Manuscript Running Header: Language and Environmental Constraints for Learning

Article Type: EMPIRICAL STUDY

Manuscript Title: Redundancy And Complementarity In Language And The Environment: How Intermodal Information Is Combined To Constrain Learning

Authors: Padraic Monaghan[a], Heather Murray[a], Heiko Holz[b,c]

Author Affiliations: [a]Lancaster University [b]Ludwigsburg University of Education [c]University of Tuebingen

Author notes / acknowledgements

CRediT author statement – Padraic Monaghan: conceptualization (lead); methodology (lead); supervision (lead); writing – review and editing (lead). Heather Murray: investigation (lead); writing – original draft preparation (lead). Heiko Holz: software (lead); methodology (supporting); writing –review and editing (supporting).
A one-page Accessible Summary of this article in non-technical language is freely available in the Supporting Information online and at https://oasis-database.org
Correspondence concerning this article should be addressed to Padraic Monaghan, Department of Psychology, Lancaster University, Lancaster LA1 4YF, UK, p.monaghan@lancaster.ac.uk.

Abstract

To acquire language, learners have to map the language onto the environment, but languages vary in terms of how much information is present within the language to constrain how the sentence relates to the world. We investigated the conditions under which information within the language and the environment is combined for learning. In a cross-situational artificial language learning study, participants listened to transitive sentences and viewed two scenes, and selected which scene was described by the sentence. The language had free word order, and varied in terms of whether or not it contained morphosyntactic information in order to define the subject and object roles of nouns in the sentence. We found that participants were able to learn information about word order and vocabulary from each language, demonstrating that information within the language only was not necessary for learning. Instead, participants can combine constraints from language and environment to support acquisition.

Data and analyses are available at:
https://osf.io/hxqzc/?view_only=ea6ba6fff6bb468e8de2e8596f029dca

Keywords: vocabulary learning, morphology, language acquisition, language variation, cross-situational learning.

Introduction

Information in language structure and environmental structure

Language learning, under natural ecological conditions, occurs in context, such that speech is heard at the same time as events unfold around the learner. However, determining which sounds within the speech refer to precisely what in the world is a substantial challenge. This complexity is due to there being multiple words within the utterance, and an infinite number of possible referents within the world for each of those words (Quine 1960). For instance, in an analysis of a corpus of child-directed speech, Yu and Ballard (2007) investigated co-occurrences of words spoken by a caregiver with actions and objects within the child's view, and found that matching word-referent pairs comprised between 2.4 and 2.7% of the possible pairings in the child's environment, with the vast majority being pairings between words and other referents in the environment in addition to the target. That is, more than 97% of possible pairings between words and referents in the child's environment need to be disregarded by the learner, and only the small minority of correct co-occurrences are to be acquired.

One key means by which the language learner begins to get a grasp on pairing words with meanings is to track the cross-situational correspondences between words in the utterance and the objects, events, and relations that occur around them in the environment (McMurray et al., 2012; Siskind, 1996; Yu et al., 2021; Yurovsky et al., 2013). Cross-situational correspondences describe co-occurrences between particular words and particular features of the environment that are referred to by those words. From a single communicative situation, there may be several words and a huge panoply of possibilities for referents to each word that the language learner experiences. However, over multiple situations, learners may observe that a particular environmental feature is present whenever a certain word is heard (Yu & Ballard, 2007). Experimental studies with adults and infants have shown that both noun-object mappings (Yu & Smith, 2007; Smith & Yu, 2008; Smith et al., 2011) and intransitive verb-action mappings (Childers et al., 2012; Scott & Fisher, 2012) can be acquired from these cross-situational statistics. Monaghan et al. (2015) extended these studies to demonstrate that both nouns and intransitive verbs could be learned simultaneously from cross-situational statistics, demonstrating how both these grammatical categories could be mapped from a single utterance to various objects and actions in the environment.

Learning individual word-referent mappings, as Gleitman (1990) noted, is an instance of a difficult problem, but language learners tend to experience substantially more complex situations than those tested in laboratory settings. Learners tend not to hear just words or very short phrases and tend not to see constrained environments containing small sets of selected objects or actions. Instead, learners hear sentences and observe varied scenes, where nouns, verbs, and (for some languages) morphosyntactic features have to be mapped onto objects, actions, and relations among those objects and actions. Yet, in order to fully understand the role of a word in a sentence, its meaning has to be known, but in order to understand the meaning of a word, its role in a sentence must also be known. How does the language learner solve this paradox?

Artificial language learning explorations of acquisition of vocabulary and syntactic structure have tended to solve this “chicken-and-egg” problem for learners by presenting the tasks in sequence. For example, Friederici et al. (2002) provided training in the meaning of words, before they were embedded in multi-word utterances. However, an alternative perspective is that there are multiple cues within the language that can support learning of grammatical structure as well as vocabulary, and these cues may mitigate the need for prior, explicit learning of meanings of individual words.

Morphology and word order across languages

One source of information to indicate the role of words in the sentence is morphological case-marking of nouns, which can indicate which is the subject and which the object, and thus constraining the nouns' relation to the verb. So, developing knowledge about the correspondence between these morphological markers and grammatical roles can provide an avenue toward learning words' meanings. Alternatively, languages may provide this information in terms of word order (e.g., that the subject, object, and verb occur in reliable orderings within the sentence) which can also support derivation of this information. For example, in English, which does not mark subject and object case for nouns (except pronouns), the subject noun tends to occur before the verb and the object noun tends to occur immediately after the verb. Even for languages that permit freer word order, there is often a tendency to conventionalise the word order, as in the case of Japanese where both subject-object-verb (SOV) and subject-verb-object (SVO) word orders are permitted, but SOV is dominant (Shibatani, 1990).

The Competition Model (MacWhinney & Bates, 1989) contends that information about grammatical structure is carried either in word order or in morphology, and certainly occurrence of either case marking or fixed word order in a language has been shown to support learning of that language (Hawkins, 2004; Lupyan & Christiansen, 2002; van Everbroeck, 1999). Table 1 reports the results of an analysis of a wide set of languages in the World Atlas of Languages (Dryer &

Haspelmath, 2013) in terms of whether they use word order to indicate subject, object, and verb, and/or whether they use morphological case marking. Of the 239 languages with information on both word order and case marking, there is some support for the Competition Model, in that the majority of languages without case marking had a dominant word order, and the majority of languages without a dominant word order had case marking. Koplenig et al. (2017) found a similar trade-off in distribution between occurrence of case marking and fixed word order across 1,200 languages, though Levshina (2021), in a causal analysis of 30 languages, found that sociolinguistic influences between languages, rather than information processing constraints, were predominant in explaining the trade-off of morphology and word order constraints.

However, as shown in Table 1, there remain 11 languages of the 239 classified that have neither dominant word order nor morphological case marking (Dryer, 2013). In addition to the difficulty of acquiring language without prior knowledge of either vocabulary meaning or grammatical roles, how is language learning possible when subject and object roles are underspecified within the language? In other words, how could learners begin to relate words to their referents without information about the role of the word in the sentence?

Table 1. Analysis of number of languages by word order and case marking properties in the World Atlas of Languages (Dryer, 2013; Iggersen, 2013).

	Case marking	Borderline case marking	No case marking
Dominant word order	99	17	86
No dominant word order	21	5	11

Note: “Borderline case marking” is when the language contains case marking but not of syntactic roles such as subject and object (Iggersen, 2013).

Combining environmental and linguistic cues

One possibility for resolving this problem of learning languages (with underspecified information within the language about subject and object roles of words) is that information from the environment about the subject and object roles of *referents* could contribute to constraining interpretation of the language. In a cross-situational experimental study, Rebuschat et al. (2021) tested whether participants were able to keep track of correspondences among multiple words in sentences of an artificial language and objects, properties of those objects, and actions illustrated in a scene. The language comprised transitive sentences with word order varying between SOV and OSV. The nouns were marked for subject or object case, and adjectives occurred optionally before the nouns. In their Experiments 2 and 3, each utterance occurred with two scenes, one of which corresponded to the referents for the nouns, case markers, adjectives, and verb in the sentence, and the other which contained objects, actions, and features of objects that did not correspond to the sentence. The study tested whether, over multiple learning trials without feedback, cross-situational learning was a sufficiently powerful mechanism for the learners to be able to keep track of the multiple possible mappings between words and referents in the scenes. Rebuschat et al. (2021) found that participants could successfully learn the language, acquiring information about word order, as well as the meanings of nouns, verbs, adjectives, and the roles of the subject and object case-markers.

Rebuschat et al. (2021) thus demonstrated that cross-situational statistical learning is a powerful source of information to aid learners to acquire grammatical structure and meanings of words from multiple grammatical categories. Yet, the combination of sentences and scenes in Rebuschat et al.'s (2021) study provided more information than was strictly needed in order to highlight subject and object roles of nouns in the sentence. There was overlapping information present in terms of subject and object roles indicated in the scene alongside morphosyntactic markers indicating subject and object nouns in the sentence. Such overlapping information in language is not rare (Dale & Lupyan, 2012). Indeed, as noted above in Table 1, 99 languages in the analysis of 239 languages in the World Atlas of Languages have both case marking and a dominant word order indicating subject and object roles in the sentence. However, the effect of this overlapping information in the language and the environment has not, to our knowledge, been previously examined.

Previous studies have, however, investigated the effect of overlapping information *within* the language. Tal and Arnon (2022) conducted a study to determine whether overlapping information present in the language (referred to as redundancy) can confer an advantage for learning subject and object roles in an artificial language. Hebrew-speaking children and adults were exposed to a language with Hebrew nouns and verbs, but with a different word order than Hebrew. Participants were assigned to one of two conditions, one where subject and object roles were indicated with morphological markers, and one where only word order provided this information. Participants heard sentences and saw scenes relating to those sentences. Tal and Arnon (2022) found that children learned the redundant language better than the non-redundant language, whereas adults performed at ceiling in both conditions. Thus, they found that redundancy can promote learning of grammatical structure, at least for children. But what of the situation when grammatical information is specified neither within word order nor morphology? It remains an understudied question as to how learning can proceed when grammatical structure is underspecified in the language.

Grounding language in the environment – that is, accompanying sentences with scenes to which those sentences refer – means that a far broader array of information is available for supporting language learning beyond the co-occurrences of words and morphemes within the language itself. In the case of a transitive sentence, such as “the penguin chases the monkey”, when the learner also observes a scene where penguin is observed as the chaser (i.e., the subject role), and monkey is observed being the chased (i.e., the object role), this provides valuable semantic information about the subject and object of the sentence, in turn constraining the interpretation of the verb “chases”[1]. If the learner has some knowledge of the referent of penguin, or monkey, then bootstrapping of the grammar of the language and, simultaneously, the meaning of the verbs as well as other words in the sentence can commence. This combination of information about roles from the environment with the language means that learning, even from a language that is under-specified with respect to subject and object roles (as in those natural languages without dominant word order and without morphological case marking) is feasible.

Furthermore, there may be advantages to minimising the amount of overlapping information available for learning. According to the Competition Model (MacWhinney & Bates, 1989), providing overlapping, redundant information increases the burden on the learner by adding to the quantity of information that is present within the language. For instance highlighting subject and object roles with both word order and morphosyntax, as in Tal and Arnon’s (2022) study, means that the language is more internally complex than a language which indicates subject and object roles only with morphosyntax or with word order. So, if there is a benefit of overlapping information for learning, then this is likely a trade-off with the increased quantity of information that the learner has to process (Gold, 1967).

The current study

In this study, our first research question was how can languages that are underspecified with regard to subject and object roles within the language still be learned? We predicted that a combination of environmental and linguistic constraints would be sufficient for learning, even if linguistic constraints alone were insufficient. Our second research question asked whether the overlapping information carried between sentence and scene in terms of determining subject and object roles can be exploited to support language learning. Based on studies of the benefit of redundancy (e.g., Tal & Arnon, 2022), we predicted that overlapping information between environment and linguistic features may result in learning that is advantageous over a language that requires information about subject and object roles to be supported by environmental information, where the information is complementary rather than redundant across the modalities. This boost to learning may be observed in acquisition of the nouns and verbs around which the marker words provide information about the respective roles of the two nouns. Thus, any benefit of information overlap could be observed in overall learning of the language – either in terms of acquiring the referents for the vocabulary or in terms of supporting learning of the word order constraints. Alternatively, the additional complexity of including morphosyntactic information about subject and object roles of the words may affect the ability of participants to acquire the word order and the multiple co-occurrences among words and potential referents in the environment, though Tal and Arnon's (2022) study suggests that redundancy may proffer an advantage for learning.

We conducted a cross-situational language learning task, where sentences and scenes co-occurred, but we varied the extent to which information in the language about subject and object nouns was present. Our study took as its starting point the language and environment used in Rebuschat et al. (2021), Experiment 2, where a sentence occurred along with two scenes, one of which was referred to by the sentence.

In the first condition (informative marker condition), we included morphological markers to subject and object nouns in the sentence, as in Rebuschat et al. (2021). In this version of the language, there was sufficient information within the language itself to constrain the roles of words in the sentence. We anticipated that, as in Rebuschat et al. (2021), the word order and the referents (both actions and objects) for individual words in the language would be learnable, as well as the meaning of the subject and object markers themselves, though in Rebuschat et al. (2021) this learning effect was very small.

In a second condition (no marker condition), we omitted the morphological markers, such that the language was now under-specified and required complementary information from the scene about subject and object roles to be processed to determine the meaning of the actions and objects in the scene. Thus, there was no linguistic information about the subject or object roles of nouns in the sentence. If, in line with our prediction from the first research question, this partial information from language and scene can be combined effectively, then we should see evidence of learning, thus gaining evidence for how an under-specified language lacking both word order and morphological case marking information can be acquired (i.e., via co-occurrences among words in sentences and objects, actions and roles in the environment). Further, to test research question 2, if learning benefits from a fully-specified language then we should see learning at lower levels for this no marker condition than the informative marker word condition.

We also included a third condition (uninformative marker condition), where the language contained morphological markers, but these were ambiguous with respect to the subject and object nouns in the sentence. Hence, this language was underspecified but also contained (uninformative) marker words that the learner would have to learn to ignore. This condition was included to control for the complexity of the language of the informative marker condition, because the no marker condition resulted in shorter sentences than the informative marker condition, and so differences between the no marker and informative marker conditions could be mitigated, or exaggerated, by differences in the sentence length[2]. As for the no marker condition, we anticipated that if learning could proceed by combining partial information from the language and the environment for this uninformative marker condition, then we should see performance greater than chance in terms of accuracy. If participants had difficulty learning to ignore the uninformative marker words then we should see performance lower than the no marker condition.

Pre-registration of the analyses, data, and code for analyses are available on the Open Science Framework (OSF): https://osf.io/hxqzc/?view_only=ea6ba6fff6bb468e8de2e8596f029dca

Method

Participants

There were 60 participants in this study, recruited through opportunity sampling via social media. Twenty participants were each randomly assigned to the three conditions, with this sample size for each group the same as previous studies using a similar paradigm to the marker word condition (Rebuschat et al., 2021)[3]. Twenty-eight participants identified as female and 32 identified as male, and their ages ranged from 19 to 68 (*M* = 24.2 years, *SD* = 9.7)[4]. Five of the participants had knowledge of a language that uses morphology to identify the subject and object of the sentence (one each speaking Albanian, Greek, both Kurdish and Turkish, Malayalam, and Russian). Three of these participants were in the informative marker condition, and one participant was in each of the other conditions. The other participants were monolingual English speakers[5].

Ethical approval for this study was received from the Department of Psychology Research Ethics Committee at Lancaster University. Participants were informed of their right to withdraw and gave consent after receiving information about the general characteristics of the task they were to conduct. Participants automatically received a debrief sheet in their email one hour after completing the questionnaire.

Materials

Eight colourful cartoon animals were used in the study. These were adapted from https://freepik.com and https://vhv.rs by adding eyes when these did not appear and editing the background on which they appeared. Two of these (cat and warthog) were used in the practice trials, and the other six occurred in the main experiment (chicken, cow, elephant, owl, tortoise, and zebra). The animals were selected so that they were visually distinct from one another in terms of shape and colour.

The cartoon animals were animated to perform 6 different actions. Two of these were used in the practice trials (circling: where one animal moves around the other; chasing: where one animal moves toward the other then the other animal moves away with the first animal continuing to move towards the other). The other four actions were retained for the main part of the experiment, and involved: hiding, where one animal moved behind the other; jumping, where one animal moved in an arc over the other and stopped on the other side of the animal; bumping, where one animal approaches the other, touches it, and then the other animal moves rapidly away; and lifting, where one animal moves underneath the other, then both animals move upwards. The animals and actions occurred in front of a colourful outdoor scene, again taken from https://freepik.com. Figure 1 shows a still frame from an example of a learning trial.

Table 2. Example sentences in the three language conditions.

Condition	Example sentence
No Marker	bahget gorshal dingep zebra cow jumps "cow jumps zebra" *or* "zebra jumps cow"
Informative Marker	melbai noo bimdah tha racken owl SUBJECT turtle OBJECT lifts "owl lifts turtle"
Uninformative Marker	gorshal noo limebah tha mackot cow SUBJECT/OBJECT hen SUBJECT/OBJECT bumps "hen bumps cow" *or* "cow bumps hen"

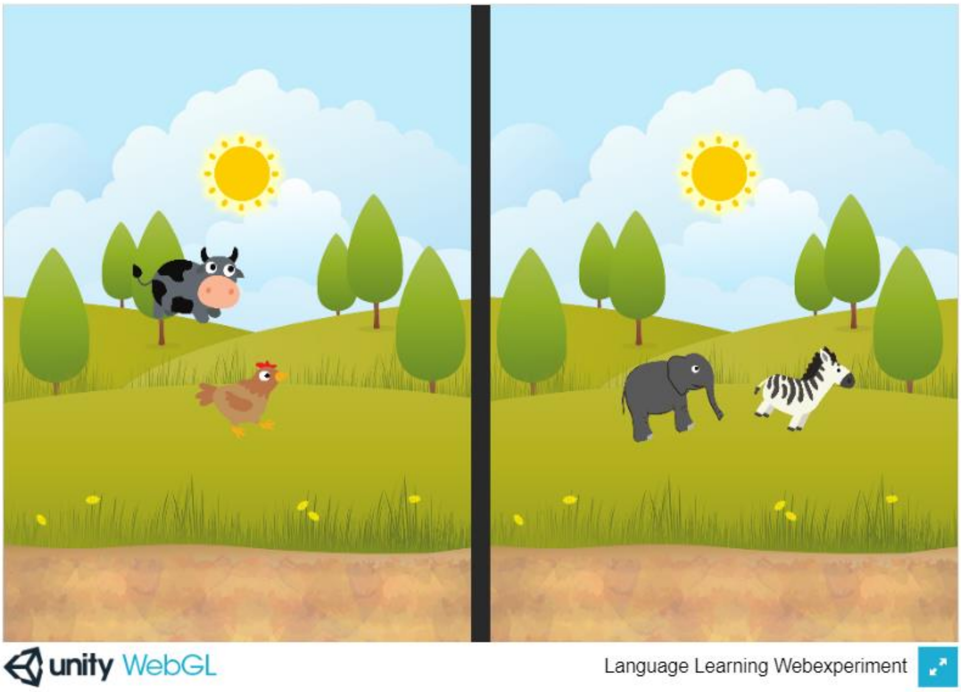

Figure 1. Still from an example learning trial. Participant hears a sentence in the artificial language and sees two scenes, in this case the left depicts the cow jumping the hen, and the right scene depicts the elephant bumping the zebra.

The language comprised 12 words, taken from Monaghan and Mattock (2012). There were six bisyllabic nouns that corresponded to the six animals (“bahget”, “bimdah”, “chilad”, “gorshal”, “limebah”, and “melbai”), four bisyllabic verbs (“dingep”, “mackot”, “racken”, and “thislin”), and two monosyllabic subject/object marker words (“noo” and “tha”). Words were spoken by a female native English speaker, and were recorded individually in a monotone.

The grammar of the language was verb-final with subject and object free to vary. Three different versions of the language were used, one for each condition of the experiment. For the no marker condition, there were no marker words, and so whether the first noun was the subject or object of the sentence was ambiguous. For the informative marker condition, the marker words occurred immediately after the noun, and one marker word was always assigned to indicate subject, and one marker word reliably indicated object of the sentence. For the uninformative marker condition, the marker words occurred after the nouns, but there was no reliable correspondence as to whether each marker word indicated subject or object. Table 2 provides an example sentence for each of the three conditions, indicating that for the no marker and uninformative marker conditions two meanings are consistent with the sentence, but just one meaning is consistent in the informative marker condition.

The items used for testing what participants had learned about the word order constraints of the language comprised 12 grammatical and 12 ungrammatical sentences. The grammatical sentences were constructed in the same way as for the training trials, with nouns and verbs occurring an equal number of times across the sentences. The ungrammatical sentences had a different word order, with 3 each having the order VSO, VOS, SVO, and OVS. Depending on the condition, marker words occurred in these sentences alongside the nouns, and so this word order constraints test measured knowledge of the relative word order of the verb and noun phrases. Again, nouns and verbs occurred an equal number of times within the ungrammatical sentences.

Procedure

Participants first visited a webpage hosted on Qualtrics, an online questionnaire platform. Participants were provided with information about the task, indicated their consent, and gave background information on which language(s) they spoke. The information provided about the task was that participants would learn a new language, where they would hear a sentence and have to select from two scenes which they felt the sentence referred to. Qualtrics then automatically referred participants to a webpage where the experimental task was delivered, programmed as a WebGL application using Unity 3D (version 2019.1.0f2, https://unity3d.com/) where the animations were implemented. The experiment was thus delivered entirely online.

The experiment commenced with participants indicating their gender, age, and handedness, and then an audio check was conducted by participants clicking on a series of four animals in the order that they were named using their English names (the animals were the same as those that later appeared in the main experiment).

Then the experiment moved on to the practice trials. The instructions about selecting the left or right scene corresponding to the sentence were reiterated. Then there were four practice trials. In each, participants saw two scenes, each comprising a cat and a warthog but interacting in different ways: with one either chasing or circling the other. After one completion of the action, participants heard the sentence, whilst the action continued. After the sentence completed, a sun appeared in the upper centre of each scene and participants clicked on the sun to indicate their scene selection.

After the four practice trials, the instructions were again reiterated, and participants pressed a button indicating Continue to move on to the training phase. For the training phase, participants saw two scenes, each containing two different animals and two different actions that had not been seen during the practice trials. After the action had completed, the sentence was heard whilst the action repeated. After the sentence was heard, the two suns appeared, and the actions repeated until the participant made a selection. After this selection, the next trial began immediately.

There were a total of 6 training blocks, each comprising 12 trials, with each animal occurring twice in the target scene and twice in the foil scene, and each action occurring 3 times in the target scene and 3 times in the foil scene. Half the trials in each block had SOV word order, and the other half had OSV word order. Each noun and verb occurred an equal number of times in each block of training, each noun occurred an equal number of times as subject and object of the sentence, and co-occurrences of nouns with verbs were balanced over the whole training set. Marker words were similarly balanced within the training sets. We selected 6 training blocks, as previous studies using this paradigm had indicated reliable levels of learning after this amount of exposure (Rebuschat et al., 2021).

After the training trials completed, the vocabulary test items began. These were similar to the training trials, except that the two scenes differed in only one aspect. For the noun trials, the action was the same in the two scenes and just one of the animals differed; for the verb trials, the animals were the same in the two scenes, and the actions differed; for the marker word trials, the actions and animals were the same in the two scenes, but the subject and object roles of the animals were switched between the two scenes. There were 12 noun test trials, with each noun as target two times, and there were 8 verb test trials, with each verb occurring twice as target. For the marker word condition, there were 8 marker word test trials, but these trials did not occur for the other two conditions. Half the trials had SOV and half had OSV word order in the testing trials. Further examples of training and vocabulary test trials are shown in Appendix S1.

After the vocabulary testing trials, the experiment began the word order constraints testing trials. Participants were instructed that they would hear a sentence and see one scene, and they had to indicate if the language was the same or different than the language they had already heard. After hearing the sentence, participants clicked a similar or different button, then the next trial began. There were 2 practice trials, then the instructions were reiterated, then the 24 word order test trials were presented, which balanced the occurrence of nouns and verbs. Two of the 24 word order test trials had previously been heard during training, but the majority of these test sentences were novel and so accurate performance in this test could only be achieved by generalising the participants' language knowledge to these novel sentences, rather than rote learning the sentences. However, accurate responses could be achieved by recognising when the verbs occurred word finally or not – hence the test is of the word order with respect to relative positions of noun phrases and verbs rather than the full grammatical structure of the language. Results of analyses omitting the two test sentences that had been previously heard are reported in Appendix S4, and these results are similar to those reported here.

Results

For training trials and noun, verb, and marker word test trials, accuracy of correct sentence-scene pairs was scored 1 for correct selections and 0 for incorrect choices. These values were then entered into a series of general linear mixed effects models (GLMEs; using lme4 package v.1.1-31 in R v.4.2.2) as the dependent variable, with random effects of participant and of the sentence. Note that including random effects of the sentence was important to ensure that learning was not greater for certain items than others, because the pairing between particular words and actions and words and objects was the same across all participants.

All modelling commenced with the baseline model containing only random effects, including intercepts and slopes for fixed effects. If the model failed to converge or resulted in singularity then slopes were omitted one at a time until the model converged. Models were then constructed incrementally including additional fixed effects of training block (for the training trials) which was treated as a numeric variable from 1 to 6, and of language condition (for training and test trials) which was a categorical variable. Each subsequent model was compared to the previous model using log-likelihood comparisons (Barr et al., 2013).

For the training trials, we additionally tested the fit of a quadratic effect of block, and compared this to the linear fit of block, according to our preregistration. However, the linear fit was found to be the better fit and so only linear fit of block was considered in the results reported here (see Appendix S2 for details of these analyses).

For the word order constraints test trials, to control for response bias, we analysed d-prime values for responses to grammatical over ungrammatical sentences. These d-prime values were thus produced for each participant and subsequently analysed using a linear regression model (*lm* in R)[6].

Training trials

Figure 2 shows the proportion of correct responses by participants for each of the three language conditions over the six training blocks. As explorative statistics, we first determined whether participants could learn to select the correct scene given the sentence for each condition individually over the six blocks using one-sample t-tests. The results are shown in Table 3, and suggest that the correspondence between scenes and sentences could be acquired in all conditions, with slightly quicker acquisition for the no marker condition than the other two conditions, though note that there is a drop in accuracy for the uninformative marker condition after block 3.

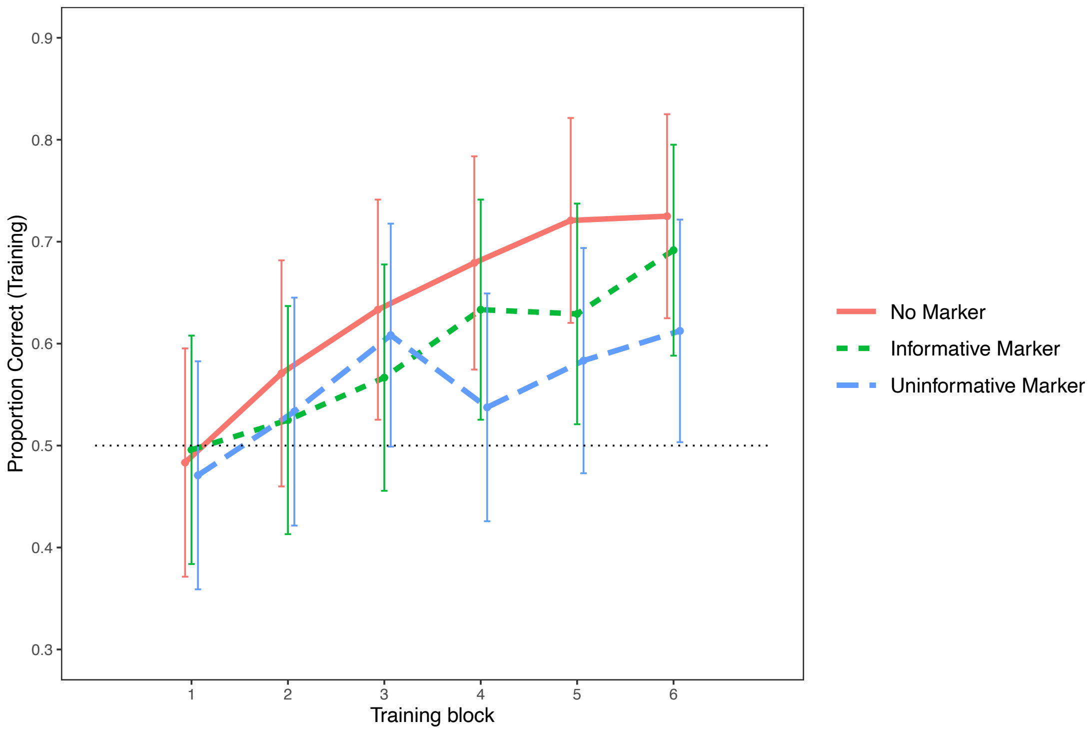

Figure 2. Mean proportion correct across training for each language condition. Error bars indicate standard error of the mean. The dashed black line shows chance performance.

Table 3. One-sample t-tests comparisons for participants' performance against chance and Cohen's d effect size (95% confidence intervals for Cohen's d values in parentheses) for each language condition in each of the six training blocks.

Block	No Marker		Informative Marker		Uninformative Marker	
	t	*d*	*t*	*d*	*t*	*d*
1	-0.52	0.12 (-0.79, 0.46)	-0.11	0.03 (-0.66, 0.59)	-0.65	0.15 (-0.82, 0.43)
2	1.67	0.37 (-0.15, 1.18)	0.49	0.11 (-0.47, 0.79)	0.75	0.17 (-0.40, 0.86)
3	2.91*	0.63 (0.15, 1.65)	1.25	0.28 (-0.26, 1.03)	2.51*	0.56 (0.06, 1.50)
4	3.23*	0.72 (0.23, 1.78)	2.59*	0.58 (0.08, 1.53)	0.71	0.16 (-0.41, 0.85)
5	3.93*	0.72 (0.39, 2.06)	2.19*	0.49 (-0.02, 1.37)	1.53	0.34 (-0.19, 1.13)
6	3.44*	0.77 (0.28, 1.87)	3.61*	0.81 (0.31, 1.94)	2.06	0.46 (-0.05, 0.43)

Note: * indicates $p < .05$.

Table 4. Fixed effects in the generalised linear mixed effects model of training trial accuracy, with no marker condition as reference category.

Fixed effects	estimate	SE	*z*	*p*
(intercept)	-0.188	0.231	-0.816	.414
Block	0.246	0.035	6.920	<.001
Language condition (informative marker)	0.012	0.326	0.038	.970
Language condition (uninformative marker)	0.150	0.324	0.461	.645
Block: Language condition (informative marker)	-0.058	0.049	-1.184	.236
Block:Language condition (uninformative marker)	-0.142	0.049	-2.918	.004

Note. Marginal R^2 = .030, Conditional R^2 = .211.

Syntax in R: glmer(`Accuracy ~ blocks * Condition + (1 | Participant) + (1 | item), family = binomial)`

For the generalised linear mixed effects models, for training block, the model with random slopes resulted in singularity, and so only random intercepts were included. There was a significant effect of training block, showing that over all three conditions, participants learned through the experiment, $\chi^2(1) = 80.42$, $p < .001$. There was no significant effect of condition, $\chi^2(2) = 1.41$, $p = .494$, but there was a significant interaction between block and condition, $\chi^2(2) = 8.66$, $p = .013$, with the no marker condition resulting in a greater increase in accuracy with training than the uninformative marker condition, with no observable difference between the no marker and the informative marker conditions. We relevelled the effect of condition in the analyses to compare the uninformative marker and informative marker conditions, but the interaction with block was not significantly different, estimate = -0.084, SE = 0.048, z = -1.744, p = .081. The final model, reporting R^2 values for model fit (Nakagawa et al., 2017), is shown in Table 4.

Table 5. Testing trials accuracy against chance, showing one-sample t-tests and Cohen's d effect size (95% confidence interval of effect size in parentheses).

Language Feature	No Marker		Informative Marker		Uninformative Marker	
	t	*d*	*t*	*d*	*t*	*d*
Nouns	2.77*	0.62 (0.12, 1.60)	2.08	0.46 (-0.05, 1.33)	2.48*	0.56 (0.05, 1.49)
Verbs	3.52*	0.79 (0.29, 1.90)	2.91*	0.65 (0.15, 1.65)	3.62*	0.81 (0.32, 1.94)
Markers	.	.	-0.16	0.04 (-0.67, 0.57)	.	.
Word order	13.41*	3.00 (2.21, 6.25)	7.34*	1.64 (1.08, 3.53)	20.44*	4.57 (3.47, 9.46)

Note: * indicates $p < .05$.

Testing trials

It is possible to discern the target scene from the foil scene based on only partial information from the language – for instance, if the participant had learned only the verb then for most trials, it would be possible to accurately select the correct scene in the training trials. In order to determine what features of the language were acquired, we assessed the testing trials for each language feature separately. For each testing trial (where the two possible scenes differ by just one property) selecting the scene that corresponds to the sentence was assigned 1, and 0 if the incorrect scene was selected.

First, in explorative analyses, we determined whether performance was above chance using one-sample t-tests for each language feature for each condition separately. The results are shown in Table 5. The results indicate that learning was effective for nouns, verbs, and word order in all three conditions – though this was not significant for nouns for the informative marker condition (p = .051). For marker words, this was only tested in the informative marker condition, and this resulted in no significant evidence of learning[7]. Figure 3 shows the mean accuracy by participant for each language feature.

For the generalised linear mixed effects model of accuracy for noun test trials, only random intercepts and no slopes were included in the final model. The addition of language condition to the null model containing only random effects resulted in no significant effect, $\chi^2(2) = 0.202$, $p = .904$, marginal $R^2 = .001$, conditional $R^2 = .116$, thus nouns were not learned significantly better in any of the language conditions. For the verb test trials, again only random intercepts were included in the final model. The addition of language condition did not significantly improve model fit, $\chi^2(2) = 0.662$, $p = .718$, marginal $R^2 = .006$, conditional $R^2 = .362$, with none of the conditions resulting in significantly different verb learning.

For the word order trials, the linear regression model with language condition as a predictor was significant (see Table 6), $R^2 = .11$, $p = .034$. D-prime values were significantly higher (so discrimination between grammatical and ungrammatical sentences better) for the no marker condition than the informative marker condition, beta = 0.67, SE = 0.27, $t = 2.49$, $p = .016$, and also higher for the uninformative marker condition than the informative marker condition, beta = 0.57, SE = 0.27, t =

2.09, *p* = .041. The no marker and uninformative marker conditions did not differ significantly when the language condition was relevelled so that the no marker condition as the reference level, beta = 0.11, SE = 0.27, *t* = 0.40, *p* = 0.691.

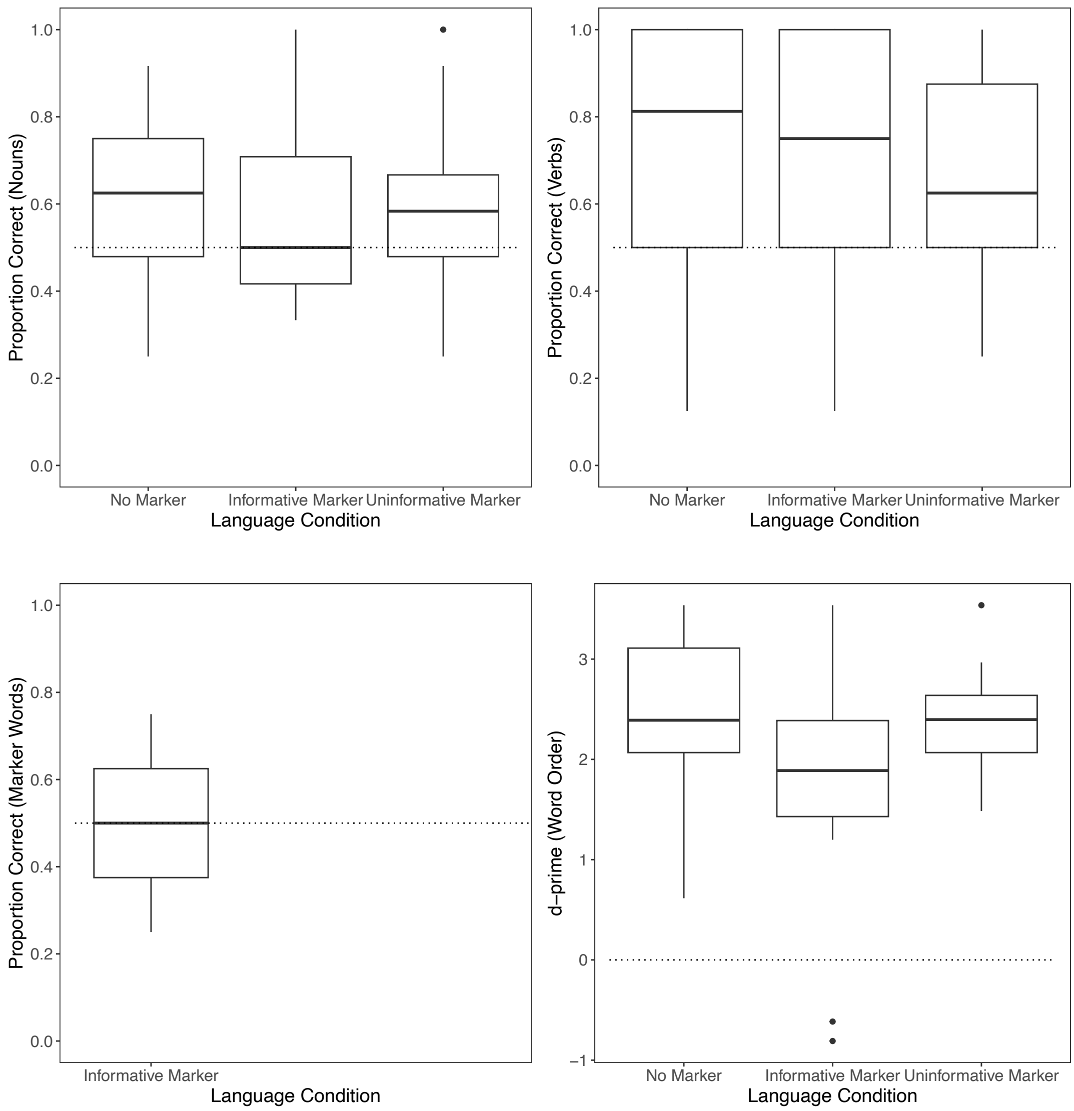

Figure 3. Proportion correct for nouns, verbs, marker words, and d-prime for word order in the test trials. Dashed line indicates chance performance.

Table 6. Linear regression of d-prime values for word order testing trials.

Fixed effects	*estimate*	*SE*	*t*	*p*
(*intercept*)	2.493	0.191	13.045	<.001
Language condition (informative markers)	-0.674	0.270	-2.494	.016
Language condition (uninformative markers)	-0.108	0.270	-0.399	.691

Syntax in R: lm(`dprime ~ Condition`)

Discussion

This study assessed the conditions under which languages can be learned from correspondences between complex multi-word sentences and scenes, when varying the role of morphological markers to indicate subject and object roles in the language. The languages either contained morphological markers to subject and object roles of sentences, or contained morphological markers that did not indicate subject and object roles, or had no morphological markers at all. In each of these languages, word order was not a cue to subject or object roles, in contrast to Tal and Arnon's (2022) study. In terms of identifying which scene corresponded to a heard sentence during training, learning was found to be general across the three language conditions. In particular, when there were informative markers indicating subject and object roles and when there were no markers at all, performance on selecting the target scene corresponding to the sentence was accurate, and learning was rapid. After hearing only 36 sentences, participants were already above chance in detecting the scene. The evidence for the condition where the markers were uninformative about subject and object roles was more mixed, with performance above chance by block 3, then not increasing substantially with more training.

These results replicate previous studies (Rebuschat et al. , 2021) that have shown that artificial languages containing multiple grammatical categories can be learned when presented alongside scenes indicating transitive actions. Multiple possible mappings between each of the words in the language and the actions and objects in the environment are tracked by the learner, where the intended mapping between words and their referents are rapidly acquired, and the word order in which those words occur is also ascertained quickly and accurately by learners. In Rebuschat et al. (2021), the meaning of the (informative) marker words was also found to be learned above chance, however, in the current study, there was no direct evidence of learning of the marker words (in the informative marker condition for the marker word test items).

These previous results by Rebuschat et al. (2021) are also extended by demonstrating that learning is robust to variations in the language itself. The no marker condition provides an underspecified language to the learner, in that the first word may be the subject or the object of the sentence, and there is no clue in the language itself to indicate the role of the nouns. However, information about the subject or object role of the animals with respect to the action they perform is present in the environment. The overlapping constraints from information contained in the language and the environment are shown to support learning, even when there is not sufficient information within the language itself to constrain the interpretation. This highlights that language learning cannot be isolated from the information contained in the broader environment within which speakers produce utterances. Learners are able to combine information from complex scenes to support acquisition of words and word order structure in the language.

The first key research question in our study was whether this learning of cross-situational correspondences between sentences and scenes could occur even when the language itself was underspecified with respect to the relations among words in the sentence – i.e., when subject and object roles of nouns were not indicated by either word order or morphosyntax. The results were positive. For all language conditions, learning was above chance by the end of training. For the tests of individual features of the language at the end of training, again for all conditions performance was above chance for recognising the word order and the verbs of the language. There was also evidence of learning the nouns in the no marker and uninformative marker conditions at a level that was greater than chance, though performance was not significant for the informative marker condition, but the mixed effects modelling results indicated that the three conditions were not significantly different than one another in terms of overall noun learning. However, there was limited evidence of learning the role of the morphological markers to subject and object roles in the informative marker condition, and so it remains a possibility that attention to these markers was not relevant, and certainly not required, for acquisition of the language.

Overall, we found evidence of learning that extended across both word order and vocabulary, demonstrating that the meanings of words and the relative positions of words in sentences can be acquired at the same time, even when structure within the language is formally insufficient for learning. Hence, acquisition occurs through integration with information present in the environment. The results therefore demonstrate that apparently impossible languages – those that are underspecified with respect to the relations between words in the sentence – can be learned (or at least they can begin to be acquired in terms of learning the nouns, verbs, and aspects of word order) when co-occurrences with information in the environment that the language describes is integrated into learning.

The second research question was whether overlapping information between the language structure and the environment promoted learning. We compared languages where subject and object roles are indicated in both the environment and the morphological structure within the language to languages where the morphological structure was absent, so subject and object roles were only evident in the environment. In studies which have included multiple overlapping cues *within* the language (e.g., word order and morphosyntactic markers) learning grammatical structure appears to be promoted, at least in the case of children learning the language (Tal & Arnon, 2022), whereas in the current study when these cues were distributed across modalities (within the language and outside the language in the broader environment) then there was no boost to learning observed.

In contrast to studies finding an advantage for redundant information when that redundancy was present within the language structure, in our study we found evidence that the extra simplicity that results from omitting morphological information may instead result in an advantage for learning. For the no marker condition omitting morphosyntax enabled a simplification of the structure of the language, resulting in a language without subject and object markers, and consequently shorter sentences. This greater simplicity resulted in a steeper trajectory of learning during training when compared to the more complex (but also under-specified) uninformative marker condition. Note that the greater simplicity exerted its effect both on the training trials, and on the test trials, which were all shorter in the no marker condition than the other conditions. The uninformative marker condition provided a control for the omission of morphological markers regarding subject and object roles to the no marker condition, but also a control for overall sentence length with the informative marker condition. The uninformative marker and informative marker conditions did not result in a significant difference in learning during training, supporting evidence that simplicity (i.e., the advantage of the shorter language in the no marker word condition) was to a degree driving the results during learning.

For the other test trials, simplicity of language structure did not fully explain the observed results. The noun and verb test trials did not differ according to any condition, but the word order testing trials did: both the no marker and uninformative marker conditions resulted in more accurate performance than the informative marker condition. In this case, the lack of informative morphological information seemed to drive more accurate detection of the word order of the nouns and verbs in the sentence. Note that the uninformative marker condition, controlled in terms of sentence length and number of marker words with the informative marker condition, may have elicited a different learning approach: requiring unlearning that the marker words indicate any structure, and the slight drop in learning during training from block 3 to block 4 may indicate this reconfiguration of learning (see Figure 2). Though the effect of learning the function of the marker words in the informative marker word condition was not replicated from Rebuschat et al. (2021), there was some indirect evidence that these marker words were inducing an effect on learning, even though this seemed to be impeding the acquisition of knowledge about the word order of the language. This may be because the participants' focus is on determining the role of these marker words in the informative marker condition, prior to learning exactly how those markers relate to subject and object roles. Walker et al. (2020) showed that doubling the length of training compared to the current study increased the ability of participants to detect the subject and object roles of marker words, and so extending the training time in an extension of the current study may increase the fidelity of the direct and indirect effects of language structure on learning.

Our participants' learning when there is no dominant word order and no morphological marking of subject and object roles highlights that, when the language and the environment in which the language is spoken are taken as joint contributors to learning, there are more possibilities for how and where information about subject and object roles are represented in the language. This learnability of an underspecified language points to flexibility in language structure that permits a broader space within which language structure can change over time, and provides a potential explanation for how the 11 languages without reliable word order or case marking, reported in Table 1, could be acquired (Iggersen, 2013). Languages that develop conventionalised word order are more likely as a consequence to lose morphological case marking, though diachronic studies have suggested there are also numerous cases of language change from fixed to free word order (Gell-Mann & Ruhlen, 2011). Yet, even when languages are in transition – where cues to grammatical roles are unreliable – learning is likely still to be possible by coordination of the language information with the environment that the language describes.

The investigation here into languages lacking redundancy within the linguistic features of the language provides a counterpoint to studies that have investigated how linguistic redundancy can support learning. According to the Competition Model, languages will tend to include information about grammatical structure either in word order or in morphology (MacWhinney & Bates, 1989), whereas our studies have shown that this information can also be included in links between the linguistic structure and broader information within the visual environment around the learner. Dale and Lupyan (2012) theorised, and Tal and Arnon (2022) demonstrated experimentally, that redundancy in the linguistic structure – indicating subject and object roles in both word order and morphology – could improve learning, at least for some participants. Hence, redundant information within the language can promote learning. Yet, our studies indicate that when there is overlapping information that spans across modalities – linguistic and environmental – then this did not provide evidence of learning at higher levels of accuracy than when there was no redundancy in the subject and object roles of nouns in the language.

It is possible that the lack of an advantage from including morphological information (which provides overlapping information between language structure and environment) in our study is due to the predominance of speakers of English as a first language where there was no expectation of information about morphology marking subject and object roles. However, in Tal and Arnon's (2022) study, child Hebrew-speakers could learn better from the redundant information even though Hebrew also does not have morphology marking subject and object roles. Yet, Tal and Arnon (2022) found that adult Hebrew-speakers did not improve with redundancy, though they considered that this was likely because learning was high in both redundant and non-redundant conditions. In our study, learning was not at ceiling levels, and so it remains a possibility that overlapping information between morphological information and subject and object roles in the environment does not boost learning, with redundancy within language only able to provide this increase to learning. It could be that providing this overlapping information between language and environment in terms of word order constraints, rather than morphological cues, may promote learning, particularly as the participants were speakers of English, where word order, but not morphology (except in the case of pronouns), provides this information.

Determining how participants approach the task, and what they learn explicitly about the task, would be a valuable addition to future studies using this, and related, paradigms. For instance, Monaghan et al. (2021) found that, for a cross-situational learning study of a simpler language, explicit knowledge about language structure was related to improved learning, especially during the middle stages of exposure. Relatedly, Reber (1993) and Wonnacott et al. (2012) described how differences in task understanding may affect what is learned. Furthermore, how participants are tested may have affected the information that they used for learning. In Tal and Arnon's (2022) study, the subject and object roles were required as a part of the participants' response to the task, whereas in our paradigm, this was not explicitly required (except for the marker word trial tests for the informative marker word condition). Thus, the task requirements may affect the "experimental pragmatics" (Wonnacott et al., 2012) contributing to the influence of redundancy to what is learned from the language.

The learning of correspondences between complex sentences and scenes provides a proof of concept for how the early stages of language acquisition can proceed when the learner has no knowledge of what words mean nor what syntactic roles occur in the language nor how they are expressed in the language. Our learners can acquire information both about word-referent mappings and the relative position of (at least some) words in sentences. Learning meanings of words is an example of a "difficult problem", but not all words are equal, and some are harder to learn than others (Gleitman et al., 2005). As an example of words that are harder to acquire, Gleitman (1990) describes cases where there is no feature of the environment to define unambiguously the meaning of transitive verbs, such as for the distinction between "flee" versus "chase", or "buy" versus "sell". In these cases, the action observed in the environment will be identical, but the meaning of the verb requires the subject and object roles of the nouns to first be resolved. This "impossible" problem thus requires the syntax of the utterance to be understood prior to the meaning of the words being acquired.

Though we have not directly addressed this in our study, the results of our study provide an inkling as to how meanings of these "hard words" (Gleitman et al., 2005), which require integration of syntactic information to constrain word meanings, can be learned – from subtle cross-situational statistics where cues about how events unfold in the environment are combined with constraints that are present within the language itself. Our participants showed how they can begin to acquire word-referent mappings as well as some of the word order constraints of the language. However, our language did not contain any of these hard words, such as "flee" versus "chase". Determining how these verbs could be acquired using a similar paradigm to that used here would be an interesting point of future study. For these "hard word" verbs, the action observed in the environment would be identical, with novel words referring to "chase" and "flee" both presented for learning by participants, where the meaning of the verb requires the subject and object roles of the nouns to first be resolved. The underspecified language without marker words would not be able to provide this information.

Interestingly, even the informative marker condition requires additional information to constrain interpretation of the verb. That is, subject and object marker words on their own would not be sufficient to determine which is the subject and which is the object – only that they are distinct. For instance, for the sentence in Table 2, "melbai noo bimdah tha racken", meaning "owl subject-marker turtle object-marker lifts", if noo is learned as an object marker, and tha is learned as a subject marker, then the sentence could be interpreted as “turtle lifted by owl”. Thus, either subject markers or verbs need to be learned in unambiguous contexts, such as intransitive sentences, or the environment would need to provide additional information about communicative focus, such as indicating the intended agent of the action, and thus subject of the sentence (Tomlin, 1995).

A further limitation, and potential point for extension, of the current study is that the language and environment are still simple compared to the complexity of children’s language input and environment to which the language refers. Though the current language exceeded typical utterance length for children’s input (e.g., Hoff & Naigles, 2003), the sentence structure was limited to transitive sentences, the vocabulary was small (comprising only 12 words), and the complexity of the scenes was very restricted (4 objects, and 2 actions, across the 2 scenes). How learners scale up their learning to more complex scenes, and a larger vocabulary, though the current study does indicate how this learning may at least commence for frequently occurring words in restricted scenes (e.g., Yu & Ballard, 2007).

The studies of learning in this experiment involve comprehension of a sentence with the referent of the sentence concretely available. A further challenge to communication comes from describing events that are not present. In this case, as for the example of the “hard words”, a language with free word order and no morphological marking cannot convey subject and object cases, and so a degree of ambiguity in whether “penguin chases monkey” or “monkey chases penguin” is the intended meaning would be present. However, during language learning, language is frequently produced when referents are present in the environment (e.g., Yu & Ballard, 2007; Yu et al., 2021), meaning that acquisition of an underspecified language is still possible even if further communicative tools are later required to convey meaning of events distant in time or space (Snow, 1990).

In summary, we have shown how some aspects of vocabulary and word order constraints – an important aspect of grammar – can be acquired from integration of linguistic and environmental information about events to which the language refers. Though there was no direct evidence for use of morphological information to constrain subject and object roles in this study, the results point the way toward a potential solution to a long-standing question about how language can be learned given the interconnectedness of vocabulary and grammar. Further, the study illustrates that considering a broader environmental context that extends beyond the structure present in the language itself highlights a broader space of possibilities for how languages can express grammatical roles and how this expression can change diachronically.

Footnotes

1. Note that further determining the meaning of *chase* in contrast to *flee* would require additional knowledge of the subject and object roles. We return to this point in the Discussion.

2. Note that this deviates from the structure of a natural language, where uninformative cues do not occur, though morphological cues that refer to features other than subject or object roles (such as gender, number or tense) may need to be acquired.

3. We conducted post hoc power analyses, and simulations of power where the sample size was varied, and found that power was effective for determining the key effect of learning during training (power = 100%), as well as the key interaction between training time and comparisons between the no marker and uninformative marker conditions (power = 84%). We acknowledge that post hoc power analyses are not ideal for justifying participant numbers, though this is a better alternative than conducting power analyses based on simulations of studies that are less closely related to the current study. Data simulations revealed that distinguishing between the no marker and informative marker condition would require 100 or more participants per group, and so this was an effect that is likely underpowered in our current study. However, as this was just one effect investigated in the current study, we opted not to extend further the sample size for the current study, which is not recommended after data analysis has commenced (Simmons et al., 2011) and is consistent with our pre-registration of the analysis (see https://osf.io/hxqzc/?view_only=ea6ba6fff6bb468e8de2e8596f029dca for details of the power simulations).

4. Age distributions were not different across the three conditions: Informative marker condition mean age = 25.3 years (SD = 9.5); No marker condition mean age = 23.4 years (SD = 9.1); uninformative marker condition mean age = 24.1 years (SD = 10.1). |*t*(38)| <= 0.65, *p* >= .522, |*d*| <= 0.11.

5. Unfortunately, our anonymisation of the data meant that we were unable to identify exactly which participants these were, so we could not repeat the analyses omitting these participants. However, if this morphological variation had a substantial effect on learning then we would see it distributed across the three conditions adding variation to the results, rather than skewing them substantially.

6. Note that in the preregistered analysis we stated that d-prime values would be analysed using linear mixed effects models, however, this overlooked the fact that only one d-prime value per participant could be generated and so analysis with item level effects could not be conducted.

7. An alternative way to report the overall effect of learning for the marker words in the marker word condition would be to construct a generalised mixed effects model with random effects but no fixed effects. In this case, the model results provide an intercept only, which corresponds to whether learning is different than chance. The results were consistent with the values reported in Table 5: intercept estimate = -0.025, SE = 0.158, z = -0.158, *p* = .874. marginal R^2 = 0.00.

References

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

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278. https://doi.org/10.1016/j.jml.2012.11.001

Bickel, B., & Nichols, J. (2013). Exponence of selected inflectional formatives. In Dryer, M.S., & Haspelmath, M. (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.(Available online at http://wals.info/chapter/21, Accessed on 2021-09-07.)

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. https://doi.org/10.1080/15475441.2011.585285

Clerkin, E.M., Hart, E., Rehg, 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 Science, 372*(1711). https://doi.org/10.1098/rstb.2016.0055

Dale, R., & Lupyan, G. (2012). Understanding the origins of morphological diversity: The linguistic niche hypothesis. *Advances in Complex Systems, 15*(3), 1150017. https://doi.org/10.1142/S0219525911500172

Dryer, M.S. (2013). Order of subject, object and verb. In: Dryer, M.S., & Haspelmath, M. (eds.), *The world atlas of language structures online*. Leipzig: Max Planck

Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/81, Accessed on 2021-09-22.)

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

Gell-Mann, M., & Ruhlen, M. (2011). The origin and evolution of word order. *Proceedings of the National Academy of Sciences*, *108*(42), 17290-17295. https://doi.org/10.1073/pnas.1113716108

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1*(1), 3-55. 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/s15473341lld0101_4

Gold, E. M. (1967). Language identification in the limit. *Information and Control*, *10*(5), 447-474. https://doi.org/10.1016/S0019-9958(67)91165-5

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language*. Greenberg, J. H. (ed.). Cambridge, MA.: MIT Press.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*(2), 418-433. https://doi.org/10.1111/1467-8624.00415

Lupyan, G., & Christiansen, M. H. (2002) Case, word order, and language learnability: insights from connectionist modeling. *Proceedings of 24th Annual Conference of the Cognitive Science Society* (pp. 596–601). Mahwah, NJ: Lawrence Erlbaum. https://doi.org/10.4324/9781315782379-138

Iggesen, O.A. (2013). Number of cases. In Dryer, M.S., & Haspelmath, M. (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology.(Available online at http://wals.info/chapter/49, Accessed on 2021-09-07.)

Koplenig, A., Meyer, P., Wolfer, S., & Müller- Spitzer, C. (2017). The statistical trade-off between word order and word structure – Large-scale evidence for the principle of least effort. *PLoS ONE 12*(3), e0173614. https://doi.org/10.1371/journal.pone.0173614

Levshina, N. (2021). Cross-linguistic trade-offs and causal relationships between cues to grammatical subject and object, and the problem of efficiency-related explanations. *Frontiers in Psychology, 12,* 648200. https://doi.org/10.3389/fpsyg.2021.648200

MacWhinney, B., & Bates, E. (1989). *The crosslinguistic study of sentence processing*. New York: Cambridge University Press.

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*(4), 831–877. https://doi.org/10.1037/a0029872

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.

Monaghan, P., Ruiz, S., & Rebuschat, P. (2021). The role of feedback and instruction on the cross-situational learning of vocabulary and morphosyntax: Mixed effects models reveal local and global effects on acquisition. *Second Language Research*, *37*(2), 261-289. https://doi.org/10.1177/0267658320927741

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface, 14*(134), 20170213 https://doi.org/10.1098/rsif.2017.0213

Quine, W.V.O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Reber, A. S. (1993). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. Oxford, UK: Oxford University Press.

Rebuschat, P., Monaghan, P., & Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition*, *206*, 104475. https://doi.org/10.1016/j.cognition.2020.104475

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

Shibatani, M. (1990). *The languages of Japan*. Cambridge: Cambridge University Press.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science, 22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

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

Smith, K., Smith, A. D. M., & Blythe, R. A. (2011). Cross-situational learning: An experimental study of word-learning mechanisms. *Cognitive Science*, *35*(3), 480–498. https://doi.org/10.1111/j.1551-6709.2010.01158.x

Smith, L. B., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568. https://doi.org/10.1016/j.cognition.2007.06.010

Snow, C. E. (1990). The development of definitional skill. *Journal of Child Language*, *17*(3), 697-710. https://doi.org/10.1017/S0305000900010953

Tal, S., & Arnon, I. (2022). Redundancy can benefit learning: Evidence from word order and case marking. *Cognition, 224*, 105055. https://doi.org/10.1016/j.cognition.2022.105055

Tomlin, R. (1995). Focal attention, voice, and word order: an experimental, cross-linguistic study. In Downing, P, & M. Noonan (Eds.), *Word order in discourse*, 517-554. Amsterdam: Benjamins. https://doi.org/10.1075/tsl.30.18tom

Van Everbroeck, E. (1999). Language type frequency and learnability. A connectionist approach. *Proceedings of the 21st Annual Conference of the Cognitive Science Society* (pp. 755-760). Mahwah, NJ: Erlbaum. https://doi.org/10.4324/9781410603494-136

Walker, N., Monaghan, P., Schoetensack, C., & Rebuschat, P. (2020). Distinctions in the acquisition of vocabulary and grammar: An individual differences approach. *Language Learning*, *70*(S2), 221-254. https://doi.org/10.1111/lang.12395

Wonnacott, E., Boyd, J. K., Thomson, J., & Goldberg, A. E. (2012). Input effects on the acquisition of a novel phrasal construction in 5 year olds. *Journal of*

Memory and Language, 66(3), 458-478. https://doi.org/10.1016/j.jml.2011.11.004

Yu, C., & Ballard, D. H. (2007). A unified model of early word learning: Integrating statistical and social cues. *Neurocomputing*, *70*(13), 2149–2165. https://doi.org/10.1016/j.neucom.2006.01.034

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

Yu, C., Zhang, Y., Slone, L. K., & Smith, L. B. (2021). The infant's view redefines the problem of referential uncertainty in early word learning. *Proceedings of the National Academy of Sciences*, *118* (52), e2107019118. https://doi.org/10.1073/pnas.2107019118

Yurovsky, D., Yu, C., & Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, *37*(5), 891–921. https://doi.org/10.1111/cogs.12035

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Additional Information on Materials and Instructions for Participants.

Appendix S2. Additional Examples of Training Trials.

Appendix S3. Additional Analyses of Quadratic Versus Linear Effects during Training.

Appendix S4. Additional Analyses of Syntax Test Trials, Omitting Items that Occurred During Training.