Simultaneous acquisition of vocabulary and grammar in an artificial language learning task

Neil Walker (n.walker@lancaster.ac.uk)
Department of Linguistics and English Language
Lancaster University, Lancaster LA1 4YL, UK

Christine Schoetensack (s1649700@sms.ed.ac.uk)
School of Philosophy, Psychology and Language Sciences
Edinburgh University, Dugald Stewart Building, 3 Charles Street, Edinburgh, EH8 9AD

Padraic Monaghan (p.monaghan@lancaster.ac.uk)
Department of Psychology
Lancaster University, Lancaster LA1 4YF, UK

Patrick Rebuschat (p.rebuschat@lancaster.ac.uk)
Department of Linguistics and English Language
Lancaster University, Lancaster LA1 4YL, UK

Abstract

Learning syntax requires determining relations between the grammatical categories of words in the language, but learning those categories requires understanding the role of words in the syntax. In this study, we examined how this chicken and egg problem is resolved by learners of an artificial language comprising nouns, verbs, adjectives and case markers following syntactic rules. We found that the language could be acquired through cross-situational statistical correspondences with complex scenes and without explicit feedback, and that knowledge was maintained after 24 hours. Results also showed that verbs and word order were the first to be acquired, followed by nouns, adjectives and finally case markers. Interdependencies in learning were found for word order and verbs, and also for nouns, adjectives and case markers. Grammar and vocabulary can be acquired simultaneously, but with distinctive patterns of acquisition – grammar and the role of verbs first, then the vocabulary of other lexical categories.

Keywords: language acquisition; grammar; vocabulary; artificial language learning; statistical learning.

Introduction

The early stages of language learning involve a great deal of ambiguity as learners must make sense of the stream of input they hear by noticing words boundaries, decoding the meanings of words, identifying lexical categories and understanding the relations between categories defined by the syntax. How this is achieved and the order in which vocabulary and grammatical knowledge is acquired has been a critical question in the cognitive sciences (Marcus, 1996).

Cross-situational learning

Recent research has shown that it is possible for children and adults to learn vocabulary within basic categories of words when they are presented across numerous ambiguous learning situations without any feedback, a mechanism known as cross-situational learning. Smith and Yu (2008) showed that 12 to 14-month old infants could learn the meanings of novel nouns by keeping track of cross-trial statistics. Scott and Fisher (2012) also demonstrated that it is possible for 2.5 year-old toddlers to learn novel verbs, utilising syntactic cues, knowledge of nouns and other situational referents. Monaghan et al. (2015) found that nouns and verbs could be learned simultaneously without any syntactic cues, although nouns were learned slightly more quickly. They suggested that this prioritisation of nouns could be explained by the greater saliency and stability of object versus action referents.

However, these studies on the cross-situational learning of nouns and verbs are a substantial abstraction from the complexity of natural language acquisition. In child-directed speech, children are generally exposed to multi-word utterances containing many word categories (Mintz, 2006). With every new word category or syntactic phrase added, the number of possible referents for any given word increases, making the tracking of statistical probabilities more complex.

On the other hand, with greater complexity comes greater interdependency between content words, function words and word order. And so conversely, this extra complexity may also provide additional cues from which to constrain learning. Monaghan and Mattock (2012), for example, found that function words could aid the learning of nouns in a cross-situational learning paradigm.

A key question is, therefore, how learners can break into this complex stream, given the difficulty of attempting to acquire the syntax to indicate the role of grammatical categories, and learning the vocabulary to populate those categories. This chicken and egg problem has proven difficult to resolve, and has led to proposals either for independence of learning grammar and vocabulary (e.g., Marcus, 1996), or their inter-relatedness (Bates & Goodman, 1997). Under these latter accounts, learning a

few words can give rise to syntactic bootstrapping (Gleitman, 1990), which can then be used to promote further vocabulary acquisition, predicting correlations in children's performance for vocabulary and grammatical processing tasks. However, evidence from actual language learning tasks in which both vocabulary and grammar must be acquired has not been extensively explored. Many previous studies of artificial language learning have trained participants on vocabulary before testing them on a language structure.

In the current study, we investigated whether it is possible to learn more complex artificial languages that combine content words, function words and syntactic structures under cross-situational learning conditions without any feedback.

The acquisition order of linguistic categories

The second question this study addressed was that of acquisition order: When are nouns, verbs, adjectives, case markers, and syntactic constraints on word order acquired and are some aspects learned before others? The vast majority of studies into early childhood language learning support Gentner's (1982) claim that across all languages children learn nouns before verbs and adjectives (e.g, Bornstein et al., 2004). One reason might be that predicates are more semantically complex as they modify and depend on nouns (Dixon, 1982), whether that be adjectives (*the black dog*) or verbs (*the cat pounced on the mouse*). Therefore, in order to learn the verb or adjective, learners need to also encode information about the noun (Gleitman et al., 2005).

However, there is some evidence that in languages such as Korean and Japanese, where the verb is found in a highly-salient sentence-final position and subjects and objects are often left out of utterances by caregivers, verbs may be learned earlier than in SVO languages such as English (Choi & Gopnik, 1995). There is, however, a lack of consensus as to whether verbs in these languages are learned at the same time as nouns (see Bornstein et al., 2004).

Regarding adjectives, Booth and Waxman (2003) demonstrated that 14-month-old children could extract the meaning from nouns but not adjectives when presented with basic syntactic and visual frames. In a large corpus-based study, Behrens (2006) found that German children aged 1 year 11 months produced more verbs than adjectives.

Finally, case markers which indicate the agent and patient of a sentence have been shown to be understood by children as young as two years old (Göksun et al., 2008). However, in this and other studies, a small vocabulary of nouns and verbs had been acquired before comprehension of case markers was demonstrated.

The participants in the current study were adults who have already mastered their first language, and so it is arguable that the acquisition order observed in child language development may not apply to these learners. An alternative strand of research comes from first exposure studies of adults learning a second language (L2; for an overview, see Indefrey & Gullberg, 2010). In a study into the initial stages of learning an L2 by adults in a classroom setting, Shoemaker and Rast (2013) found that it was the words in sentence-initial and sentence-final positions that were most easily recognised in a stream of speech. They argued that this was due to not only silence bordering the initial and final words, but that working memory is less burdened for the final word of the utterance.

Another factor that influences whether a word can be picked out of a stream of speech is the number of syllables it contains (Gullberg et al., 2012). With many function words monosyllabic, this could render them less easily noticed than highly salient content words, despite the frequency with which they occur in utterances.

Overall, if the learning of nouns and verbs follows the findings of child language research, then we can expect nouns to be learned before verbs. Alternatively, if sentence position is a more important factor, we could expect verbs, which in our current study occupy the sentence-final position, to be learned before nouns, which are mostly in medial position. We then predict that adjectives will be learned next, followed by case markers, although given the short duration of the learning paradigm, it is possible that the latter may not be learned at all (e.g., DeKeyser, 2005).

The learning mechanisms of vocabulary and syntax

A final aim of this study was to investigate how the different types of language structure cohere. Is the meaning of vocabulary items (nouns, verbs, adjectives) learned in the same way as grammatical items (word order, case markers) or do they depend on different mechanisms? Research from models of learning data (Frost & Monaghan, 2016), neuropsychology patient studies (Alario & Cohen, 2004), theoretical models (Bock & Levelt, 1994), and memory models (Ullman, 2004) treat vocabulary and syntax as distinct. If this were the case, we might expect word order and case markers to be interdependent, with nouns, verbs and adjectives also grouped together. Alternatively, if syntax and vocabulary share the same learning mechanism, as is postulated in single-system models (MacWhinney, 1987), we might expect to see no interdependency of word order and case markers. Instead, as word order is determined by the position of the verb, it is possible that learning which word is the verb and the word order will be linked.

Method

Participants

Sixty-four native speakers of English (47 women) were randomly assigned to two conditions (*massed* vs *distributed*, each n = 32) which varied in terms of whether there were pauses between blocks of training on an artificial language learning task. Participants were students or graduates of universities in the North West of England. The mean age was 26.0 years (*SD* = 7.1). None of the participants had previously studied any verb-final languages. Participants in

the massed group received 20 GBP and participants in the distributed group received 28 GBP. The difference was due to the extra time involved in the distributed condition.

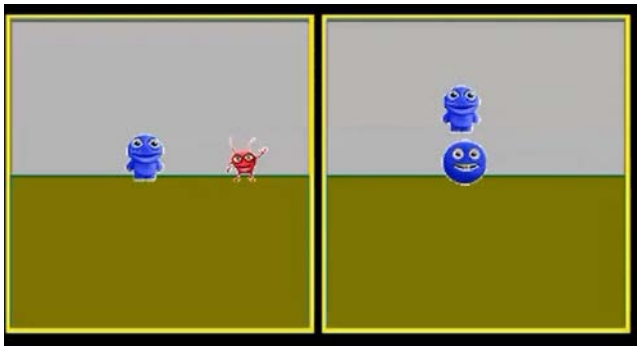

Figure 1: Screenshot of the cross-situational learning task. Participants see two dynamic scenes and hear a sentence and decide which scene the sentence refers to.

Materials

A novel artificial language was created for this experiment. The lexicon consists of 16 pseudowords, taken from Monaghan and Mattock (2012). Fourteen bisyllabic pseudowords were content words: Eight nouns, four verbs, and two adjectives. Two monosyllabic pseudowords served as function words that reliably indicated if the preceding noun referred to the subject or the object of the sentence. The words were recorded by a female native speaker of British English who was instructed to produce the words in a monotone.

In terms of syntax, the artificial language was based on Japanese. Sentences could either be SOV or OSV, i.e. verbs had to be placed in final position but the order of subject and object noun phrases (NPs) was free. NPs had to contain a noun as its head and a post-nominal case marker that indicated if the preceding noun was the agent or the patient of the action. Adjectives were optional and only occurred in half the NPs. Adjectives occurred pre-nominally.

Eight alien cartoon characters served as referents for the language (see Figure 1). The aliens could either appear in red or blue and were depicted performing one of four actions (hiding, jumping, lifting, pushing) in dynamic scenes generated by E-Prime (version 2.0). Figure 1 shows a sample screen shot, containing the target scene and a distractor scene. Each noun referred to one alien, the adjectives referred to the colours of aliens, and the verbs referred to the actions. Word-referent mappings were randomly generated for each participant to control for preferences in associating certain sounds to objects, motions or colours.

For training, there were 12 blocks of 16 trials each. In each trial, two scenes were presented and an artificial language sentence played. The sentence described only one of the scenes and the participants had to match the sentence to the correct scene. Within each block, each alien and action occurred an equal number of times; half the utterances in each block were SOV, the other half OSV. In the distractor scene, no actions were the same and the aliens and their colours were randomly selected. The locations of the target scene were counterbalanced.

For testing, each type of information in the language was assessed by presenting an utterance and varying the target and distractor scenes by one piece of information: For testing nouns, target and distractor scenes were identical except for one of the aliens; for testing verbs, only the scenes' actions differed; for testing adjectives, one of the colours of an alien was changed; and for testing marker words, the two scenes depicted the same aliens performing the same actions but with opposite agent-patient assignment. Testing trials were intermingled with every third training block. The purpose of this was to make it less likely that participants would know they were being tested. For testing word order, grammatical and ungrammatical sentences were presented: Half the trials followed the licensed SOV or OSV order in sentences that had not been presented in the cross-situational learning trials, whereas the other half contained syntactic violations (*VSO, *VOS, *OVS, *SVO).

Procedure

Participants were trained and tested on the artificial language on two days. Participants first completed 16 training and testing blocks. Twenty-four hours later, they returned to the lab to complete a delayed post-test. There were eight pure training blocks, four mixed training and testing blocks, and four grammaticality judgment test (GJT) blocks. In the cross-situational learning task, participants were instructed to observe the two scenes on the screen and listen to the sentence played over headphones. Their task was to decide, as quickly and accurately as possible, which scene the sentence referred to. Participants received no feedback regarding the accuracy of their choice. For the word order trials, participants were told that they would see only one 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 listen carefully and decide if the sentence sounded "good" or "funny".

Presentation order of trials within each block was randomized but all participants completed blocks in the same sequence. There were two training blocks, then one mixed training and vocabulary testing block, then a word order test block. This sequence was then repeated four times.

The massed group completed the first 16 blocks consecutively while the distributed group had three 20-minute breaks after every four blocks, in which they watched a natural history documentary on mute. Training and testing on day 1 took between 70 and 90 minutes for the massed group and between 130 and 150 minutes for the distributed group. The delayed test on day 2 comprised a final block of vocabulary testing trials and then a block of word order testing, and five cognitive tests (not reported here), and lasted approximately 90 minutes.

Results

Performance on training trials

We first performed a mixed analysis of variance on accuracy within each block, with training block as within subjects factor and the two training conditions (massed and distributed) as between subjects factor. There was a significant main effect for block, using the Greenhouse-Geisser correction, $F(4.55, 282) = 42.0$, $p < .001$, $\eta_p^2 = .40$. This indicates that subjects improved with more training. However, there was no significant main effect for group, nor was there a significant interaction between time and group, both $F < 1$. We therefore pooled the data from the two training conditions for the remaining analyses.

In order to ascertain when learning had taken place during the training blocks, a one-sample t-test was conducted to compare the mean scores for each block to a chance score of .5. Participants performed significantly above chance from block two ($M = .57$, $SD = .18$) onwards, 95% CI [.028 to .12], $t(63) = 3.27$, $p = .002$. In other words, 32 trials of exposure (without feedback) were enough to lead to above-chance performance in the cross-situational learning task.

Performance on test trials

In order to determine performance for each type of information in the language, we performed one sample t-tests to establish the first test block at which accuracy was above chance (at .5). We then carried out a series of repeated measures ANOVAs in order to determine the effects of test block on the scores for word order, nouns, verbs, adjectives and markers. Finally, we conducted further repeated measures ANOVAs for test blocks 4 (immediate post-test) and 5 (delayed post-test) for each word type and word order to assess the role of the 24-hour delay. The results are displayed in Figure 2 and Table 1 and Table 2.

Table 1: Summary of repeated measures ANOVA over test blocks 1 to 4 showing effect for block**.**

Test	F	p	h_p^2
Word order	7.82	<.001	11
Noun	14.5	<.001	.19
Verb	2.46	.064	.038
Adjective	2.76	.043	.043
Case marker	.63	.60	.010

Table 2: Summary of repeated measures ANOVA over blocks 4 to delayed test block 5 showing effect for block.

Test	F	P	h_p^2
Word order	.025	.88	.00
Noun	2.90	.59	.005
Verb	4.61	.036	.069
Adjective	2.50	.12	.040
Case marker	2.63	.11	.041

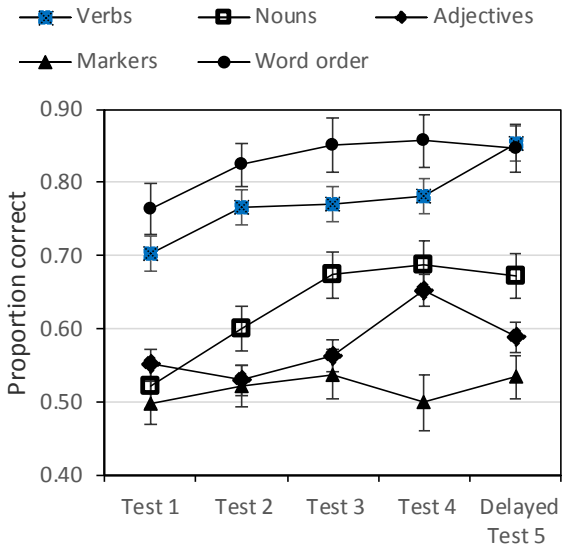

Figure 2: Proportion of correct trials across the five test blocks. Test blocks 1 to 4 were completed on day 1. Test block 5 was administered with a 24-hr delay.

Participants performed significantly above chance from test block 1 onwards for both the word order tests, ($M = .76$, $SD = .19$), 95% CI [.22 to .31], $t(63) = 10.9$, $p = < .001$ and also verb tests, ($M = .70$, $SD = .25$, 95% CI [.13 to .26], $t(63) = 6.16$, $p < .001$). For noun tests, participants performed significantly better than chance from test block 2 onwards, ($M = .60$, $SD = .19$, 95% CI [.052 to .15], $t(63) = 4.16$, $p < .001$). Adjective test results were significantly above chance from test block 4 onwards, ($M = .64$, $SD = .27$), 95% CI [.076 to .21], $t(63) = 4.22$, $p < .001$). Finally, case markers only reached significantly above chance on test block 5, ($M = .54$, $SD = .16$), 95% CI [.001 to .079], $t(63) = 2.06$, $p = .043$.

Determining relations between learning different information types

In order to determine the factors driving performance in the task – whether learning was independent or interdependent for different types of information, we conducted a principal components analysis on test performance for the final test block for word order, nouns, verbs, adjectives, and marker words. There were two components with eigenvalues greater than 1, and the loadings of the individual tests on these components, with varimax rotation, showed a simple solution (i.e., each test loaded > 0.4 on only one component). The components and their loadings are shown in Table 3.

The first component related to learning nouns, adjectives, and marker words, and the second component related to learning word order and verbs. This indicated that performance across the five information types was effectively explained by two aspects of the data: The first relates to learning the vocabulary items of nouns and

Table 3: Loadings of the five delayed tests on the two principal components.

Test	First component	Second component
Noun	**.778**	.104
Adjective	**.769**	.034
Marker words	**.604**	.081
Verb	.322	**.718**
Word order	-.090	**.873**

adjectives and how the marker words affected the role of the adjective-noun phrases, and the second indicated a close relation between learning the identities of verbs and learning that the word order of sentences was verb-final.

Discussion

In this study we investigated whether adult learners could acquire the syntax and vocabulary of a novel language by keeping track of cross-trial statistics, without feedback and without any explicit instruction about the structure of the language or its vocabulary. We also provided a delayed post-test after 24 hours to determine whether any acquired knowledge had been maintained. Furthermore, we examined the order of acquisition and investigated how learning of syntax and of vocabulary cohered.

Simultaneous learning of words and syntax

Our results indicated that adult learners can rapidly acquire both syntax and vocabulary of the language simultaneously. Previous cross-situational learning studies only investigated nouns (Smith & Yu, 2008) or verbs (Scott & Fisher, 2012) or nouns and verbs simultaneously (Monaghan et al., 2015). Our results extend these findings to demonstrate that it is possible for adults to acquire a wider range of information, including adjectives and case markers. We cannot rule out, however, that this occurred because of the nature of the lexical test design, in which the two scenes presented differed only in terms of the lexical item being tested, artificially making these word categories more salient. It also remains to be seen whether children can also learn such a complex system via cross-situational learning, and this is an important question that we are currently addressing. The results also show that the learning effects can be retained overnight. This is an important methodological observation as the majority of studies in statistical learning do not have a delayed post-test, which means that it is unclear whether the learning is robust. By including a 24-hr delayed post-test, we show that learning is indeed robust and that this applies to words and syntax.

For the case markers, it was only after 24 hours that test scores were significantly above chance. This corresponds with Grey, Williams and Rebuschat's (2015) study that found no learning effect for Japanese morphology on an immediate post-test, but a significant effect after a two-week delay. These findings suggest that consolidation may be valuable, particularly for the function words' role in the language. Indeed, there is evidence that sleep aids in the generalization of grammatical rules (Walker & Stickgold, 2010). The case marker results also raise another important methodological consideration. Without the delayed test, we would have underestimated the amount of learning that had taken place and would have concluded that case markers had not been learned at all. Whereas, with the delayed test, there is evidence, albeit a small effect, that learning of case markers does in fact take place. It is recommended, therefore, that future studies into cross-situational learning include delayed post-tests to show that learning is robust and to catch any learning effects brought on through consolidation.

Order of acquisition

Although learners were exposed to both vocabulary and syntax simultaneously, they performed above chance on different aspects of the language at distinct stages: First, verbs and word order were acquired, then nouns, then adjectives, and finally case markers (see Figure 2). It is interesting to note that verbs were learned before nouns in this artificial language and thus differed from the majority of first language acquisition studies. One possible reason for this is the saliency of the final-position verb compared to the mostly medial-position nouns (Shoemaker & Rast, 2013). Another possibility is that adult learners already possess syntactic and lexical knowledge of word categories in their L1 and so can transfer them to their L2. This would then allow the learner to concentrate on deciding which words map onto the different lexical categories, rather than also working out the lexical categories as infants do.

The coherence of vocabulary and syntax

Regarding the coherence of learning of syntax and vocabulary, we found that acquisition of word order and verb learning were interdependent. Upon learning that the final word in the sentence was a verb, participants were able to gain an understanding of the basic word order of the sentence. It is conceivable that such an understanding could be gained by breaking into the stream of input through any word category, with the greater salience of verbs due to final utterance position promoting this acquisition. In addition, we found that nouns, adjectives and case markers were also interdependent but acquired somewhat independently of verbs and word order. This result supports an emergentist view that syntactic knowledge associated with case markers develops only after a core vocabulary of content words has been learned (Bannard et al., 2009).

These results demonstrate that the chicken and egg problem of acquiring grammar and vocabulary can be resolved by the learner through using cross-situational statistics with events in the environment. An alternative explanation is that once verbs are learned, this knowledge is then bootstrapped to aid the acquisition of the other lexical categories. The patterns of results we found for this verb-final language in our experimental paradigm did not neatly correspond with a distinction between grammar and vocabulary learning (e.g., Ullman, 2004), with word order

being related to verb acquisition, and case marking being related to noun and adjective learning. Complex interactions between grammatical categories and grammar do not appear to lend themselves to a neat distinction in acquisition of these sources of linguistic knowledge.

References

Alario, F., & Cohen, L. (2004). Closed-class words in sentence production: Evidence from a modality-specific dissociation. *Cognitive Neuropsychology*, *21*(8), 787-819.

Bannard, C., Lieven, E., & Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proceedings of the National Academy of Sciences*, *106*(41), 17284-17289.

Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia, and real-time processing. *Language and Cognitive Processes, 12*, 507–584.

Behrens, H. (2006). The input–output relationship in first language acquisition. *Language and Cognitive Processes*, *21*(1-3), 2-24.

Bock, K., & Levelt, W. (1994). Language production: Grammatical encoding. Handbook of psycholinguistics, ed. by Morton Ann Gernsbacher, 945-84.

Booth, A. E., & Waxman, S. R. (2009). A horse of a different color: Specifying with precision infants' mappings of novel nouns and adjectives. *Child Development*, *80*(1), 15-22.

Bornstein, M. H., Cote, L. R., Maital, S., Painter, K., Park, S., Pascual, L., & Vyt, A. (2004). Cross-Linguistic analysis of vocabulary in young children: Spanish, Dutch, French, Hebrew, Italian, Korean, and American English. *Child Development*, *75*(4), 1115-1139.

Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of child language*, *22*(03), 497-529.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, 55(S1), 1-25.

Dixon, R. M. (1982). *Where have all the adjectives gone?: and other essays in semantics and syntax* (Vol. 107). Walter de Gruyter.

Frost, R. L., & Monaghan, P. (2016). Simultaneous segmentation and generalisation of non-adjacent dependencies from continuous speech. *Cognition*, *147*, 70-74.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. Technical report no. 257.

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.

Göksun, T., Küntay, A. C., & Naigles, L. R. (2008). Turkish children use morphosyntactic bootstrapping in interpreting verb meaning. *Journal of Child Language*, *35*(02), 291-323.

Grey, S., Williams, J. N., & Rebuschat, P. (2015). Individual differences in incidental language learning: Phonological working memory, learning styles, and personality. *Learning and Individual Differences*, *38*, 44-53.

Gullberg, M., Roberts, L., & Dimroth, C. (2012). What word-level knowledge can adult learners acquire after minimal exposure to a new language? *International Review of Applied Linguistics in Language Teaching*, 50(4), pp. 239-276

Indefrey, P., & Gullberg, M. (2010). The earliest stages of language learning: Introduction. *Language Learning*, 60(s2), 1-4.

MacWhinney, B. (1987). The competition model. *Mechanisms of Language Acquisition*, 249-308.

Marcus, G. (1996). Why do children say "breaked"? *Current Directions in Psychological Science*, 5, 81–85.

Mintz, T. H. (2006). Finding the verbs: Distributional cues to categories available to young learners. *Action Meets Word: How Children Learn Verbs*, 31-63.

Monaghan, P., & Mattock, K. (2012). Integrating constraints for learning word–referent mappings. *Cognition*, *123*(1), 133-143.

Monaghan, P., Mattock, K., Davies, R. A., & Smith, A. C. (2015). Gavagai is as gavagai does: Learning nouns and verbs from Cross-Situational statistics. *Cognitive Science*, *39*(5), 1099-1112.

Scott, R. M., & Fisher, C. (2012). 2.5-year-olds use cross-situational consistency to learn verbs under referential uncertainty. *Cognition*, *122*(2), 163-180.

Shoemaker, E., & Rast, R. (2013). Extracting words from the speech stream at first exposure. *Second Language Research*, *29*(2), 165-183.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558-1568.

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1), 231-270.

Walker, M. P., & Stickgold, R. (2010). Overnight alchemy: sleep-dependent memory evolution. *Nature Reviews Neuroscience*, *11*(3), 218-218.