainty or not about the scene to which the sentence related.

Whereas Experiments 1 and 2 show that adult learners can simultaneously acquire multiple lexical categories and the syntactic structure of a complex artificial language via cross-situational statistics, it is not clear whether the interleaving of training and testing blocks may have influenced performance. As mentioned above, in Experiments 1 and 2, participants were tested repeatedly (after each of the four training blocks). This enabled us to determine acquisition sequences, but the existence of these tests might have impacted on performance. In the lexical test trials, the two scenes differ by one aspect – either one of the aliens, one of the actions, the colour of one alien, or the subject and object roles of the aliens. These scenes thus create minimal pairs, i.e. items that differ in one regard only. This focus on particular aspects of the scene may have alerted participants to the language structure, and as a consequence promoted their learning. In natural environments, such minimal pairs are unlikely to occur frequently, and so sensitivity to cross-situational statistics, if it is to support natural language learning, may have to function without these minimal distinctions. In Experiment 3, we test whether both vocabulary and grammar could be learned through cross-situational statistics without exposing participants to these minimal pairs, or any other tests, before the end of training.

### 3. Experiment 3: Testing cross-situational statistics without interleaved testing

In this experiment, we repeated the design of Experiment 2, with the exception that the acquisition of vocabulary and grammar was only tested at the end of training. That is, in contrast to Experiments 1 and 2, participants did not complete any test blocks prior to completion of exposure and thus did not encounter any minimal pairs that might have influenced learning (see Fig. 2). If learning from cross-situational statistics is dependent on these minimal pairs in the scenes, then learning of vocabulary or grammar may not be observed for this study. However, if cross-situational statistics are sufficiently powerful for learning without focusing the learner on one aspect of the language or the scene, then learning ought to be observed in this study, as in Experiments 1 and 2.

#### 3.1. Method

##### 3.1.1. Participants

Twenty university students participated (Mean age = 21.7 years,  $SD = 5.0$ , 12 women), and were paid for their participation. All were native speakers of English. None had a background in Japanese or any other verb-final language, and none participated in Experiments 1 or 2.

##### 3.1.2. Materials

The materials were identical to Experiments 1 and 2.

##### 3.1.3. Procedure

The procedure was identical to Experiment 2, except that there was just one testing block for vocabulary and grammar, which occurred at the end of the training. The testing block was the equivalent in number of trials to a single test block in Experiments 1 and 2.

#### 3.2. Results and discussion

We conducted one-sample *t*-tests and Cohen's *d* for performance against chance level (0.5) for each vocabulary type and word order in Experiment 3. The results were qualitatively similar to those from

Experiment 2. By the end of training, there was evidence of learning for nouns,  $t(19) = 2.580, p < .0125, d = 0.58$ , verbs,  $t(19) = 8.966, p < .001, d = 2.01$ , and word order,  $t(19) = 11.485, p < .001, d = 2.57$ , but no clear learning effect for adjectives,  $t(19) = -0.037, p > .05, d = -0.01$ , and marker words,  $t(19) = -0.427, p > .05, d = -0.10$ . An ANOVA comparing performance at the final testing block for Experiments 2 and 3, with experiment as between subjects factor, and language feature (noun, verb, adjective, marker word, word order) as within subjects factor resulted in no significant effect of experiment,  $F(1, 37) = 1.98, p = .168, \eta_p^2 = 0.05 (\eta^2 = 0.05)$ . There was a significant effect of language feature,  $F(4, 148) = 16.69, p < .001, \eta_p^2 = 0.311 (\eta^2 = 0.45)$ , due to verbs and word order learned to a similar degree ( $p = .805$ , *p*-values Bonferroni-corrected) and better than nouns ( $p = .001$ ), which were in turn learned significantly better than marker words ( $p = .003$ ) but not significantly better than adjectives ( $p = .157$ ). Adjectives and marker words were not learned significantly differently from one another ( $p = .309$ ). The interaction between experiment and language feature was significant,  $F(4, 148) = 4.00, p = .004, \eta_p^2 = 0.10 (\eta^2 = 0.11)$ . This interaction was due to testing at the end of training compared to testing interspersed throughout the study enhancing the acquisition of verbs,  $t(38) = -2.67, p = .011, d = -0.84$ , and word order,  $t(37) = -3.27, p = .002, d = -1.04$ , but not affecting learning nouns,  $t(38) = 0.100, p = .92, d = 0.03$ , adjectives,  $t(38) = 0.94, p = .351, d = 0.30$ , or marker words,  $t(37) = -0.06, p = .950, d = 0.08$ , see Fig. 3.

#### 4. General discussion

There is a chicken-and-egg problem in language acquisition in terms of determining the meaning of vocabulary items while simultaneously discovering the grammatical role of words in the utterance in terms of the language's grammar. Once the learner understands the intended referent of the word then its grammatical category is also evident which can then provide evidence about the grammatical structure of the language. Similarly, knowing the grammar of the language can usefully constrain the possible referents for a given word. For example, if the learner knows that a novel word functions as a noun, then possible mappings to an action or a property of an object are avoided. At the very least, vocabulary and grammar are intertwined, and in some cases it may be essential to acquire one before the other (Gleitman, 1990).

Cross-situational statistics have proven to be a powerful information source available to aid language learning. In the current study, we determined whether previous demonstrations of learning words from one grammatical category (e.g., Chen et al., 2018; Scott & Fisher, 2012; Smith & Yu, 2008) could scale up to acquire mappings between multi-word sentences composed of several grammatical categories appearing with complex transitive scenes, a closer approximation to the ecology of the task facing language learners. In this paper, we explored whether cross-situational statistics could provide a possible source of information to explain how vocabulary and grammar can be acquired simultaneously from complex multi-word utterances and scenes that contain depictions of many semantic featural properties relating to the target language. Our study shows that adult learners can solve the chicken-and-egg problem in language acquisition by keeping track of cross-situational statistics. However, it does not allow us to determine if syntactic knowledge preceded lexical knowledge, or whether the two types of knowledge developed in parallel. Further adaptation of our experimental paradigm could be used to address this question by more fine-grained analyses of the point at which knowledge develops of each language feature.

In Experiment 1 we showed that both vocabulary and grammar could be acquired via cross-situational learning. Without explicit information about the grammatical categories, nor the meaning of the individual words, participants were able to acquire information about the word order of the artificial language. As one of our reviewers pointed out, it is not clear what aspects of the word order participants

actually acquired. For example, to solve the grammaticality judgment trials, participants could attend to the position of the verbs, not paying attention to the flexible word order of subject and object NPs. That is, learning of word order might be quite limited. Nonetheless, it is clear that participants have begun developing a syntactic representation of the language, even if it is unclear how well-developed the representation is at this early stage of acquisition. Participants were also able to quickly discover the word-referent mappings for nouns and verbs and, in the latter part of the study, they also displayed learning of both adjectives and subject/object markers. Though performance for grammatical role markers was not significantly better than chance at any individual test point, there was a significant improvement in learning throughout the study, as reflected in the significant linear contrast for the effect of block for marker words in Experiment 1, demonstrating that this aspect of the language was gradually acquired throughout the study. Acquisition of grammatical role markers was a particularly impressive feat, as these do not occur in the participants' native language and because there were no concrete referents in the scenes for these words. The fact that new grammatical terms could be acquired further suggests that learning in these experiments was not limited to the discovery of words belonging to grammatical categories that participants had already acquired in the course of the development of their first language.

Whereas the link between particular words and aspects of the environment was highly ambiguous in Experiment 1, the set of words and the set of possible referents for those words was provided during training. In Experiment 2, we addressed this pre-specification of the set of referents by doubling the level of ambiguity by presenting two scenes in each learning trial, with just one of those scenes relating to the sentence. Even under these conditions of greater uncertainty, participants were still able to reliably acquire the referents for nouns and verbs, and determine the word order in the syntax of the language. Experiment 3 demonstrated that this finding was robust even when there was no testing until the end of the participants' training on the task – so learning was not dependent on attention being brought to bear on minimal distinctions between scenes presented to participants during lexical test trials. However, learning of adjectives and marker words was less stable in Experiments 2 and 3. This absence of evidence for learning may highlight language features related to the relative order of acquisition of different grammatical categories in natural language, where salience in speech, predictability of words and categories in the utterance, and prominence of the referent in the environment may each contribute to learning (Behrens, 2015; Ellis, 2006; MacWhinney, 2012).

Taken together, these studies demonstrate that cross-situational learning is a mechanism sufficiently powerful to acquire words from different grammatical categories and the word order of the language, under conditions of substantially greater ambiguity than present in previous studies of cross-situational word learning. In these previous studies, participants are typically given a set of words from the same grammatical category (Scott & Fisher, 2012; Yu & Smith, 2007) and a set of referents all of which are referred to by one of the words. Even in more challenging cross-situational word learning situations where not all words refer to referents that are present (e.g., Monaghan et al., 2015; Monaghan & Mattock, 2012) or where words refer to either the basic or superordinate category of an object (Chen et al., 2018), the degree of ambiguity is still highly constrained. The ambiguity present in Experiment 1, further increased in Experiment 2 where there are two scenes accompanying a given sentence, demonstrate that learning is still possible even under conditions that resemble more closely children's experience of utterances and of the environments these utterances refer to.

Previous studies training participants on artificial languages have tended to pre-train participants on the vocabulary included in the language (Amato & MacDonald, 2010; Friederici et al., 2002; Morgan-Short et al., 2014). This may be for practical reasons because the focus

of these studies was on acquisition of the syntax rather than the vocabulary, but it raises the question as to the extent to which grammar and vocabulary can be acquired simultaneously from the language learner's input (Frost & Monaghan, 2016; Marchman & Bates, 1994), or whether one type of knowledge necessarily precedes the other (Gleitman, 1990; Peña et al., 2002; Pinker, 1998). Whether learning is simultaneous or successive for vocabulary and grammar is not yet answered by the results of our study. Our findings show that both can be learned from information present in cross-situational statistical correspondences, but not whether learning of each property of the language is acquired at the same time. Indeed, there are hints of an order of acquisition in the three experiments. Word order and verbs appear to be acquired earliest, followed by nouns. Adjectives and marker words were only acquired only later during training, showing the smallest learning effect. Such a pattern of results suggests that word order, supported by identifying the verb referents, provided the learner access to the language structure, yet this mutuality in acquisition of word order and verbs might instead suggest that the learner bootstraps gradually from information about both word order and vocabulary, with learning proceeding in tandem.

The effective learning of nouns in the experiments was unsurprising, given that the noun advantage in language development is well-documented in the literature (Gentner, 1982; Imai et al., 2008). The verb advantage in our experiments could be an artefact, in part, of the fixed word order of the language (SOV and OSV), which facilitated identification of the verb in final position – a salient position in speech (Freudenthal et al., 2010; Jones & Rowland, 2017). Once the verb was identified, this could then be used to support learning of words from the other grammatical categories. An additional contributor to the verb learning advantage in the current study was that there were also fewer verbs to learn than nouns (four vs eight). In the case of the adjectives, there were only two words to learn, but this did not transfer into an adjective learning advantage over the other content words possibly in part because this was an optional lexical category, with only half the training sentences featuring adjectives, and likely also because nouns are easier to learn than adjectives in natural language situations (e.g., Gasser & Smith, 1998; Sandhofer & Smith, 2007). Finally, the reduced learning of marker words is worth considering. There were only two marker words, which reliably indicated the subject and object of the sentence, and these occurred in each training trial, and as such they were the most frequently occurring words in the artificial language. Yet, learners displayed relatively little knowledge of these markers. In part, this difficulty in learning marker words could be due to their lower salience. The marker words were monosyllabic (in contrast to bisyllabic nouns, verbs and adjectives) and they only occurred within the utterances, i.e., in less prominent positions than nouns and verbs that could occur at utterance boundaries. It is also worth considering that in our artificial language, there was a 250 ms pause between the nouns and the post-nominal markers, which is in contrast to what happens in natural head-final languages like Japanese. It is conceivable that the marker words could have been more readily learned if they functioned like affixes to the nouns (without pause), but future research will need to explore this possibility. The greater difficulty of marker word acquisition also aligns with studies of child language development and second language acquisition (e.g., DeKeyser, 2005; Shi et al., 2006), where the delay and difficulty of learning function words has been well documented. However, differences in the frequency and the variability of words within different grammatical categories in the current artificial language make determining order of acquisition less straightforward.

Nevertheless, our experiments show that, without explicit instruction as to the grammar, and without explicit prior knowledge of vocabulary, both can be learned by determining the co-occurrences between particular features of a scene and individual words in sentences. The apparent co-dependence of learning vocabulary and grammar – the chicken-and-egg problem of language acquisition (Childers et al., 2012; Gleitman, 1990; Gleitman et al., 2005; Marchman & Bates, 1994;



Monaghan & Christiansen, 2008) – is shown to be resolvable by learners tracking cross-situational statistics. Our experiments focused on adult participants, i.e. learners who had already acquired languages and who were likely to possess metalinguistic knowledge of grammatical categories and syntactic relations between them. However, we believe the insights from our study are likely to be relevant to language acquisition by younger learners, too. Previous research (e.g., Scott & Fisher, 2012; Smith & Yu, 2008) has clearly indicated that cross-situational learning can play a role in child language development. Moreover, prior knowledge or experience in acquiring a language in some cases, in fact, hinders learning of a novel system (e.g., Ellis, 2006), so there is no reason to believe that adults would necessarily outperform child learners in this type of experiment. Further studies of cross-situational learning of complex sentence-scene correspondences in infants and children would be necessary to determine the role of this source of information in language acquisition, extending our demonstration of its role in implicit acquisition of multiple language features in adults acquiring a novel language.

### CRedit authorship contribution statement

P.Rebuschat and P.Monaghan developed the study concept and study design. Testing, data collection and preliminary data analysis were performed by all three authors, with additional testing support. P.Rebuschat and P.Monaghan completed the data analysis and interpretation. P.Monaghan and P.Rebuschat drafted the manuscript, and all authors approved the final version of the manuscript for submission.

### Acknowledgements

This research was supported by the LEAD Graduate School and Research Network (grant DFG-GSC1028), a project of the Excellence Initiative of the German Federal and State Governments, by the Economic and Social Research Council, UK (grant ES/L008955/1), and by Lancaster University's Early Career Small Grant Research scheme. Thanks to Niamh Bush, Heather Gray, Helen Joules, and Simon Ruiz for help with data collection.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cognition.2020.104475>.

### References

- Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language*, 19(57), 347–358. <https://doi.org/10.1177/014272379901905703>.
- Amato, M. S., & MacDonald, M. C. (2010). Sentence processing in an artificial language: Learning and using combinatorial constraints. *Cognition*, 116(1), 143–148. <https://doi.org/10.1016/j.cognition.2010.04.001>.
- Behrens, H. (2015). The acquisition of grammatical categories. In E. L. Bavin, & L. R. Naigles (Eds.), *The Cambridge handbook of child language* (pp. 250–270). Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9781316095829.012>.
- Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, 110(28), 11278–11283. <https://doi.org/10.1073/pnas.1309518110>.
- Chen, C. H., Zhang, Y., & Yu, C. (2018). Learning object names at different hierarchical levels using cross-situational statistics. *Cognitive Science*, 42, 591–605. <https://doi.org/10.1111/cogs.12516>.
- Childers, J. B., & Paik, J. H. (2009). Korean- and English-speaking children use cross-situational information to learn novel predicate terms. *Journal of Child Language*, 36, 201–224. <https://doi.org/10.1017/S0305000908008891>.
- Childers, J. B., Heard, M. E., Ring, K., Pai, A., & Sallquist, J. (2012). Children use different cues to guide noun and verb extensions. *Language Learning and Development*, 8(3), 233–254. <https://doi.org/10.1080/15475441.2011.585285>.
- Clerkin, E. M., Hart, E., Reh, J. M., Yu, C., & Smith, L. B. (2017). Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of The Royal Society B: Biological Sciences*, 372(1711), <https://doi.org/10.1098/rstb.2016.0055>.
- Cunillera, T., Laine, M., Càmarà, E., & Rodríguez-Fornells, A. (2010). Bridging the gap

- between speech segmentation and word-to-world mappings: Evidence from an audiovisual statistical learning task. *Journal of Memory and Language*, 63(3), 295–305. <https://doi.org/10.1016/j.jml.2010.05.003>.
- DeKeyser, R. (2005). What makes second-language grammar difficult? A review of issues. *Language Learning*, 55, 1–25.
- Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, saliency, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164–194. <https://doi.org/10.1093/applin/aml015>.
- Freudenthal, D., Pine, J., & Gobet, F. (2010). Explaining quantitative variation in the rate of optional infinitive errors across languages: A comparison of MOSAIC and the variational learning model. *Journal of Child Language*, 37(3), 643–669. <https://doi.org/10.1017/S0305000909990523>.
- 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. <https://doi.org/10.1073/pnas.012611199>.
- Frost, R. L. A., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, 147, 70–74. <https://doi.org/10.1016/j.cognition.2015.11.010>.
- Gasser, M., & Smith, L. B. (1998). Learning nouns and adjectives: A connectionist account. *Language and Cognitive Processes*, 13, 269–306. <https://doi.org/10.1080/016909698386537>.
- Gentner, D. (1982). *Why nouns are learned before verbs: Linguistic relativity versus natural partitioning*. In Kuczaj, S. A. (Ed.), *Language development: Vol. 2. Language, thought, and culture* (pp. 301–334). Hillsdale, NJ: Erlbaum.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3–55. [https://doi.org/10.1207/s15327817la0101\\_2](https://doi.org/10.1207/s15327817la0101_2).
- Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language Learning and Development*, 1(1), 23–64. [https://doi.org/10.1207/s15473341lhd0101\\_4](https://doi.org/10.1207/s15473341lhd0101_4).
- Hamrick, P., & Rebuschat, P. (2012). How implicit is statistical learning? In P. Rebuschat, & J. N. Williams (Eds.), *Statistical learning and language acquisition* (pp. 365–382). Berlin: Mouton de Gruyter.
- Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, 79, 979–1000. <https://doi.org/10.1111/j.1467-8624.2008.01171.x>.
- Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, 98, 1–21. <https://doi.org/10.31234/osf.io/wdbpc>.
- MacWhinney, B. (2012). The logic of the unified model. In S. M. Gass, & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 211–227). London: Routledge. <https://doi.org/10.4324/9780203808184.ch13>.
- Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language*, 21, 339–366. <https://doi.org/10.1017/S0305000900009302>.
- Monaghan, P., & Christiansen, M. H. (2008). Integration of multiple probabilistic cues in syntax acquisition. In H. Behrens (Ed.), *Corpora in language acquisition research: History, methods, perspectives* (pp. 139–164). Amsterdam: John Benjamins. <https://doi.org/10.1075/tilar.6.08mon>.
- Monaghan, P., & Fletcher, M. (2019). Do sound symbolism effects for written words relate to individual phonemes or to phoneme features? *Language and Cognition*, 11(2), 235–255. <https://doi.org/10.1017/langcog.2019.20>.
- Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition*, 123, 133–143. <https://doi.org/10.1016/j.cognition.2011.12.010>.
- Monaghan, P., Mattock, K., Davies, R., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from cross-situational statistics. *Cognitive Science*, 39, 1099–1112. <https://doi.org/10.1111/cogs.12186>.
- Morgan, J., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.
- 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. <https://doi.org/10.1017/S1366728912000715>.
- Peña, M., Bonatti, L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, 298, 604–607. <https://doi.org/10.1126/science.1072901>.
- Pentimonti, J., O'Connell, A., Justice, L., & Cain, K. (2015). The dimensionality of language ability in young children. *Child Development*, 86(6), 1948–1965. <https://doi.org/10.1111/cdev.12450>.
- Pinker, S. (1998). Words and rules. *Lingua*, 106, 219–242. [https://doi.org/10.1016/S0024-3841\(98\)00035-7](https://doi.org/10.1016/S0024-3841(98)00035-7).
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48, 21–69. [https://doi.org/10.1016/0010-0277\(93\)90057-3](https://doi.org/10.1016/0010-0277(93)90057-3).
- Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press. <https://doi.org/10.7551/mitpress/9636.001.0001>.
- Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive Psychology*, 66(1), 30–54. <https://doi.org/10.1016/j.cogpsych.2012.09.001>.
- Sandhofer, C., & Smith, L. B. (2007). Learning adjectives in the real world: How learning nouns impedes learning adjectives. *Language Learning and Development*, 3(3), 233–267. <https://doi.org/10.1080/15475440701360465>.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime (Version 2.0)*. Pittsburgh,

- PA: Psychology Software Tools Inc.
- Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*, 163–180. <https://doi.org/10.1016/j.cognition.2011.10.010>.
- Shi, R., Werker, J. F., & Cutler, A. (2006). Recognition and representation of function words in English-learning infants. *Infancy*, *10*(2), 187–198.
- Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, *61*, 39–61. [https://doi.org/10.1016/s0010-0277\(96\)00728-7](https://doi.org/10.1016/s0010-0277(96)00728-7).
- Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*, 1558–1568. <https://doi.org/10.1016/j.cognition.2007.06.010>.
- Suanda, S. H., & Namy, L. L. (2012). Detailed behavioral analysis as a window into cross-situational word learning. *Cognitive Science*, *36*, 545–559. <https://doi.org/10.1111/j.1551-6709.2011.01218.x>.
- Tomblin, J. B., & Zhang, X. (2006). The dimensionality of language ability in school-age children. *Journal of Speech, Language, and Hearing Research*, *49*, 1193–1208. [https://doi.org/10.1044/1092-4388\(2006/086\)](https://doi.org/10.1044/1092-4388(2006/086)).
- Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375–382. <https://doi.org/10.1016/j.cognition.2013.02.015>.
- Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. <https://doi.org/10.1111/j.1467-9280.2007.01915.x>.