

The role of feedback and instruction on the cross-situational learning of vocabulary and morphosyntax: Linear mixed effects reveal local and global effects on acquisition

Padraic Monaghan^(1,2), Simón Ruiz⁽³⁾, and Patrick Rebuschat^(1,3)

⁽¹⁾ Lancaster University, UK

⁽²⁾ University of Amsterdam, The Netherlands

⁽³⁾ University of Tübingen, Germany

Word count: 7827

Published in Second Language Research

<https://doi.org/10.1177/0267658320927741>

Acknowledgments: This work was supported by the International Centre for Language and Communicative Development (LuCiD) at Lancaster, funded by the Economic and Social Research Council (UK) [ES/L008955/1], and by the LEAD Graduate School and Research Network [DFG-GSC1028], a project of the Excellence Initiative of the German Federal and State Governments.

Corresponding author: Professor Padraic Monaghan, Department of Psychology, Lancaster University, U.K., E-mail: p.monaghan@lancaster.ac.uk

Abstract

First language acquisition is implicit, in that explicit information about the language structure to be learned is not provided to children. Instead, they must acquire both vocabulary and grammar incrementally, by generalizing across multiple situations that eventually enable links between words in utterances and referents in the environment to be established.

However, this raises a problem of how vocabulary can be acquired without first knowing the role of the word within the syntax of a sentence. It also raises practical issues about the extent to which different instructional conditions – about grammar in advance of learning or feedback about correct decisions during learning – might influence second language acquisition of implicitly experienced information about the language. In an artificial language learning study, we studied participants learning language from inductive exposure, but under different instructional conditions. Language learners were exposed to complex utterances and complex scenes and had to determine the meaning and the grammar of the language from these co-occurrences with environmental scenes. We found that learning was boosted by explicit feedback, but not by explicit instruction about the grammar of the language, compared to an implicit learning condition. However, the effect of feedback was not general across all aspects of the language. Feedback improved vocabulary, but did not affect syntax learning. We further investigated the local, contextual effects on learning, and found that previous knowledge of vocabulary within an utterance improved learning but that this was driven only by certain grammatical categories in the language. The results have implications for theories of second language learning informed by our understanding of first language acquisition as well as practical implications for learning instruction and optimal, contingent adjustment of learners' environment during their learning.

Introduction

The processes by which children acquire their first language have important implications for theories of second language acquisition. In order to understand an utterance, the language learner has to not only develop an understanding of both the meaning of words within the utterance through acquisition of the vocabulary, but also determine the grammatical roles of those words from the syntactic structure of the sentence. Such an issue faces both first and second language learners, and has raised a long-standing theoretical debate about how these two interlinked aspects of language can be learned simultaneously (Gentner, 1982; Gleitman, 1990; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005). Consider, for instance, the transitive verb “give”. Observing one person handing a gift to another accompanied by the sentence “Patrick gives the present to Simón”. Without already knowing the vocabulary for Patrick and Simón as well as the grammar that specifies word order in English it would not be possible to ascertain from the scene whether “give” means giving or receiving. Only prior acquisition of vocabulary and grammar can result in accurate performance.

So how do learners resolve this “chicken and egg” problem of requiring vocabulary to understand grammar and requiring grammar to determine the meaning of the vocabulary? One solution is to focus the learner on acquiring one aspect of the language. For instance, many previous laboratory-based studies that train participants to acquire vocabulary and grammar typically expose learners to the vocabulary first, and then present this pre-acquired vocabulary in sentences to support development of the grammar (e.g., Friederici, Steinhauer, & Pfeifer, 2002, in a study with adults). Other studies similarly distinguish stages of learning vocabulary from grammar (De Graaff, 1997; DeKeyser, 1995; Morgan-Short et al., 2010, 2012). However, it is not known to what extent this separation of vocabulary and grammar training is necessary, or even useful, for supporting language learning, and it seems to violate the situation that occurs in first language acquisition under naturalistic conditions.

In first language acquisition, infants tend to hear words spoken in multi-word utterances (Cameron-Faulkner, Lieven, & Tomasello, 2003), and are simultaneously surrounded by a multitude of possible referents to which these words may relate. Children do not receive this explicit pre-training in vocabulary before they receive the vocabulary embedded in the utterance. Furthermore, they are not provided with information about the referent for each word (Yu et al., 2009). This has to be acquired implicitly. Indeed, determining how each of the words in the utterance refer to aspects of the environment is a profoundly difficult problem because the possible referents are unconstrained (Quine, 1960). For any individual word, there are infinite possible referents in the environment to which that word refers – so if a speaker of a native language utters “gavagai” when a rabbit runs past, the hearer cannot know if the utterance refers to the rabbit, the rabbit’s ear, its colour or texture, the action of running, a tasty meal, or the entire scene.

To quantify this ambiguity, Yu and Ballard (2007) analysed a small corpus of child-directed speech, at the same time encoding which potential objects were around the child as each utterance was spoken. They found that multiple potential objects were present when the child heard each word, but over multiple occurrences of the word particular words tended to co-occur with particular objects that were within the child’s view at that moment (Siskind, 1996). Yu and Smith (2007) and Smith and Yu (2008) showed that adults and infants, respectively, could learn particular word-referent mappings from these cross-situational statistics. When sets of words and sets of objects co-occurred, it is not possible to determine which word refers to which object, but over multiple trials, as the words and objects vary, it becomes detectable to the learner that certain words always occur when certain objects are in view.

This cross-situational learning proves to be a powerful mechanism for acquiring vocabulary from multiple words and multiple objects presented simultaneously. However,

this experimental situation still does not reflect the complexity facing the learner of vocabulary and grammar in an unknown language. In these studies of cross-situational learning, the words were all nouns and their referents were always present. In naturalistic language acquisition, words sometimes occur without any concrete referents. Yu and Ballard (2007), for instance, showed that verbs were also present in the multi-word utterances of parents speaking to infants, as well as function words which served a grammatical role but without any precise link to objects in the environment.

Monaghan and Mattock (2012) showed that adult language learners were able to cope with the added complexity of words occurring without referents for all words present in the environment. They presented learners with an artificial language comprising one word referring to one of two objects in a visual display and one word that did not refer to anything in the display. The referring and non-referring word varied in order across utterances, and participants were required to respond whether they felt the left or right object was referred to by the artificial language sentence. For a single trial, participants would not be able to determine which was the target object, but over multiple trials cross-situational statistics would give information about co-occurrences between individual words and objects. A condition where additional function words that indicated which was the referring word was also tested. Participants were able to learn the meaning of the referring words, and learning was boosted by the condition where additional function words were present (see also Koehne & Crocker, 2015). Thus, even though participants were not explicitly told about the language, they were able to use the grammatical information to support their acquisition of the vocabulary.

Studies of cross-situational word learning for verbs, similar in design to those by Yu and Smith (2007), have shown that child learners can acquire word-action mappings in a similar manner to acquisition of nouns (Childers, Heard, Ring, Pai, & Sallquist, 2012; Scott

& Fisher, 2012). Monaghan et al. (2015) further showed that adult learners can cope with further complexity of referring words from multiple grammatical categories being present in the language. They presented participants with two scenes of a shape object performing a movement (one per scene) while an intransitive sentence in an artificial language comprising a noun and a verb played in the background. Participants had to decide whether the left or the right scene corresponded to the sentence. The shape object and the action of one of the scenes were the target referents for the noun-verb utterance, and over multiple trials participants could learn again the co-occurrence between particular nouns and objects and verbs and actions, if they were able to identify that two grammatical categories occurred in the utterance and referred to different aspects of the scene. Monaghan et al. (2015) found that both the nouns and the verbs could be learned from these cross-situational statistics, without participants needing prior information about the grammar of the language.

In each of these previous studies of cross-situational learning, the focus had been on vocabulary acquisition, with simple grammatical structures tested for the extent to which they can support vocabulary acquisition. But none of these studies tested or questioned participants about what aspects of the grammar were acquired at the same time as the vocabulary – they provide only accumulating evidence that the grammar could be used to support vocabulary learning. Nonetheless, these studies of cross-situational learning demonstrate that vocabulary can be acquired from utterances that comprise more than lists of words from the same grammatical category, so it is not necessary for vocabulary to be acquired prior to those words occurring in multi-word utterances.

However, the utterances so far tested in cross-situational studies remain extremely simple (comprising only intransitive sentences), and as such do not address the complexity of natural language grammar, where utterances may be substantially longer and be composed of words from several grammatical categories. The use of only intransitive sentences also does

not address the ambiguity that can come from resolving subjects and objects of verbs, as in the give/receive example.

Rebuschat et al. (under review) recently extended the cross-situational paradigm to a more complex design using a language with transitive sentences comprising nouns, verbs, adjectives, and grammatical marker words that indicate the subject and object of the sentence. The grammar was based on Japanese, with word order either SOV or OSV, with the marker words occurring after the subject or object noun to indicate its syntactic function. As well as the grammar being more complicated, the scenes viewed by participants were also more complex, involving two aliens (referred to by the nouns in the language) with different colours (referred to by the adjectives in the language) undertaking an action (described by the verb). As with the other studies of cross-situational learning, two scenes appeared, and participants had to select which scene was described by the utterance. If participants were able to track the cross-situational statistics between each word in the utterance and objects and their roles and colours, and the action that the objects performed, then learners should be able to select the target scene with increasing accuracy. Despite the complexity of the language and the scene, learning in an adult population was successful: vocabulary in each grammatical category was acquired greater than chance.

Rebuschat et al. (under review) also tested whether participants could learn the grammar of the language, by testing ability to recognize grammatical versus ungrammatical word sequences. They found that this was also successful. Thus, the chicken and egg problem of acquisition of vocabulary and of grammar was shown to be resolvable through cross-situational statistics, with learners tracking multiple possible mappings between words and aspects of complex scenes in order to hone in on the co-occurring features of the environment and words that labeled these features.

Effects of Feedback and Explicit Information on Language Learning

These previous studies of cross-situational learning demonstrate that learners are adept at detecting complex co-occurrences between words in utterances and multiple features of scenes. Furthermore, these studies show that feedback on whether or not the learner is making correct assumptions is not necessary for acquisition. In all these studies, input is presented with no feedback given on responses. This shows the power of language learning – that for acquisition of a simple vocabulary and grammar, it can proceed in the absence of feedback – but what is not known is whether feedback can promote language learning under these cross-situational learning conditions that first language learners find themselves. Certainly, infants acquiring their first language receive some feedback (implicit and explicit) on their attempts to communicate and use words referentially (Baldwin, 1993; Miller & Lossia, 2013; Tamis-LeMonda, Kuchirko, & Song, 2014), but are rarely provided with unambiguous feedback about target referents for words (Yu et al., 2009).

In second language learning, feedback is also known to provide a boost to learning (for reviews see, e.g., N. Ellis, 1990; Goo et al., 2015; Li, 2010; Nassaji, 2016; and Plonsky & Brown, 2015, for review of these reviews). This is particularly the case for more explicit types of feedback (Lightbown & Spada, 1990; Mackey, 2006). Feedback helps learners notice the gap between their representation of the second language and that of the target (Nassaji, 2016). Feedback can also assist learners in acquiring difficult target forms, including rare, non-salient, or semantically redundant forms (Loewen, 2012; White, 1991).

However, until now, the role of explicit feedback about correct selection of referents in cross-situational learning tasks has not been comprehensively tested. In this study we tested the extent to which feedback about correct scene selections supported learning. Theoretically, this is important to know how such external information about learning inter-relates with acquisition of inductively derived statistical information about vocabulary and

grammar. Practically, it is also vital to determine what manipulations of the learning environment support language learning, and whether these affect in particular vocabulary or grammar, or apply equally to both. In order to test the role of feedback on acquiring statistical information about language structure, we used a minimal form of feedback: a bell sound indicating that the participant had made a correct mapping between utterance and scene in the environment. This is distinct from the more explicit, directive feedback about language structure that tends to occur in second language learning (see review by Nassaji, 2016); however, we aimed to determine as a starting point how minimal feedback – without providing information about particular words or grammatical structures – may help guide the learner.

The cross-situational learning studies, such as Rebuschat et al. (under review), have shown that learning vocabulary and grammar is possible simultaneously, and without explicit instruction about either vocabulary or grammar. However, there is a wealth of data showing that explicit information about the language to be learned supports learning (R. Ellis, 2015; N. Ellis, 2005; Goo et al., 2015; Monaghan, Schoetensack, & Rebuschat, 2019; Norris & Ortega, 2000; Spada & Tomita, 2010). Yet, how precisely this explicit information about language structure impacts on language representation in second language acquisition is not fully understood. It could, for instance, affect the participants' understanding of the grammatical structure, which then facilitates accumulation of vocabulary (thus, solving the chicken-and-egg problem of grammar and vocabulary learning by prioritising initial grammar acquisition), or it could be that it only improves grammar learning and vocabulary acquisition proceeds largely independently of this grammatical knowledge. Monaghan, Schoetensack and Rebuschat (2019) made some progress in addressing this question of how explicit information about language structure affects language learning. Using a paradigm similar to Monaghan et al. (2015), where utterances comprised a noun and a verb, in free word order,

with function words indicating the grammatical categories preceding the function word, participants were instructed about the role of the function words, or were left to acquire the language cross-situationally with no information about the grammar. Participants were able to learn the meaning of the nouns and the verbs more effectively in the condition with explicit grammar instruction.

However, this study only tested acquisition of vocabulary, and was limited in the range of grammatical categories included in the language. In order to enrich our understanding of the points at which explicit information impacts on otherwise implicit acquisition of language from cross-situational statistics, we require a test of how explicit information about the grammar can affect both grammatical knowledge and vocabulary knowledge, preferably of vocabulary present in a variety of grammatical categories, including both content and function words, the latter of which have a grammatical role in the language, and are less likely to be accompanied by explicit information about their meaning (Paradis, 2009), such as reliable co-occurrence with a feature of the environment, and less likely to be governed by explicit control over their usage (Groom & Pennebaker, 2002).

A second aim of the current study was thus to test the role of explicit instruction about the grammatical structure of a more complex artificial language, taken from Rebuschat et al. (under review), involving tests of grammatical structure as well as different vocabulary types. We therefore compared groups that were given explicit information about the grammatical structure to groups that were given no such information and had to derive this knowledge from information provided in the input.

A further advance in determining how and where explicit and implicit information can impact on vocabulary and grammar learning is to investigate the very local learning context effects applying during acquisition. Classic statistical approaches to behavioural studies (e.g., N. Ellis, 2006; Monaghan & Mattock, 2012; Nassaji, 2016; Yu & Smith, 2007)

examine summary statistics in order to determine whether there are differences between one group of responses and another group of responses (e.g., for implicit or explicit conditions in a learning study). These statistical approaches measure the global effects of learning. However, they fail to take into account the specific situation that learners experience from one learning trial to the next. Classic statistics can show that explicit instruction improves vocabulary acquisition, but to what extent does the learner's previous exposure to a linguistic structure affect their learning of this structure the next time they come across it? Using contemporary statistical techniques – specifically, mixed effects modelling approaches (Baayen et al., 2008) – enables studies of language learning to investigate the contingency of learning based on previous performance as learning proceeds (Cunnings, 2012; Linck & Cunnings, 2015), providing insight from both the global effects of instruction and feedback on learning, but also from the very local effects of previous trial behaviour on current trial performance.

In the analyses we undertake in this paper, to investigate the role of feedback and the consequences of implicit and explicit instruction, we use these mixed effects methods to hone in on precisely how learning proceeds trial by trial. We investigate how performance on the current learning trial is affected by whether participants made a correct or incorrect response on a previous trial containing the same information in terms of the nouns, verbs, and adjectives that appeared. This enables us to test whether instruction and feedback affects the way in which this previous information is used in the current trial, and how learning transfers across different grammatical categories. It could be that knowledge of a particular noun is influenced by whether the noun was correctly interpreted in the previous trial. The transfer effects could also be more complex: For instance, it could also be the case that knowledge of the verb in a particular trial influences knowledge of the noun that the verb now appears with, thus local contextual knowledge may be direct (supporting accuracy for the same word) or

this context may have broader effects (supporting accuracy for words occurring alongside the word that is previously correctly responded to). This local information enables us to determine how learning from one source of information scaffolds learning across the language as a whole. To implement this local context, we included as fixed effects the participants' response to the previous utterance containing the same verb, noun, or adjective as the current trial. We used previous accuracy to be in line with analyses performed by Roembke and McMurray (2016) and Trueswell et al. (2013) who included a predictor of previous accuracy on the previous noun in their cross-situational statistical studies analysis.

We next report the experimental study varying feedback and instruction conditions in learning a complex artificial language from cross-situational statistics. We predict that feedback about performance will improve learning of both vocabulary and grammar as training proceeds. We also predict that explicit information about the language structure will have a direct influence on representation of the grammatical information, as the explicit instruction pertains to the grammatical structure of the language. However, we predict that explicit instruction will also have an indirect influence on acquisition of vocabulary within this grammar, due to the interdependency of vocabulary and grammatical structure in early stages of learning. Finally, we perform explorative analyses to determine where these different conditions of learning (feedback, implicit or explicit instruction) affect the way in which information transfers between language structures within the language by analyzing performance at the item by item level, considering how global learning conditions affect local learning.

Method

Participants

Ninety university students (Mean age = 22.1, SD = 3.3, 57 women) volunteered to participate. All participants were native speakers of English, and none had a background in Japanese. Participants were remunerated for their time. 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. Data collection took place in Lancaster, UK, and in Tübingen, Germany.

Materials

The materials were closely based on the study of Rebuschat et al. (under review). 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 by E-Prime 2.0 (Schneider, Eschman, & Zuccolotto, 2002). Figure 1 shows a sample trial.

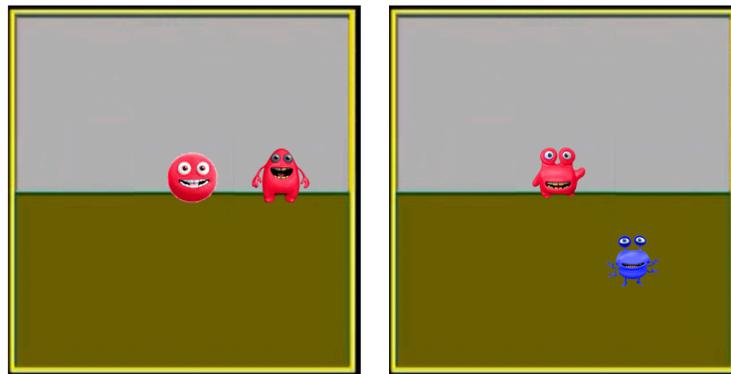


Figure 1. Example of a training trial. An agent and a patient appear in each scene, with varying colours (red or blue) performing an action. One of the scenes is referred to by the spoken sentence accompanying the trial.

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 role markers and reliably indicated if the preceding noun referred to the subject or the object of the sentence. We used more nouns than verbs and more verbs than adjectives to mimic the relative type frequency of these grammatical categories in natural language (see Walker, Monaghan, & Rebuschat, in press, for more discussion on this point). Word-referent mappings were randomly generated for each participant to control for preferences in associating certain sounds to objects, actions, or colours.

The grammar of the artificial language was verb final with variable word order, similar to the grammar of Japanese. Sentences could either be SOV or OSV, i.e. the verb (V) 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 role 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 thus ranged between five and seven words. We generated 192 unique 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 unique 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.

Sequence

Word order	First	Second	Third
SOV	NP _{subj} (A N SUBJECT)	NP _{obj} (A N OBJECT)	V
	NP _{subj} (A N SUBJECT)	NP _{obj} (N OBJECT)	V
	NP _{subj} (N SUBJECT)	NP _{obj} (A N OBJECT)	V
	NP _{subj} (N SUBJECT)	NP _{obj} (N OBJECT)	V
OSV	NP _{obj} (A N OBJECT)	NP _{subj} (A N SUBJECT)	V
	NP _{obj} (A N OBJECT)	NP _{subj} (N SUBJECT)	V
	NP _{obj} (N OBJECT)	NP _{subj} (A N SUBJECT)	V
	NP _{obj} (N OBJECT)	NP _{subj} (N SUBJECT)	V

Procedure

Participants were randomly distributed into one of three conditions, implicit ($n = 30$), explicit ($n = 31$), and feedback ($n = 29$). After providing informed consent, participants in all conditions were informed that they would learn a new language, spoken by the “friendly inhabitants of a distant planet”.

Participants in all three conditions first completed two practice trials with two scenes involving aliens performing actions, accompanied by a sequence that followed the grammar of the language (so containing the grammatical role words) and contained nonsense words not used in the main part of the study. The aliens, their colours (green), and the actions used in the practice also did not occur in the main part of the study.

Before these practice trials, participants in the explicit condition were given information about the grammatical role words, as follows: “In the sentence, there are two marker words that tell you who is the subject (= the person who does something) and who is the object (= the person to whom something happens). These marker words are ‘tha’ and ‘noo’.” Participants in

the implicit, and feedback conditions received no such instructions as to the structure of the language.

Participants in all conditions were then trained and tested on the artificial language over twelve blocks of training, which were then immediately followed by two vocabulary testing blocks and a grammatical structure testing blocks. The testing blocks occurred at the end of training so as not to influence performance during training.

Training blocks. Participants viewed the animated scene in which two alien characters performed an action. They then heard the sentence describing the scene, e.g., for referring to the target scene shown on the left panel of Figure 1:

garshal chilad tha garshal sumbad noo thislin
red alien₂ OBJECT red alien₁ SUBJECT jumps

gloss: “red alien₁ jumps over red alien₂”

This was then followed by another presentation of the action. Participants were required to make a response to either the left or the right scene as being referred to by the sentence, by pressing a button on a computer keyboard. In the feedback condition, an auditory bell sound was played if the participant responded correctly. In the other conditions, no feedback was given. There were 16 trials in a training block, with each alien, action, and colour, and each word occurring a balanced number of times within each block.

Test blocks. The testing procedure was the same for all conditions, except that in the feedback condition, as in the training blocks, participants were given auditory feedback as to whether their response was correct or not. The acquisition of vocabulary was assessed in a test block after every training block, by means of a two-alternative forced-choice task. Participants were presented with two animated scenes and played a test sentence. Their task was to decide as quickly and accurately as possible which the scene the sentence referred to. Each lexical category was 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 role 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.

Each vocabulary test block consisted of 24 trials, four for testing each verb, four for testing the adjectives (two in each position), eight for testing each noun, and eight testing the marker words so as to include a wide variety of contexts in which the marker words were tested. Trials occurred in randomised order. An example test trial for a verb test is shown in Figure 2.

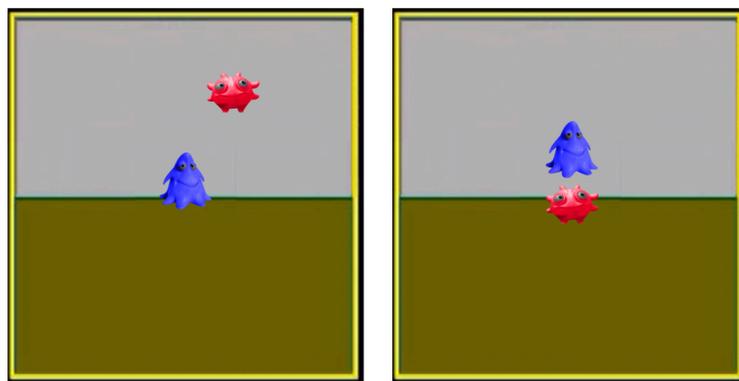


Figure 2. Example of a lexical test trial, measuring knowledge of the verb (aliens, colours, and subject/object roles are the same in both scenes).

After the vocabulary testing, acquisition of word order was then tested by a grammaticality judgment task where word order was varied between grammatical and ungrammatical sequences. 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. Feedback was provided on response accuracy only for participants in the feedback condition. Half the trials followed the grammar of the artificial language, with SOV and OSV sentence patterns carefully counterbalanced. The other half involved sentences with syntactic violations (*SVO, *OVS, *VSO, *VOS). Presentation order within each block was randomised.

After the final test block, participants completed a debriefing and language background questionnaire.

Statistical analyses

The use of mixed-effects regression modelling is increasingly advocated as a tool to analyse second language data (e.g., Cunnings, 2012; Cunnings & Finlayson, 2015; Godfroid, 2020; Gries, 2015; Linck & Cunnings, 2015; Loewen & Godfroid, 2020; Murakami, 2016), since it allows accounting for the effects of multiple factors on the dependent variable as well as accounting for individual variation between different participants and stimuli in the same study (Linck, 2016). Hence, experimental condition effects and inter-individual differences can be estimated within a single analysis. In this study, given our aim of assessing contextual effects at the item level, we used generalized linear mixed-effects modelling (Baayen et al., 2008), specifically logistic mixed-effects regression models (Jaeger, 2008). This type of model framework allows for modeling accuracy (correct vs. incorrect) for each participant on each item instead of estimating the average or total score across participants, and thus allows us to examine effects of broad learning conditions as well as the local context effects of the particular learning trial by investigating variation at the participant and the item level simultaneously.

Consequently, logistic mixed-effects regression models were used to assess both global and local effects of acquisition, i.e., the effects of the different exposure conditions (feedback, explicit, and implicit) and the effects of the more immediate context in which a stimulus is processed, respectively. For the analysis of word order in the testing data, however, linear regression models were built to model the data from the grammaticality judgement task, as the raw accuracy scores obtained from this task were transformed into d' scores (Wickens, 2002) to control for response bias, and thus there was just one dependent variable measure per participant in the analysis. In the logistic mixed-effects models, on the other hand, accuracy was modelled as a binary dependent variable (correct = 1, incorrect = 0). The mixed-effects models were implemented using the lme4 package (version 1.1-21; Bates et al., 2015) of R (version 3.6.0; R Core Team, 2019).

In the mixed effects model of training, Block (12 blocks) was entered as a fixed effect, and both linear and quadratic effects were tested, using orthogonal polynomials (Mirman, 2017). For the analysis of global effects in the training trials, Group (Feedback vs. Explicit vs. Implicit) was also entered, as was the interaction between Group and Block, to determine whether global training conditions affected learning over time. In order to test local effects of learning, the accuracy for the previous trial in which the nouns, adjectives, and verb occurred (encoded for each word in the current trial as previous noun, previous verb, and previous adjective accuracy) were simultaneously entered as fixed effects in the same logistic mixed-effects model. We also tested the interaction between the local context in terms of the different linguistic features (previous nouns, previous verb, and previous adjectives) with Block to determine if the influence of previous knowledge varied over training. In addition, we tested the effect of the local effects interacting with Group (the global effect), but when including this we omitted the interaction of local effects with Block as the model failed to converge when all 2-way interactions were included. As random

crossed effects (Baayen et al., 2008), intercepts for subjects and items as well as by-subject random slopes for Block and local effects (the model failed to converge when their interaction was also included), and by-item random slopes for Block, Group and their interaction.

In the testing trials, the variable Linguistic Feature (Adjective vs. Marker word vs. Noun vs. Verb) was included as a fixed effect in the logistic mixed-effects models in order to determine learning of the different grammatical categories in the language. We also included the global effect of Group as a fixed effect, and the interaction between Linguistic Feature and Group to determine whether feedback and instruction affected acquisition of one linguistic feature more than another. It was not possible to test the effect of local context effects during testing, due to the sparsity of the occurrence of particular words in the testing set. For subject random effects, we included intercepts and all fixed effects and interactions as slopes. However, including random effects over items resulted in the model being rank deficient (with too many variables included in the model fitting to the number of data points). In order to effectively measure the role of the different Linguistic Features, we therefore did not include item intercepts and slopes from this model.

For the word order test, in order to examine global effects (i.e., effects of feedback and instruction), Group was entered as a predictor in the linear regression model.

The categorical variables Group and Test were represented by orthogonal (i.e., independent) contrast-coded variables. Orthogonal contrast codes were used to compare means of combined groups, with each uncorrelated regression coefficient representing test of significance of group mean differences, or combinations of group means, and the intercept representing the overall average of all groups (i.e., grand mean) (see Cohen et al., 2003, p. 332) Importantly, their use may increase the statistical power to detect effects if they truly exist, compared to less focused, omnibus tests (Cohen et al. 2003). In the case of Group, it

was entered as two orthogonal contrast codes. The first code (CG1: Feedback = 2, Explicit = -1, Implicit = -1) compared the Feedback group to the Explicit and Implicit groups combined (i.e., average over these two groups). The second code (CG2: Feedback = 0, Explicit = 1, Implicit = -1) compared the Explicit and Implicit groups to each other. Regarding Test, it was entered as three orthogonal contrast codes. The first (CT1: Adjective = 3, Marker word = -1, Noun = -1, Verb = -1) compared the adjective test to all other tests. The second (CT2: Adjective = 0, Marker word = 2, Noun = -1, Verb = -1) compared the marker word test to the combination of the noun and verb tests. The final contrast code (CT3: Adjective = 0, Marker word = 0, Noun = 1, Verb = -1) compared the noun and verb tests to each other.

Results

Performance across training blocks

The performance during the training task is summarized in Figure 3 and Table 1.

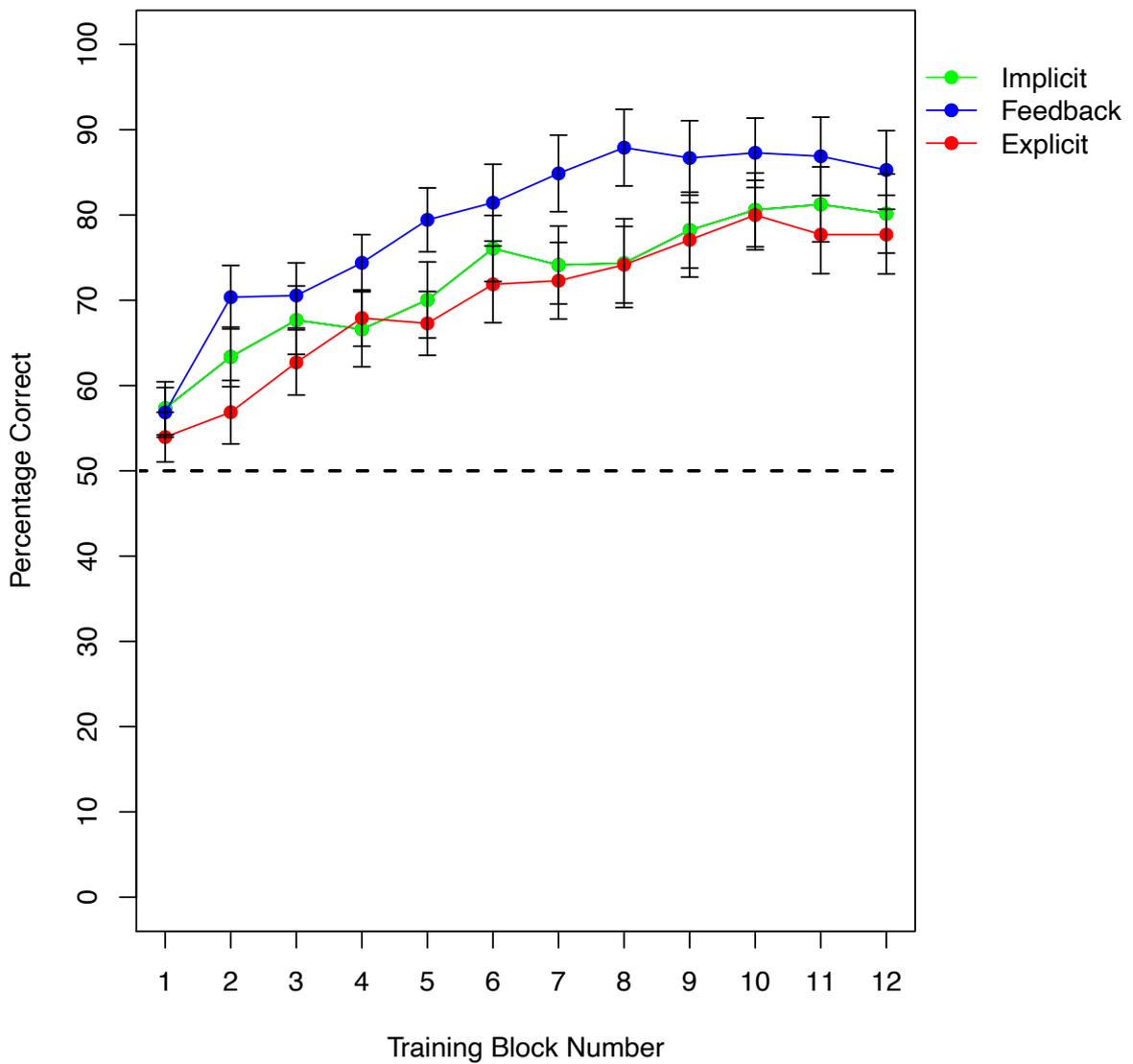


Figure 3. Performance for training blocks. Error bars show standard error of the mean. Dotted line at 50 shows chance performance.

Table 1. Descriptive statistics for training blocks.

Groups	Training Block												
	1	2	3	4	5	6	7	8	9	10	11	12	
Implicit ^a	<i>M</i>	57.33	63.36	67.67	66.59	70.04	76.08	74.14	74.35	78.23	80.60	81.25	80.17

<i>SD</i>	16.79	18.80	21.59	23.70	24.00	20.87	24.65	28.02	23.95	23.29	23.68	25.00
<i>SE</i>	3.12	3.49	4.01	4.40	4.46	3.88	4.58	5.20	4.45	4.32	4.40	4.64
Feedback ^b												
<i>M</i>	56.85	70.36	70.56	74.40	79.44	81.45	84.88	87.90	86.69	87.30	86.90	85.28
<i>SD</i>	12.95	22.07	20.17	22.90	19.51	22.33	19.82	18.39	20.20	21.50	19.59	19.80
<i>SE</i>	2.33	3.96	3.62	4.11	3.50	4.01	3.56	3.30	3.63	3.86	3.52	3.56
Explicit ^c												
<i>M</i>	53.96	56.88	62.71	67.92	67.29	71.88	72.29	74.17	77.08	80.00	77.71	77.71
<i>SD</i>	15.95	20.39	20.92	18.11	20.48	24.66	24.60	24.61	23.92	22.29	25.09	25.25
<i>SE</i>	2.91	3.72	3.82	3.31	3.74	4.50	4.49	4.49	4.37	4.07	4.58	4.61

^a $n = 29$. ^b $n = 31$. ^c $n = 30$.

We found that performance improved over training, with the main effect of Block significant as both a linear (estimate = 2.677, $SE = 0.260$, $p < .001$) and a quadratic (estimate = -0.299, $SE = 0.129$, $p = .021$) effect. In terms of global properties of learning, there was a significant main effect of Group, with accuracy higher for the Feedback group than the combined Explicit and Implicit group (estimate = 0.192, $SE = 0.096$, $p = .045$), which did not differ from one another (estimate = 0.001, $SE = 0.163$, $p = .994$). The interaction between Group and Block was significant in the quadratic contrast (estimate = -.243, $SE = .085$, $p = .004$) comparing the Feedback group to the other groups. This indicates that the groups began at similar levels of accuracy, and that participants in the Feedback group learned more rapidly initially before beginning to converge toward the end of training, as shown in Figure 3.

For the local effects, that is, estimating how likely it is to predict the correctness of the current trial based on the correctness of the previous trial in which the same nouns, verb, or adjectives (treatment-coded: 0 = incorrect, 1 = correct) occurred, there were main effects of previous verb accuracy (estimate = .344, $SE = .069$, $p < .001$) and the previous adjective in

the second noun position (estimate = .156, SE = .053, $p = .003$). This indicated that these local contextual effects were not equally distributed across all words in the utterance, but rather knowledge only of the verb and the second adjective (which occurred adjacent to the verb) affected performance in the current trial. These local effects interacted significantly with Block, with previous verb accuracy interacting with Block in both a linear (estimate = .554, SE = .163, $p = .001$) and quadratic (estimate = -.455, SE = .153, $p = .003$) effect, and previous second adjective accuracy interacting with Block as a linear effect (estimate = .511, $p = .180$, $p = .005$). These interactions indicated that previous verb accuracy had a larger effect in the middle stages of training, and that previous second adjective accuracy increased in its predictiveness during training.

There was no significant interaction between global and local learning effects. The full model is reported in Appendix 1A.

Performance across test blocks

Accuracy for acquisition of nouns, verbs, adjectives, marker words, and word order (syntax) for each of the four test blocks is shown in Figure 2, and Tables 4 and 5.

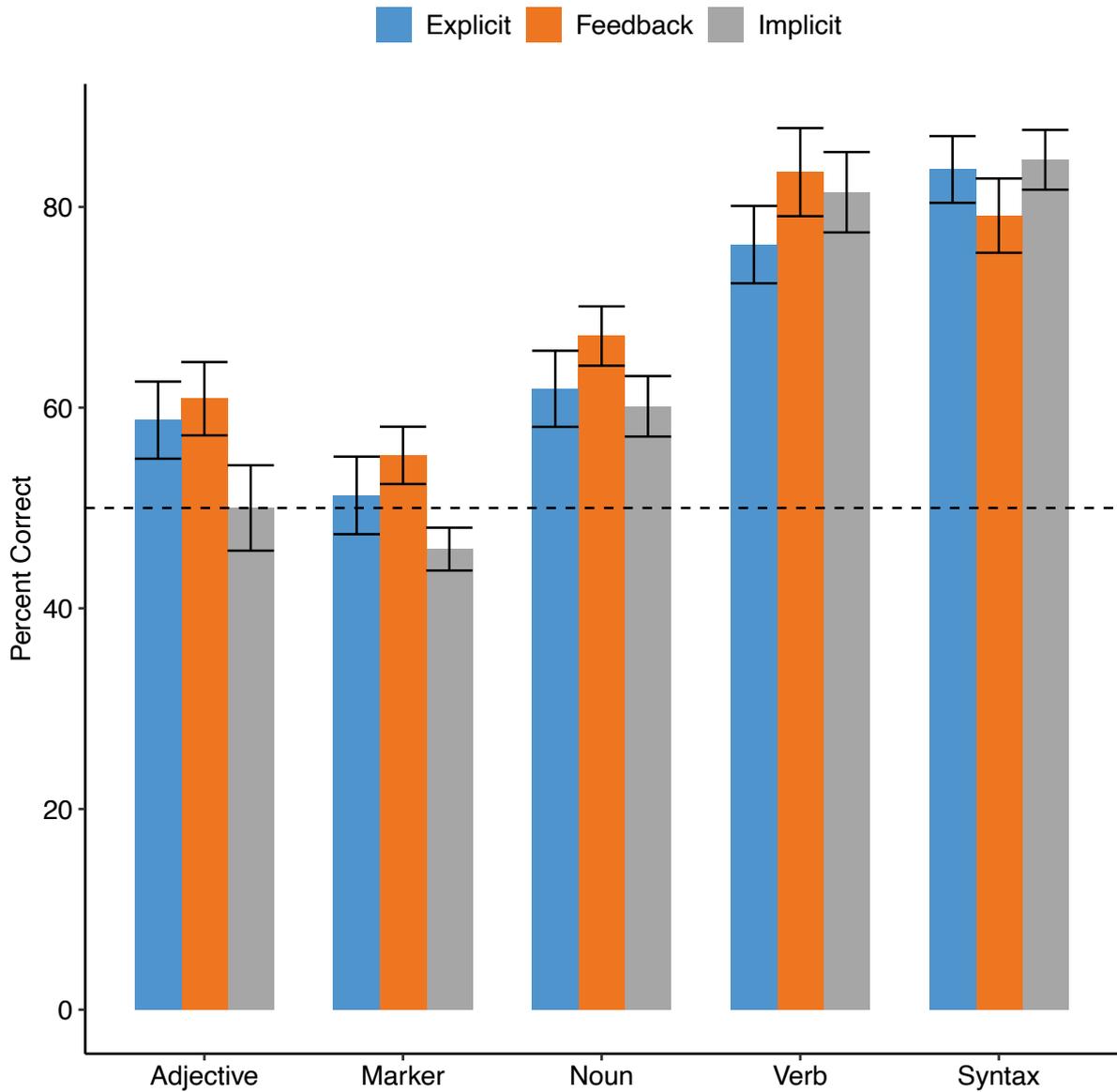


Figure 2. Performance for testing blocks. Error bars show standard error of the mean.

Table 4. Descriptive statistics for performance on adjective, marker word, noun and verb tests

Test		Implicit ^a	Feedback ^b	Explicit ^c
Adjectives	<i>M</i>	50.00	60.89	58.75
	<i>SD</i>	22.90	20.35	21.06

	<i>SE</i>	4.25	3.65	3.84
Marker words	<i>M</i>	45.91	55.24	51.25
	<i>SD</i>	11.49	15.90	21.17
	<i>SE</i>	2.13	2.86	3.87
Nouns	<i>M</i>	60.13	67.14	61.88
	<i>SD</i>	16.23	16.45	20.78
	<i>SE</i>	3.01	2.96	3.79
Verbs	<i>M</i>	81.47	83.47	76.25
	<i>SD</i>	21.55	24.45	21.11
	<i>SE</i>	4.00	4.39	3.85

^a n = 29. ^b n = 31. ^c n = 30.

Table 5. Descriptive statistics for performance on word order tests

Test		Implicit ^a		Feedback ^b		Explicit ^c	
Word Order		<i>d'</i>		<i>d'</i>		<i>d'</i>	
	<i>M</i>	84.70	2.44	79.13	2.05	83.72	2.41
	<i>SD</i>	16.03	1.30	20.62	1.55	18.21	1.41
	<i>SE</i>	2.98	0.24	3.70	0.28	3.32	0.26

^a n = 29. ^b n = 31. ^c n = 30.

In order to test global effects (i.e., group differences) on the learning of adjectives, nouns, verbs, and marker words, a mixed-effect logistic model was applied to the testing data. We first of all determined whether performance changed from the first to the second set for the vocabulary testing. There was no significant effect of testing block (estimate = .008, SE = .077, $p = .923$). We next determined whether words from different grammatical

categories There was a significant main effect of linguistic feature, with poorer performance on the adjectives than the combination of other linguistic features (estimate = $-.154$, SE = $.094$, $p < .001$), and poorer performance on the marker words than the nouns and verbs (estimate = $-.460$, SE = $.050$, $p < .001$), and nouns were also responded to with lower accuracy than verbs (estimate = $-.668$, SE = $.117$, $p < .001$). There was also a significant effect of the global variables, with the Feedback group performing significantly better than the combined Explicit and Implicit groups (estimate = $.144$, SE = $.063$, $p = .023$). The Explicit and Implicit groups did not differ from one another (estimate = $.020$, SE = $.109$, $p = .853$). There was no significant interaction between linguistic feature and Group, demonstrating no evidence that feedback or instruction affected overall accuracy of learning of grammatical categories in different ways.

In the case of word order, there was not a significant main effect of Group, indicating that the learning condition did not influence performance, $\chi^2(2) = 0.71$, $p = .497$).

Discussion

First language acquisition proceeds initially with little explicit information about the language to be learned – either in vocabulary or in grammatical structure – and with only occasional opportunities for receiving direct, explicit, and unambiguous feedback about decisions about the referents for words within an utterance (Yu & Ballard, 2007; Yu et al., 2009). In this study, we trained participants to acquire a complex, artificial language from co-occurrences between utterances and scenes containing objects, properties of objects, actions, and a distinction between subject and object roles of those objects. We replicated previous studies demonstrating that participants can resolve this difficult task (Rebuschat et al., under review); learning the meaning of nouns, adjectives, verbs, and grammatical role function words by tracking cross-situational statistics between words and varying properties of the

scenes that they viewed. Furthermore, participants could simultaneously also acquire the grammatical structure of the language – they showed learning of the syntax in terms of sensitivity to word order regularities of words in the speech.

The power of participants' learning shows that the conditions under which children acquire language – where both vocabulary and grammar are uncertain – are not an impossible impediment to language acquisition at least for the simplified language we utilised here, and additional informational cues or structural biases within the learner are not necessary for learning to proceed (Baldwin, 1993; Gleitman, 1990; Gleitman et al, 2005; Smith & Yu, 2008; Yurovsky et al., 2013). Indeed, learning is extremely rapid under these conditions: after just a few dozen exposures to varying utterances appearing with scenes to which they refer, participants are better than chance at knowing the meaning of the words and identifying the syntax within those utterances. Hence, the chicken-and-egg problem of vocabulary and grammar acquisition is solvable as a consequence of sensitivity to cross-situational statistics (see also Abend, Kwiatkowski, Smith, Goldwater, & Steedman, 2017, for a computational demonstration of this ability).

Though the non-instructed, implicit condition demonstrates that this complex artificial language can be acquired inductively, with no explicit instruction or feedback, we predicted that both feedback and explicit information about the language structure ought to support learners further in developing understanding of the language. This expectation was partially supported by the results.

In terms of feedback, we found that informing participants about whether they had selected the correct scene improved learning during training: there was a steeper trajectory of learning in the feedback condition than the explicit and implicit learning conditions, particularly in the intermediary stages of learning, with this minimal provision of feedback consistent with effects of more explicit feedback about language structure found in studies of

second language learning (Lightbown & Spada, 1990; Nakata, 2015; Nassaji, 2016). This benefit of feedback for learning also influenced performance during testing – for vocabulary learning, participants in the feedback condition scored higher overall than participants in the other conditions.

In terms of explicit instruction, we found no improvement in learning or testing performance as compared to the implicit condition. This was surprising, given that explicit instruction about grammatical structure tends to improve performance compared to conditions where no advance information about the grammar is provided (Goo et al., 2015; Norris & Ortega, 2000; Spada & Tomita, 2010). This may be a consequence of the inductive nature of the task (Andringa & Curcic, 2015; Cintrón-Valentín & Ellis, N., 2015), where explicit knowledge of language structure emerges only gradually during learning (Monaghan et al., 2019). Participants had to learn associations between individual words in an utterance with relatively free word order, and different semantic features of a complex scene depicting a transitive action. Information about the general grammatical structure could help somewhat in determining the syntax of the language (which, as we discuss below, elicits an improvement in detecting that structure) but this does not seem to transfer to supporting learning of the precise associations between words and features of the scene.

This honing in on precisely how learning proceeds, trial by trial, highlights one advantage of the laboratory-based language learning paradigm over more formal classroom-based language learning studies where detailed measurement may not be possible (see also De Graaff, 1997; DeKeyser, 1995). Such an approach thus provides theoretical insight into what aspects of learning are promoted by instruction and feedback, but also raises practical implications for interventions in supporting language learners' progress in acquiring languages. We can, for instance, ask whether the learning benefit is focused on direct improvement of the grammatical structure, which is given by the explicit instruction, or

whether the learning benefit spreads also to vocabulary (and if so, whether this applies to words within only some grammatical categories). In short, we were able to determine, for this learning task, which aspects of language are penetrated by feedback and by instruction about syntax. This assumes, of course, that patterns observed from artificial language learning apply to the language classroom, and expanding the complexity and duration of training from the artificial language to natural language situations would be a necessary step before pedagogical implications are substantiated.

However, curiously, the general improvement in training and testing performance for the feedback condition was not found in terms of accuracy on the syntax testing. We had predicted that the explicit condition would result in better performance for the word order test, but this did not explain the effects we observed. The implicit (no instruction) and the explicit condition (where instruction about the word order was given) did not differ. Thus, feedback seemed to support vocabulary acquisition but did not affect the acquisition of word order. Thus, whereas different instructional and learning conditions may improve learning, the improvement may be focused on one property of the language – with a potential dissociation between acquisition of grammar and acquisition of vocabulary, as predicted by models of learning that distinguish cognitive processing systems serving vocabulary and grammar acquisition (Monaghan et al., 2019; Paradis, 2009; Ullman, 2004). These results are also consistent with theories of cross-situational learning that suggest the benefit to learning should be around the vocabulary – where propose-but-verify (Trueswell et al., 2013) or establishing associations between words and potential referents (McMurray et al., 2012), are both strengthened – whereas each of these theories does not make predictions about improvement in learning of grammatical structure in which these words occur. The observation that feedback improvement is limited to vocabulary and not grammar is therefore consistent with these theories. Another possible contributor to this distinction between

instructional effects on vocabulary and grammar learning is due to the differences in methods of testing. Whereas the vocabulary was tested using the same type of task that was used for training (i.e., deciding which one of two scenes an artificial language sentence was referring to), the acquisition of word order was tested by a different task, which required participants to judge whether a heard sentence corresponded to the language they had heard (grammaticality judgments). This distinction between implicit and explicit responses to the structure of the language may have contributed to the different effects of instruction on acquisition (Christiansen, 2019), with implicit learning better tested by processing-based, implicit testing, and explicit learning better tested by tests involving declarative knowledge.

Contemporary statistical methods enable further detail to be revealed about how learners acquire both language vocabulary and structure (Cunnings, 2012). However, the advantage of these approaches has previously covered their value in identifying individual differences between different language learners (Linck, 2016). We show in our analyses that a further advantage of these mixed-effects models is that the precise context of learning can be taken into account during the dynamic trajectory of learning that participants experience. The statistical methods we employed enabled us to investigate the local context of learning, as well as the global effects of instruction and feedback on acquisition. The effects of local context during training in our study showed that these context effects are complex, and not generic across all aspects of the language being learned. Participants who had previously responded correctly to a trial containing the same verb or the same adjective in the second position were more likely to be correct in a trial containing the same information. For instance, responding correctly to an utterance containing a particular verb predicted a more accurate response to the next utterance in which that same verb occurred, compared to a previously incorrect response to an utterance containing that verb. Similarly, when a learner had responded correctly to an utterance containing the same adjective in the second position

of the sentence (adjacent to the verb) performance was also more accurate. These local effects provide insight into the source of knowledge that transmits from one trial to another, and enables us to pinpoint which aspect of the utterance appears to be driving learning. Verbs are sentence final, and the second adjective is the penultimate word in the sentence. Thus, learning appears to proceed from the end of the utterance, consistent with studies highlighting the importance of utterance endings in language acquisition (Freudenthal, Pine, & Gobet, 2010; Jones & Rowland, 2017).

There are two theories of how language learners acquire word-referent mappings from cross-situational statistics (MacDonald, Yurovsky, & Frank, 2017; Yurovsky, Smith, & Yu, 2013), each of which describes different mechanisms for feedback to affect learning. The first theory (propose-but-verify, Trueswell, Medina, Hafri, & Gleitman, 2013) contends that language learners generate a hypothesis about a word-referent mapping, then search for confirmatory evidence of the link. If the proposed word-referent mapping is not correct, then the information in future learning situations to verify this will be weak, and the proposed mapping will be set aside by the learner. The alternative theory (associative learning theory, McMurray, Horst, & Samuelson, 2012) is that participants do not make explicit proposals about mappings, but instead gradually acquire associations between particular words and referents in the environment. Co-occurring words and referents become incrementally strengthened as a consequence of exposure, until the actual word-referent mappings eventually have the strongest links in the learner's representation of the vocabulary. Feedback about what is correct will have a different effect on learning according to each theory, but potentially with the same observable effect on learning. In the case of propose-but-verify, explicit feedback enables the verification to be instantaneous – if the proposal of the word-referent mapping is erroneous or correct then sufficient information is provided in feedback for this proposal to be dropped. On the other hand, for the associative learning theory of

cross-situational vocabulary learning, feedback could provide an additional boost to the associative strength between the word and the referent if the association is receptive to external information about correct mappings. If this is the case, learning of the mapping would be faster, though this would still be slow in comparison to the immediate learning under the propose-but-verify theory. Under both theories, then, feedback can support inductively derived cross-situational information. However, both theories are shown to be insufficient to account for the current effects of local context on learning. Identifying where in the utterance previous contextual knowledge affects performance enables us to show which aspect of the utterance is being proposed and verified, or alternatively, which words in the utterance are receiving the strongest associative learning signal. Thus, current theories of cross-situational learning, which have been constrained to simulate isolated word to referent mappings, will require supplementary attentional mechanisms to explain learning from more realistic, complex utterance-scene correspondences.

Taken together, these results show that inductively derived, associative learning between complex utterances and complex scenes can drive learning of an artificial language, and by extension, demonstrates how naturalistic experience of a second language – embedded in context – can drive acquisition of that language. The results show that acquisition of vocabulary and grammar – whilst interactive and inter-dependent, as argued by Gentner (1982) and Gleitman (1990) – are not affected in the same way by different instructions. Furthermore, the results show that feedback affects global learning, and that local context also affects learning in terms of the importance of previous verb and previous final adjective knowledge on responses to utterances containing the same words. This means that there is substantial opportunity for contingencies to be exploited during learning situations, as in computer language tutoring systems (Amaral & Meurers, 2011; Heift & Hegelheimer, 2017). For instance, a future study combining both explicit instruction and feedback could test

whether both acquisition of word order and vocabulary are supported by differential mechanisms during learning. The insights available from investigating the local effects provide us with details about what contingent information about learning should be encoded and taken to influence the future exposure to structures being learned by the participant.

References

- Abend, O., Kwiatkowski, T., Smith, N. J., Goldwater, S., & Steedman, M. (2017). Bootstrapping language acquisition. *Cognition*, *164*, 116-143.
- Amaral, L. & D. Meurers (2011). On using intelligent computer-assisted language learning in real-life foreign language teaching and learning. *ReCALL*, *23*(1), 4–24.
- Andringa, S., & Curcic, M. (2015). How explicit knowledge affects online L2 processing: Evidence from Differential Object Marking Acquisition. *Studies in Second Language Acquisition*, *37*(2), 237-268.
- Baayen, R. H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.
- Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(2), 395-418.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.
- Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843-873.
- Childers, J. B., Heard, M. E., Ring, K., Pai, A. and Sallquist, J. (2012). Children use different cues to guide noun and verb extensions. *Language Learning and Development*, *8*(3), 233–254.
- Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*(3), 468-481.
- Cintrón-Valentín, M., & Ellis, N. (2015). Exploring the interface: Explicit Focus-on-Form Instruction and Learned Attentional Biases in L2 Latin. *Studies in Second Language Acquisition*, *37*(2), 197-235.

- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah, NJ: Erlbaum.
- Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, 28(3), 369–382.
- Cunnings, I., & Finlayson, I. (2015). Mixed effects modelling and longitudinal data analysis. In L. Plonsky (Ed.), *Advancing quantitative methods in second language research* (pp. 159–181). Abingdon: Routledge.
- de Graaff, R. (1997). The Esperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, 19(2), 249-276.
- DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, 17(3), 379-410.
- Ellis, R. (1990). *Instructed second language acquisition: Learning in the classroom*. Oxford: Blackwell.
- Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, 27(2), 164-194.
- Ellis, N. C. (2015). *Implicit and explicit learning: Their dynamic interface and complexity*. Amsterdam: John Benjamins.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27(2), 141-172.
- Ellis, R. (2015). *Understanding second language acquisition, 2nd edition*. Oxford: Oxford University Press.
- 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.

- 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.
- Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj II (Ed.), *Language development*. Vol. 2 (pp. 301–334).
Language, thought and culture Hillsdale, NJ: Lawrence Erlbaum.
- Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, 1(1), 3-55.
- Gleitman, L.R., Cassidy, K., Nappa, R., Papafragou, A. & Trueswell, J.C. (2005). Hard words. *Language Learning and Development*, 1(1), 23-64.
- Godfroid, A. (2020). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. New York: Routledge.
- Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In Rebuschat, P. (Ed.), *Implicit and explicit learning of languages*, pp.443-482. Amsterdam: John Benjamins.
- Groom, C.J., & Pennebaker, J.W. (2002). Words. *Journal of Research in Personality*, 36, 615-621.
- Gries, S. Th. (2015). The most underused statistical method in corpus linguistics: Multi-level (and mixed-effects) models. *Corpora*, 10, 95–125.
- Heift, T., & Hegelheimer, V. (2017). Computer-assisted corrective feedback and language learning. In Nassaji, H., & Kartchava, E. (Eds.), *Corrective feedback in second language teaching and learning: Research, theory, applications, implications* (Vol. 66) (pp.51-65). New York: Routledge.
- Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.

- 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.
- Koehne, J., & Crocker, M.W. (2015). The interplay of cross-situational word learning and sentence-level constraints. *Cognitive Science*, *39*(5), 849-889.
- Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, *60*, 309–365. <https://doi.org/10.1111/j.1467-9922.2010.00561.x>
- Lightbown, P. M., & Spada, N. (1990). Focus-on-form and corrective feedback in communicative language teaching: Effects on second language learning. *Studies in Second Language Acquisition*, *12*(4), 429-448.
- Linck, J. A. (2016). Analyzing individual differences in second language: The benefits of mixed effects models. In G. Granena, D. O. Jackson, & Y. Yilmaz (Eds.), *Cognitive individual differences in second language processing and acquisition* (pp. 105–128). John Benjamins Publishing.
- Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(S1), 185–207.
- Loewen, S. (2012). The role of feedback. In *The Routledge handbook of second language acquisition* (pp. 24-40). New York: Routledge.
- Loewen, S., & Godfroid, A. (2019). Advancing quantitative research methods. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 98–107). Routledge.
- MacDonald, K., Yurovsky, D., & Frank, M. C. (2017). Social cues modulate the representations underlying cross-situational learning. *Cognitive Psychology*, *94*, 67-84.
- Mackey, A. (2006). Feedback, noticing and instructed second language learning. *Applied Linguistics*, *27*(3), 405-430.

- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*, 831–877.
- Miller, J. L., & Lossia, A. K. (2013). Prelinguistic infants' communicative system: Role of caregiver social feedback. *First Language*, *33*(5), 524-544.
- Mirman, D. (2017). *Growth curve analysis and visualization using R*. London: Taylor Francis.
- Monaghan, P. & Mattock, K. (2012). Integrating constraints for learning word-referent mappings. *Cognition*, *123*, 133-143.
- 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.
- Monaghan, P., Schoetensack, C., & Rebuschat, P. (2019). A single paradigm for implicit and statistical learning. *Topics in Cognitive Science*, *11*, 536-554.
- Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154-193.
- Morgan-Short, K., Steinhauer, K., Sanz, C., & Ullman, M. T. (2012). Explicit and implicit second language training differentially affect the achievement of native-like brain activation patterns. *Journal of Cognitive Neuroscience*, *24*(4), 933-947.
- Murakami, A. (2016). Modeling systematicity and individuality in nonlinear second language development: The case of English grammatical morphemes. *Language Learning*, *66*(4), 834–871.

- Nassaji, H. (2016). Interactional feedback in second language teaching and learning: A synthesis and analysis of current research. *Language Teaching Research*, 20(4), 535-562.
- Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50(3), 417-528.
- Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.
- Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research*, 31(2), 267-278.
- Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.
- R Core Team. (2019). R: A language and environment for statistical computing. Vienna, Austria. <http://www.R-project.org>
- Rebuschat, P., Monaghan, P., & Schoetensack, C. (under review). *Learning vocabulary and grammar simultaneously from cross-situational statistics*. Submitted for publication.
- Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, 33(4), 829-856.
- Roembke, T. C., & McMurray, B. (2016). Observational word learning: Beyond propose-but-verify and associative bean counting. *Journal of Memory and Language*, 87, 105-127.
- Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, 122, 163-180.
- Siskind, J.M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition*, 61, 39-61.
- Smith, L. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, 106, 1558-1568.

- Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 263-308.
- Tamis-LeMonda, C. S., Kuchirko, Y., & Song, L. (2014). Why is infant language learning facilitated by parental responsiveness? *Current Directions in Psychological Science, 23*(2), 121-126.
- Trueswell, J. C., Medina, T. N., Hafri, A., & Gleitman, L. R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology, 66*(1), 126-156.
- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition, 92*, 231–270.
- White, L. (1991). Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom. *Second Language Research, 7*(2), 133-161.
- Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford: Oxford University Press.
- Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing, 70*, 2149–2165.
- Yu, C., & Smith, L. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science, 18*, 414–420.
- Yu, C., Smith, L. B., Shen, H., Pereira, A. F., Smith, T. (2009). Active information selection: Visual attention through the hands. *IEEE Transactions on Autonomous Mental Development, 1*, 141–151.
- Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science, 16*, 959-966.

Appendix

Note that in the following report, the variables refer as follows:

ot1	linear effect of block
ot2	quadratic effect of block
CG1	comparison between feedback condition and explicit/implicit conditions
CG2	comparison between explicit and implicit conditions
prevnoun1	accuracy on previous utterance containing noun in the first position of the current utterance
prevnoun2	accuracy on previous utterance containing noun in the second position of the current utterance
prevverb	accuracy on previous utterance containing verb of the current utterance
prevadj1	accuracy on previous utterance containing adjective in the first position of the current utterance
prevadj2	accuracy on previous utterance containing adjective in the second position of the current utterance
CT1	comparison between adjective and the other grammatical categories
CT2	comparison between marker words and the other grammatical categories
CT3	comparison between nouns and the other grammatical categories

A1. Generalised linear mixed effect model results for training.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: binomial ( logit )
Formula: EnableInput.ACC ~ (ot1 + ot2) * (CG1 + CG2) + (CG1 + CG2) * (prevnoun1 +
  prevnoun2 + prevverb + prevadj1 + prevadj2) + (1 + ot1 +
  ot2 + (prevnoun1 + prevnoun2 + prevverb + prevadj1 + prevadj2) | Subject) + (1 +
  (ot1 + ot2) * (CG1 + CG2) | items)
Data: dat2
```

Control: ctrl

AIC	BIC	logLik	deviance	df.resid
15658.4	16472.8	-7724.2	15448.4	17157

Scaled residuals:

Min	1Q	Median	3Q	Max
-13.8640	-0.7166	0.1986	0.5870	2.1446

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
items	(Intercept)	0.049006	0.22137	
	ot1	0.045474	0.21325	-0.95
	ot2	0.064764	0.25449	-0.99 0.98
	CG1	0.000703	0.02651	0.85 -0.67 -0.80
	CG2	0.007257	0.08519	-0.58 0.33 0.50 -0.91
	ot1:CG1	0.009039	0.09507	0.00 -0.25 -0.08 -0.51 0.80
	ot1:CG2	0.003935	0.06273	-0.67 0.55 0.64 -0.56 0.34 -0.14
	ot2:CG1	0.007085	0.08417	-0.95 1.00 0.98 -0.68 0.34 -0.25 0.56
	ot2:CG2	0.004373	0.06613	0.61 -0.65 -0.63 0.54 -0.45 -0.20 0.16 -0.64
Subject	(Intercept)	1.237364	1.11237	
	ot1	4.857213	2.20391	0.80
	ot2	0.600897	0.77518	-0.60 -0.07
	prevnoun1	0.038002	0.19494	0.29 0.60 0.35
	prevnoun2	0.023474	0.15321	0.54 0.67 0.22 0.68
	prevverb	0.200434	0.44770	0.09 0.20 0.07 -0.08 0.20
	prevadj1	0.064279	0.25353	0.11 0.18 -0.11 -0.16 -0.33 -0.64
	prevadj2	0.014817	0.12173	0.83 0.70 -0.53 -0.10 0.25 0.06 0.48

Number of obs: 17262, groups: items, 181; Subject, 90

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.260552	0.135382	9.311	< 2e-16 ***
ot1	2.676975	0.260491	10.277	< 2e-16 ***
ot2	-0.298756	0.129124	-2.314	0.02068 *
CG1	0.192409	0.095957	2.005	0.04495 *
CG2	0.001292	0.162547	0.008	0.99366
prevnoun1	0.016451	0.049591	0.332	0.74008
prevnoun2	0.036528	0.048124	0.759	0.44783

prevverb	0.343537	0.068734	4.998	5.79e-07	***
prevadj1	0.030338	0.059401	0.511	0.60954	
prevadj2	0.155656	0.052505	2.965	0.00303	**
ot1:CG1	0.047151	0.180764	0.261	0.79421	
ot1:CG2	-0.074433	0.317322	-0.235	0.81455	
ot2:CG1	-0.242950	0.084621	-2.871	0.00409	**
ot2:CG2	-0.072568	0.143849	-0.504	0.61393	
CG1:prevnoun1	0.005616	0.036393	0.154	0.87737	
CG1:prevnoun2	-0.008396	0.035241	-0.238	0.81170	
CG1:prevverb	-0.047884	0.049653	-0.964	0.33485	
CG1:prevadj1	0.050424	0.043427	1.161	0.24559	
CG1:prevadj2	-0.046656	0.038636	-1.208	0.22721	
CG2:prevnoun1	0.034160	0.058297	0.586	0.55790	
CG2:prevnoun2	-0.054364	0.056382	-0.964	0.33495	
CG2:prevverb	-0.155134	0.082181	-1.888	0.05906	.
CG2:prevadj1	0.065896	0.070135	0.940	0.34744	
CG2:prevadj2	-0.046556	0.061588	-0.756	0.44970	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					

Fixed effects of training block x local features, from the following model:

Formula: EnableInput.ACC ~ (ot1 + ot2) * (CG1 + CG2) + (ot1 + ot2) * (prevnoun1 + prevnoun2 + prevverb + prevadj1 + prevadj2) + (1 + (ot1 + ot2) + (prevnoun1 + prevnoun2 + prevverb + prevadj1 + prevadj2) | Subject) + (1 + (ot1 + ot2) * (CG1 + CG2) | items)

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)	
ot1:prevnoun1	-0.26788	0.15562	-1.721	0.08517	.
ot1:prevnoun2	-0.29237	0.16067	-1.820	0.06881	.
ot1:prevverb	0.55400	0.16345	3.389	0.00070	***
ot1:prevadj1	-0.31289	0.18248	-1.715	0.08641	.
ot1:prevadj2	0.51077	0.18014	2.835	0.00458	**
ot2:prevnoun1	-0.07421	0.15071	-0.492	0.62246	
ot2:prevnoun2	0.05698	0.15450	0.369	0.71230	
ot2:prevverb	-0.45482	0.15348	-2.963	0.00304	**
ot2:prevadj1	0.05162	0.17514	0.295	0.76820	

```

ot2:prevadj2  0.10236  0.17469  0.586  0.55789
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

A2. Generalised linear mixed effect model results for testing vocabulary.

```

Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
Family: binomial (logit)
Formula: EnableInput.ACC ~ (CT1 + CT2 + CT3) * (CG1 + CG2) + (CT1 + CT2 +
      CT3) * TestingBlockCat + (CG1 + CG2) * TestingBlockCat + (1 + (CT1 + CT2 + CT3) *
TestingBlockCat | Subject)
Data: dat
Control: ctrl

```

AIC	BIC	logLik	deviance	df.resid
5489.6	5833.7	-2690.8	5381.6	4266

Scaled residuals:

Min	1Q	Median	3Q	Max
-5.1310	-0.9574	0.3972	0.8254	1.4494

Random effects:

Groups	Name	Variance	Std.Dev.	Corr
Subject	(Intercept)	0.499292	0.70661	
	CT1	0.035064	0.18725	-0.05
	CT2	0.081572	0.28561	-0.63 0.24
	CT3	0.484879	0.69633	-0.50 0.62 0.72
	TestingBlockCat	0.006892	0.08302	-0.90 -0.20 0.47 0.10
	CT1:TestingBlockCat	0.038148	0.19531	-0.52 -0.68 0.06 -0.02 0.50
	CT2:TestingBlockCat	0.017034	0.13052	-0.52 -0.33 -0.23 -0.02 0.46 0.82
	CT3:TestingBlockCat	0.111888	0.33450	0.04 -0.74 -0.63 -0.66 0.16 0.68 0.73

Number of obs: 4320, groups: Subject, 90

Fixed effects:

Estimate	Std. Error	z value	Pr(> z)
----------	------------	---------	----------

(Intercept)	0.725613	0.093826	7.734	1.05e-14	***
CT1	-0.154406	0.037946	-4.069	4.72e-05	***
CT2	-0.460114	0.049881	-9.224	< 2e-16	***
CT3	-0.668383	0.117336	-5.696	1.22e-08	***
CG1	0.143852	0.063091	2.280	0.0226	*
CG2	0.020169	0.108877	0.185	0.8530	
TestingBlockCat	0.007511	0.077226	0.097	0.9225	
CT1:CG1	-0.002521	0.019174	-0.131	0.8954	
CT1:CG2	0.050134	0.033110	1.514	0.1300	
CT2:CG1	-0.012804	0.028735	-0.446	0.6559	
CT2:CG2	0.053220	0.049326	1.079	0.2806	
CT3:CG1	-0.029668	0.064862	-0.457	0.6474	
CT3:CG2	0.110540	0.109363	1.011	0.3121	
CT1:TestingBlockCat	0.022533	0.049309	0.457	0.6477	
CT2:TestingBlockCat	0.084293	0.055688	1.514	0.1301	
CT3:TestingBlockCat	-0.040345	0.126186	-0.320	0.7492	
CG1:TestingBlockCat	-0.027659	0.047723	-0.580	0.5622	
CG2:TestingBlockCat	0.082705	0.082149	1.007	0.3140	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A3. Linear regression model results for grammatical word order test.

```
lm(formula = dprime ~ CG1 + CG2, data = dat_dprime)
```