Language Learning and Development

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/hlld20

Semantic Cues Facilitate Structural Generalizations in Artificial Language Learning

Erin Conwell & Jesse Snedeker

To cite this article: Erin Conwell & Jesse Snedeker (20 Apr 2024): Semantic Cues Facilitate Structural Generalizations in Artificial Language Learning, Language Learning and Development, DOI: 10.1080/15475441.2024.2342873

To link to this article: https://doi.org/10.1080/15475441.2024.2342873

Published online: 20 Apr 2024.

LANGUAGE LEARNING AND DEVELOPMENT
https://doi.org/10.1080/15475441.2024.2342873

Semantic Cues Facilitate Structural Generalizations in Artificial Language Learning

Erin Conwell[a] and Jesse Snedeker[b]

[a]Department of Psychology, North Dakota State University, Fargo, ND, USA; [b]Department of Psychology, Harvard University, Cambridge, MA, USA

ABSTRACT

Natural languages contain systematic relationships between verb meaning and verb argument structure. Artificial language learning studies typically remove those relationships and instead pair verb meanings randomly with structures. Adult participants in such studies can detect statistical regularities associated with words in these languages and their use of novel words will adhere to those statistical regularities. However, word use in natural languages is associated with more than distributional statistics. Using an artificial language learning paradigm, we asked how a relationship between verb meaning and sentence structure affected learning and structure generalization. Twenty-four English-speaking adults watched videos described in an artificial language with two possible sentence structures. Half of the participants (statistics-only condition) learned a language with no relationship between verb meaning and sentence structure. The other half (semantics condition) learned a language in which verb meaning predicted which structures a verb occurred in. Although all learners were able to comprehend the learned structures with novel verbs, participants in the semantics conditions made grammaticality judgments and productions with novel verbs that were more consistent with the target language than participants in the statistics-only condition. The availability of semantic cues to verb subcategory supports artificial language learning.

For over 50 years, artificial language learning paradigms have been used to probe language learning processes in people of all ages. Because artificial languages allow researchers to tightly control participants' experience with various aspects of a language, this work has provided a window into the components of language experience that are necessary and/or sufficient for language learning to occur. Such work has successfully demonstrated that learners of all ages can detect statistical and structural regularities in both auditory and visual language-like stimuli. However, the majority of this research has not examined a key component of natural languages: relationships between form and meaning. Because adults can learn artificial languages in the absence of these relationships, we might conclude that form and meaning relationships are unnecessary for language learners, provided that distributional statistics are reliable. Alternatively, learners may use form-meaning relationships to support extension of structures to new lexical items and to reinforce their understanding of how specific words are used. The present study asked how a relationship between verb meaning and sentence structure affected adults' comprehension and use of attested verbs in an artificial language, as well as how adults would use that relationship to guide generalizations about unfamiliar words.

The ability to precisely control learners' experiences makes artificial language tasks ideal for examining how learners can use various sources of information to extract structure from

CONTACT Erin Conwell erin.conwell@ndus.edu Department of Psychology, North Dakota State University, NDSU Dept. 2765, Fargo, ND 58108-6050, USA

language-like stimuli. Findings from artificial language learning studies have indicated that many levels of linguistic structure can be learned at least in part from statistical information alone. In lab settings, infants have shown impressive skill at detecting regularities in the speech stream and using that information to segment words, form distributional categories, detect relationships between adjacent words, and extract abstract patterns (Gómez & Gerken, 1999; Gómez & Maye, 2005; Marcus et al., 1999; Saffran et al., 1996). They can also use statistical information to narrow phonemic categories and to learn phonological alternations (Anderson et al., 2004; White et al., 2008). Adults and children use factors such as co-occurrence probabilities to detect phrasal structure in an artificial language and can do so in both the auditory and visual modalities (Saffran, 2001, 2002). Additionally, adults can use overlapping distributional contexts to draw conclusions about the likely distribution of a new word experienced in only one context (Reeder et al., 2013). All of these findings come from studies of learners' ability to extract patterns from non-referential language-like stimuli. While impressive, this ability is only one part of the language learning process, as true linguistic mastery additionally involves connecting forms with meanings.

A number of studies have extended distributional learning paradigms to address how distributional factors present in natural language might facilitate the acquisition of meaningful structures. Casenhiser and Goldberg (2005) found that verb frequency can affect how well children acquire construction meaning in a learning scenario involving familiar English nouns, novel verbs and a novel construction. Participants who heard one verb several times in this construction and other verbs only once performed better on a comprehension task than those who heard all of the verbs an equal number of times. This finding indicates that an uneven distribution of verbs in constructions supports construction acquisition, which dovetails well with the tendency for a small number of verbs to comprise the majority of instances of a construction in natural language. Casenhiser and Goldberg were not directly concerned with the relationship between verb meaning and constructions as part of the learning process, but rather with how distributional properties of language experience affect learning of construction meaning. Similar studies have found that skew in the number of lexical types used in a structure can affect generalization and production of novel argument structures and noun classes by both adult and child learners (Wonnacott et al., 2012, 2017).

Wonnacott et al. (2008) used an artificial language learning paradigm to demonstrate that adult learners are adept at picking up on the distribution of verbs across constructions and at generalizing that distribution to new verbs. Participants in these studies learned a small artificial language containing two argument structures and were tested on their ability to understand, produce, and judge the grammaticality of utterances containing the verbs used during the learning stage as well as some to which they had only limited exposure. Adults used the distribution of verbs in argument structures to guide their interpretation and production of the language, but they did not tend to over-generalize. When presented with novel verbs, learners used those verbs in a manner consistent with the distribution of verbs in their experience with the language, favoring the structure that they had heard more frequently (See Wonnacott, 2011, for similar results with child learners in a simplified paradigm.).

Further research using similar methods investigated other factors that might affect how adults use novel verbs in a recently learned artificial language. Perek and Goldberg (2015, 2017) showed that adults generalize novel verbs more readily in artificial constructions that alternate due to differences in construction function, as opposed to those that alternate for no reason. In these studies, the two constructions in the artificial language had different meanings, which impacted learners' use of the constructions with novel verbs, but verb meaning was not itself a predictor of argument structure. Thothathiri and Rattinger (2016) also demonstrated that adults can learn a relationship between event type and sentence structure and use that relationship to guide their use of novel verbs. In that study, verb meaning was independent from structure, but the interpretation of a third noun as either an instrument of the verb or a noun modifier depended on which sentence structure was used. Participants successfully used the association between event structure and sentence structure to

guide their descriptions of events, indicating an ability to integrate both distributional and semantic cues in language learning.

The studies described in the preceding three paragraphs all involved artificial languages with a visual reference world. The use of a reference world in these studies allowed researchers to examine a broader range of linguistic competencies than would have been possible using auditory exposure alone. The sentences had meaning and argument order was critical for distinguishing the two structures, which allowed researchers to consider comprehension and production, in addition to the grammaticality judgments usually obtained in artificial language studies that do not involve reference. Additionally, the referential context may have facilitated learning (Moeser & Bregman, 1972; Morgan & Newport, 1981).

Overall, previous work with artificial language learning by adults has revealed that learners are adept at using distributional statistics to discover a wide range of properties of artificial languages. In the particular domain of argument structure, learners use the distribution of verbs in structures to restrict their use of those verbs and to make predictions about the behavior of new verbs. Distribution is a powerful source of information about language structure and verb subcategorization. Furthermore, artificial language studies have shown that learners can leverage a relationship between sentence structures and event structures to support their use of verbs in previously unattested structures. What remains unexamined in an artificial language context is how other cues to subcategorization, such as verb meaning, might impact the learning and generalization of sentence structures. These interactions have been studied primarily in the context of natural language learning.

Verb meaning is a known predictor of argument structure (Levin, 1993; Pinker, 1989) and can contribute to natural language learning in a number of ways. For example, learners can use the argument structure of a verb to determine its meaning (i.e., Syntactic Bootstrapping; Gleitman, 1990; Yuan & Fisher, 2009). Learners might also use verb meaning to determine the meaning of a particular argument structure (i.e., Semantic Bootstrapping; Pinker, 1989). Meaning can help learners restrict their generalization of new verbs to an argument structure (Gropen et al., 1989) and affects acceptability judgments of novel verbs (Ambridge, Pine, & Rowland, 2012).

Children are aware of the relationship between verb semantics and argument structure and use it as a cue for constraining overgeneralizations. For example, in a production task, 7-year-old children used the meaning of a novel dative verb to determine whether that verb occurred in both forms of the dative alternation or whether it was restricted to only one (Gropen et al., 1989). Even younger children were able to use meaning to constrain their use of locative verbs (Gropen et al., 1991). Adults have also been shown to use verb meaning to make judgments about the grammaticality of novel verbs in the dative alternation, but children aged 10 and younger did not use that information in the same task (Ambridge, Pine, & Rowland, 2012). In grammaticality judgment tasks with familiar verbs, adults and older children were influenced by both verb meaning and verb frequency (Ambridge, Pine, & Rowland, 2012), while younger children's grammaticality judgments were influenced by meaning to a lesser degree than they were by frequency (Ambridge et al., 2014).

While such findings indicate that language users are aware of the relationship between verb semantics and argument structure, the use of familiar words and alternations makes it challenging to use these data to address how distribution interacts with meaning in language learning. In particular, the work by Ambridge and colleagues found that children and adults give different weight to distributional and meaning cues when making judgments about word use in a familiar alternation. That might lead us to conclude that distributional cues are used for the first phases of language learning and that meaning is only incorporated into language after a distributional basis has been established (See Culbertson et al., 2017; Stefanowitsch, 2008; Tomasello, 2003; Wonnacott et al., 2008, for discussion of similar ideas.). Alternatively, meaning might operate as a cue to which argument structures a verb is allowed to occur in only when distributional information cannot, as may happen with unfamiliar or low frequency words. Another possibility is that learners use both cues at all times during the process and that the availability of both semantic and distributional cues to verb

subcategorization would increase learners' consistency in their judgment and use of both learned and novel verbs in newly learned argument structures.

A small number of studies have used artificial language paradigms to investigate the learning of lexical classes, specifically noun classes, and have found that semantic information was not generally prioritized by language users. Although these studies are not focused on argument structure learning, they do inform the extent to which learners will use semantic, as opposed to statistical or phonological, cues to make generalizations about membership in a grammatical subclass. Some work on learning of natural noun classification systems has suggested that learners prioritize phonological information over semantic information in learning noun classes, even when phonological information is the less reliable cue (Gagliardi & Lidz, 2014). However, Culbertson and Wilson (2013) demonstrated that adults can learn a noun classifier system that is based solely on semantic properties, indicating that learners are able to use meaning as the basis for classification of words when that is the available cue. Culbertson et al. (2017) argued that learners prioritize other cues over semantic information because meaning is available to learners later than distributional or phonological information and is less salient. These findings together suggest that adult learners may not use meaning to guide verb subcategorization in an artificial language learning task, as long as distributional information is reliable.

Artificial language learning paradigms have mostly focused on statistical cues such as distribution and frequency, although some research on artificial noun classes has incorporated meaning as a cue to subcategorization. However, natural languages include a range of cues to structure. In particular, verb argument structure is closely tied to verb meaning. Although a number of studies on natural language have examined that relationship, those studies were unable control the distribution of particular items or the reliability of cues in the learner's experience. To assess whether language learners can use semantics to learn verb subcategories, this study directly manipulated the relationship between sentence structure and verb meaning in a miniature artificial language learning paradigm with adults.

This study used an artificial language learning paradigm modeled on that used by Wonnacott et al. (2008). Participants watched scenes described in the artificial language over three learning sessions, followed by two testing sessions. The key difference between this study and the first experiment in Wonnacott et al. (2008) was that this study had two between-subjects conditions: one which replicated the previous work by randomly associating verbs with subcategories and one in which verb meaning predicted verb subcategory. If learners do not use meaning to support learning of distributional subcategories of verbs, we would expect no difference in learning between these languages because both languages contain the same amount of distributional information. If adults only use verb meaning as a cue to subcategory when distributional information is absent, we should find that participants in both conditions show similar consistency with familiar verbs, but that learners in the meaning condition use semantic information to guide their use of novel verbs. If meaning and distributional cues are used in tandem, we should expect participants in the meaning condition to be more consistent in their judgment and use of both familiar and novel verbs than those in the distribution-only condition.

Method

Participants

Participants were 24 adults between 18 and 35 years of age, all of whom had English as their first language. Thirteen were female and 11 were male. Participants were paid per session for their participation in the study and received a monetary bonus for completing all five sessions. All five sessions were completed within a two-week span with at least 24 hours between sessions and a maximum of 4 days between consecutive sessions. An additional 3 participants began the study, but did not complete it due to illness or time constraints (2) or failure to comply with instructions (1). Two more participants completed the study, but were excluded

from the final analysis. One of these participants was removed due to experimenter error during the learning phase and the other failed to produce utterances consistent with the language on more than half of the trials in the testing phase. There were 12 participants in each of the two learning conditions. This final sample size was based on approximating the number of participants per condition in Wonnacott et al. (2008; $N = 14$ per condition), while also allowing us to balance verb types across subcategories across participants. Participants were assigned to conditions randomly by alternating conditions as participants joined the study.

Language structure

The language used in this study consisted of 5 nouns, each of which was uniquely associated with a plush animal puppet, and 18 verbs, 12 of which were introduced during the first 3 days of learning and 6 of which were withheld and used as novel verbs during the testing phase of the study. The language had two possible sentence structures: Verb-Agent-Patient (VAP) and Verb-Patient-Agent-ka (VPA-ka). The "ka" marker was purely structural and carried no meaning. In other words, in principle, any scene could be described with either the VAP or the VPA-ka forms. However, in this study, each sublanguage had three distinct subclasses of verbs: six verbs could appear in both structures, another six verbs were restricted to the VAP structure, and the remaining six were restricted to the VPA-ka structure. In the statistics-only condition, verb subclass was unrelated to verb meaning and each semantic subclass contributed an equal number of verbs to each structural subclass. In the semantics condition, the semantic subclass of the verb predicted its structural subclass. In both conditions, we counterbalanced the assignment of specific verbs to syntactic subclasses, such that, across participants, each verb appeared in each subclass. In both learning conditions, two verbs from each structural subclass were withheld during the learning period for use as novel verbs during the testing period.

Procedure

Learning phase: days 1–3

Following Wonnacott et al. (2008), sessions on the first three days of the study had 3 parts. First, participants were directly taught the nouns in the language. Each animal was presented in a still photograph and named once by the experimenter. The images were then presented in a different order and participants were asked to name each animal as it was presented. Participants received feedback on their accuracy and were told the correct name of any animal that they mislabeled.

Next, participants watched brief videos of animal puppets interacting. Each video was accompanied by an audio recording that described the scene using a semantically-appropriate, grammatical sentence from the target language. Participants viewed these videos alone in a quiet room and were encouraged to repeat the sentences aloud to help them learn the language. Participants were told to press the space bar to move to the next video, so progress through the task was self-paced. These videos were presented in a random order in two blocks, each of which lasted approximately 15 minutes. Every video in the training set was presented once on each of the three training days. Two-thirds of the verbs were presented during this phase (4 from each subclass; see the section above on Language Structure), while one-third of the verbs were withheld until the testing phase of the study.

Following the video-watching portion of the session, each participant completed three short tasks similar in format to the tasks that they would complete on days 4 and 5 of the study. These tasks each consisted of 30 trials and only included the verbs used during the familiarization period. The tasks are described in greater detail below. Although performance on these tasks was recorded, those data were not analyzed and participants received no feedback.

Testing phase: days 4 and 5

Each participant's comprehension, production, and grammatical knowledge of the language was tested on days 4 and 5 of the study. On these days, participants reviewed the animal names with the experimenter, receiving feedback for correct responses. No participant incorrectly named an animal on either testing day. Following the animal naming, each participant received a brief "refresher" on the language, in the form of 30 videos randomly selected from the learning phase of the study. After completing that refresher, participants began the full versions of the language processing and production tests. All 18 verbs in the language were included in these tasks on both testing days. Each participant completed the tasks in a random order, with a different order on each of the test days. Participants were told on the test days that they would encounter new words during the tasks and that they should assume that those new words were consistent with the language they had been learning. At no point did participants receive feedback on their performance on these tasks.

Acceptability judgment task

During the grammatical acceptability judgment task, participants viewed videos that were described using one of the two sentence structures in the language. These sentences always contained the correct nouns and verb for the scene. Participants were asked to indicate whether the sentence was "allowed" in the language by pressing one of two keys on a laptop keyboard. Participants were asked to make their decision as quickly as possible without sacrificing accuracy.

Comprehension task

During the comprehension task, participants saw two videos side-by-side and heard one sentence. They were asked to indicate which video was described by the sentence by pressing a laptop keyboard key below the video. One video was always fully consistent with the sentence, while the other served as a foil. The foil video differed from the consistent video in one of three ways: different action (verb comprehension), one different animal (noun comprehension) or reversed roles of the two animals (structure comprehension).

Production task

In the production task, participants saw a video and were asked to describe it using a complete sentence that was "allowed" in the language they were learning. Following Wonnacott et al. (2008), participants were cued with the target verb at the beginning of each video. Participants were recorded during this task and their responses were transcribed offline. Responses were coded for whether the participant produced an utterance that contained the correct words to describe the scene, whether the utterance took one of the two forms permitted in the language, which of those two forms the utterance took, and whether the utterance correctly reflected the roles of the two actors. Responses were further coded for whether the particular verb was "allowed" in the structure that the participant used. For the purposes of analysis, an accurate production had to contain the correct words for the animals and action in the appropriate order to describe the scene, using the form consistent with the verb's subcategory, and the particular order had to be allowed in the language. Sentences in VPA order had to include the "–ka" particle to be counted as correct. Minor mispronunciations of the animal names (e.g., *flumat* instead of *flugat*) were counted as correct.

Stimuli

The visual stimuli in this study consisted of brief (8–10 second) videos of plush animal puppets engaged in simple transitive actions. Those actions could be separated into three distinct categories: brief contact actions (e.g., *hug, kiss, pat*), change of posture (e.g., *turn around, knock over, flip upside down*) and caused motion along a path (e.g., *drag, roll, push*).[1] A female English speaker recorded each of the words in the language in each sentence position and those recordings were spliced together to create the auditory stimuli. The labels for the animals and the actions were consistent across the languages. The only difference between the languages was which verbs were used in each structure.

All stimuli and tests were presented on a laptop computer using EPrime2 software (in the case of the videos) or PowerPoint (in the case of the still images). Because participants' responses triggered the next stimulus, progress through the study was self-paced.

Results

Our analyses examined the effects of condition and of verb familiarity on accuracy for each test type. We used logit mixed model analyses implemented in R (R Core Team, 2015) using lme4 (Bates et al., 2015). Familiarity (familiar vs. novel) and condition (statistics-only vs. semantics) were fixed effects. Although best practice is to use a fully maximal model, including random slopes (Barr et al., 2013), no model with random slopes converged. Therefore, participant and trial were included as random intercepts without random slopes in all analyses. Fixed effects were contrast coded. All data and analysis code are available from the first author upon reasonable request.

Comprehension results

Participants in the statistics-only condition were accurate on 94% of comprehension trials and participants in the semantics condition were accurate on 91% of trials. The rate of correct responses is slightly lower in this sample than the approximately 95% accuracy rate reported by Wonnacott et al.

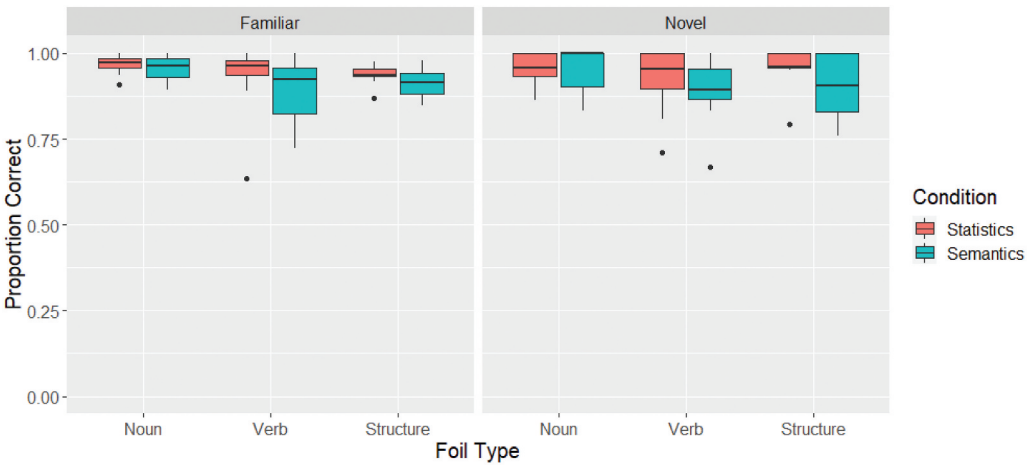

Figure 1. Average proportion correct by participant in the comprehension task.

[1]In the language used by Wonnacott et al. (2008), all verbs described simple transitive events, mostly involving brief contact, although some involved a change in location or posture. Critically, the structural properties of those verbs were completely randomized across participants and meaning never predicted subclass.

Table 1. Results for fixed effects in the mixed model of data from the comprehension task.

Effect	Estimate	Standard Error	z
Intercept	3.049	0.152	20.095***
Condition	0.452	0.267	16.91
Familiarity	0.081	0.152	0.535
Condition x Familiarity	0.133	0.243	0.547

***$p < .001$.

(2008), but still quite high. As shown in Figure 1, there was no main effect of learning condition on comprehension ($z = 1.691$, $p = .091$). The comprehension task also showed no main effect of familiarity ($z = 0.535$, $p = .593$) and no interaction of familiarity and condition ($z = 0.547$, $p = .585$). Participants in both conditions learned to understand the nouns, verbs and structures in this language and could extend that knowledge to understand sentences containing novel verbs. The results of the mixed model analysis of comprehension are presented in Table 1. These findings indicate that participants in both conditions had sufficiently verb-general understanding of the two structures in the language to comprehend both familiar and novel verbs in these structures. Additionally, we considered whether there were meaningful differences in comprehension accuracy depending on the foil type (different nouns, different action, or reversed roles; contrast coded for analysis). We found that a model including a simple effect of foil type was a better fit to the data than a model that did not contain this factor ($\chi^2(2) = 27.12$, $p < .001$). This finding indicates that participants had higher accuracy for comprehending the nouns than for comprehending the verbs and the argument structure. However, including an interaction for foil type in the model did not improve fit relative to the simple effect, suggesting that foil type did not interact with the other factors ($\chi^2(6) = 3.11$, $p = .795$). Therefore, we can conclude that trials on which the foil video contained the wrong nouns were easier for participants than trials involving other kinds of foil videos, but that this did not interact with familiarity or condition.

Grammatical acceptability results

Participants in the semantics condition judged the grammatical acceptability of sentences with greater accuracy than did participants in the statistics-only condition and both groups performed better with familiar than with novel verbs (Figure 2). Participants in the semantics condition correctly judged the

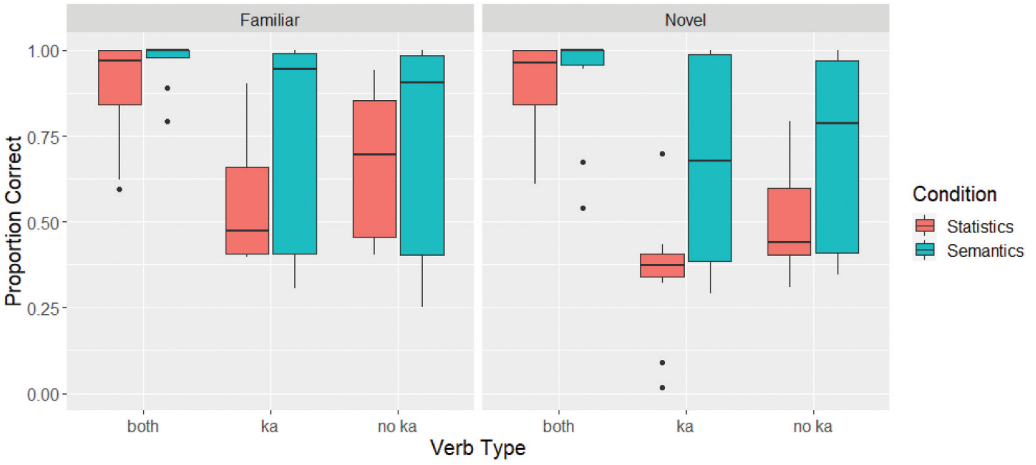

Figure 2. Average proportion correct by participant in the grammaticality task. A "correct" response was defined as either correctly categorizing a sentence as grammatical or as correctly rejecting a sentence as ungrammatical.

Table 2. Results for fixed effects in the mixed model of data from the grammaticality judgment task.

Effect	Estimate	Standard Error	z
Intercept	1.151	0.281	4.093***
Condition	1.545	0.562	2.75**
Familiarity	0.462	0.054	8.66***
Condition x Familiarity	0.37	0.1	3.7***

**$p < .01$; ***$p < .001$.

grammaticality of 76.32% of utterances, while participants in the statistics-only condition correctly judged the grammaticality of 61.5% of utterances. The mixed model analysis found significant main effects of condition ($z = 2.75$, $p = .006$) and of familiarity ($z = 8.66$, $p < .001$). It also found a significant interaction of familiarity and condition ($z = 3.7$, $p < .001$). The complete results of this analysis are shown in Table 2. Post-hoc comparisons showed that participants in the statistics-only condition were less accurate at judging grammaticality for novel verbs than participants in the semantics condition ($t(11) = 3.62$, $p = .002$). The two groups did not differ in the accuracy of their acceptability judgments of familiar verbs ($t(11) = 1.56$, $p = .13$), indicating that the participants in both conditions were comparable to one another in learning the grammatical properties of verbs encountered during the learning phase. Participants in the semantics condition extended the relationship between verb meaning and structural subclass when making judgments about the use of unfamiliar words, while those in the statistics-only condition had no basis for making these judgments. The distributions of grammaticality judgments across verb types and learning conditions are shown in Table 3. Participants in the statistics-only condition were less consistent in their grammaticality judgments than the participants in Experiment 1 of Wonnacott et al. (2008), although the study presented here cannot be directly compared because of differences in the frequency distributions of the verb. However, participants in the semantics condition showed greater systematicity in their judgments.

One possible cause of this higher level of systematicity in the grammatical acceptability judgments of participants in the semantics condition could be the nature of the errors made by participants in the statistics-only condition. Because these participants have no meaningful basis for rejecting a use of a novel verb as ungrammatical, they may simply have accepted all uses of novel verbs. This means that they may be more likely to make "false alarm" errors (incorrectly accepting a use) than "miss" errors (incorrectly rejecting a use). To evaluate this possibility and better understand the nature of the errors that participants made when judging grammaticality, we fit another model adding response type as a factor. In this model, a response of "acceptable" that was accurate would represent a "hit," a response of "acceptable" that was inaccurate would represent a "false alarm," a response of "unacceptable" that was correct would represent a correct rejection, and an incorrect "unacceptable" response would represent a "miss." In this model, learning condition, verb familiarity, and response type (hit, miss, false alarm, correct reject) were all fixed factors, and participant and item were included as random intercepts. The outcome variable was accuracy of response. All fixed effects were contrast coded.

This model found significant main effects of learning condition, verb familiarity, and response type on accuracy, as well as significant interactions of learning condition and verb familiarity and of verb

Table 3. Percentage of utterances judged as grammatical by verb type, sentence type, and learning condition.

Learning Condition	Verb Type	Sentence Type	
		VAP	VPA-ka
Statistics-Only	Alternating	91.75	85.15
	VAP only	96.86	60.37
	VPA-ka only	77.86	88.39
Semantics	Alternating	93.46	95.71
	VAP only	95.69	41.39
	VPA-ka only	45.43	98.26

Table 4. Results for fixed effects in the mixed model including response type (grammatical vs. ungrammatical) of the data from the grammaticality task. Accuracy was the outcome measure.

Effect	Estimate	Standard Error	z
Intercept	1.315	0.262	5.016***
Condition	1.51	0.528	2.861**
Familiarity	−0.708	0.073	−9.704***
Response	0.611	0.08	7.659***
Condition x Familiarity	0.382	0.146	2.614**
Condition x Response	0.121	0.16	0.76
Familiarity x Response	−0.745	0.147	−5.079***
Condition x Familiarity x Response	0.229	0.294	0.779

**$p < .01$; ***$p < .001$.

familiarity and response type, but not of response type and learning condition. The three-way interaction of these factors was also not significant. Complete results of this analysis can be found in Table 4. These findings indicate that participants are more likely to be accurate when they accept a sentence as grammatical than when they reject it as ungrammatical, and that they are more likely to incorrectly accept as grammatical utterances with unfamiliar verbs. The lack of interaction of response type with learning conditions indicates that, while participants in the semantics condition were more accurate overall, this difference was not due to different patterns of "misses" versus "false alarms" between conditions. Of errors made by participants in the statistics-only condition, 88.36% were false alarms, while the errors made by participants in the semantics condition were 91.43% false alarms. The most likely grammaticality judgment error was incorrect acceptance of a sentence as grammatical, regardless of the learning condition.

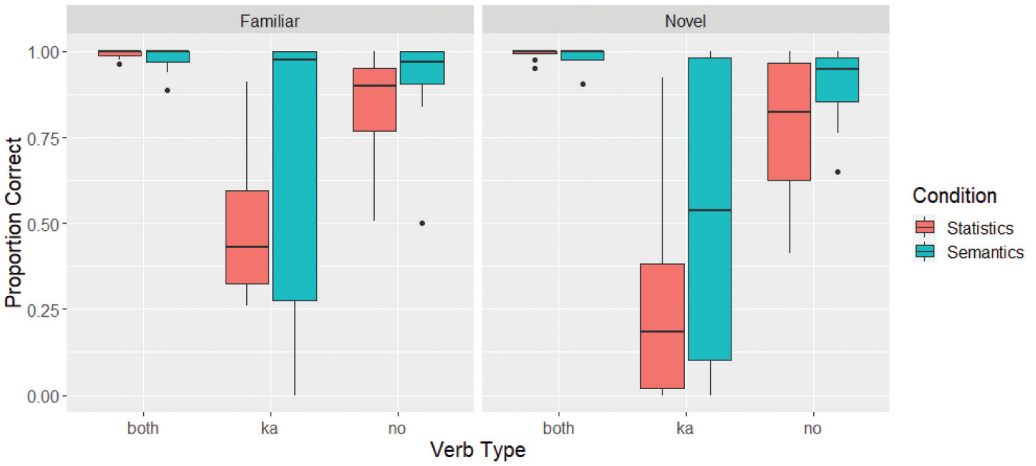

Figure 3. Average proportion correct by participant in the production task. A "correct" production was defined as containing the correct lexical items in an appropriate structure to describe the event. Incorrect productions for sentences with verbs that could be used in either structure contained one or more incorrect nouns.

Table 5. Results for fixed effects in the mixed model of data from the production task.

Effect	Estimate	Standard Error	z
Intercept	1.816	0.296	6.128***
Condition	1.573	0.592	2.655**
Familiarity	0.485	0.064	7.542***
Condition x Familiarity	0.047	0.125	0.706

**$p < .01$; ***$p < .001$.

Production results

The results of the production tests, shown in Figure 3, showed a significant main effect of familiarity ($z = 7.542$, $p < .001$) and a significant main effect of condition ($z = 2.655$, $p = .008$), but no familiarity by condition interaction ($z = 0.378$, $p = .705$). The complete analysis results are in Table 5. These findings showed that all participants were better at producing familiar verbs in appropriate sentence structures, but that participants in the semantics condition used the relationship between verb meaning and verb subcategory to extend their use of novel verbs to the target structures and to bolster their use of familiar verbs as well. Participants in both learning conditions showed a preference for the VAP structure for the alternating verbs (68% for statistics-only; 73% for semantics) and for the VAP only verbs (82% for statistics-only; 92% for semantics). While participants in the semantics condition preferred to produce the VPA-ka structure for those verbs that were only allowed in the VPA-ka structure (57%), participants in the statistics-only condition preferred the VAP structure for VPA-ka only verbs, producing the VAP structure with those verbs 59% of the time. In this case, the statistics-only participants do not mirror the behavior of the participants in Experiment 1 of Wonnacott et al. (2008), who preferred the VPA-ka structure for those verbs that only occurred in that structure. Participants in the statistics-only condition were less able to produce grammatical sentences in the language overall, even for familiar verbs. This difference from the Wonnacott et al. (2008) study may be due to the smaller number of "refresher" sentences that participants were provided with on testing days. Participants in the present study saw only 30 videos on testing days, all at the very beginning of the testing session, while participants in Wonnacott et al. (2008) were re-exposed to 120 sentences from the learning phase divided across three different exposure sessions during the testing days, potentially improving their memory for which words occurred in each structure. These findings indicate that a relationship between meaning and form can support the productive extension of new verbs to appropriate sentence structures and can facilitate participants' memory for which words occur in each structure.

Another question to explore with regard to production is whether the learning condition affected participants' likelihood of producing both structures for those verbs that were allowed to alternate. All but three participants (one in the statistics-only condition and two in the semantics condition) produced both forms for the alternating verbs; those who did not exclusively produced the VAP structure. There was no difference in the percentage of VAP productions between the two learning conditions ($t(22) = 0.444$, $p = .661$). All participants, regardless of learning condition, showed a bias toward the VAP form, producing it 70.5% of the time for alternating verbs, but learning condition did not impact the production of both forms overall. The bias to produce the VAP structure may reflect an effect of the participants' first language (English). The VPA-ka structure, which places the object before the subject and requires a particle, has similarities to the English passive, which is rare in the spontaneous speech of American English users (Roland et al., 2007). Alternatively, it could reflect the well-attested cross-linguistic bias to put subjects before objects in basic word orders.

One might also expect that participants' decisions about the grammatical acceptability of a particular structure with a given verb would affect their choice of form when producing a sentence with that verb. To investigate this issue, we looked at the relationship between the proportion of productions in the VAP form (i.e., without the "ka" particle) and participants' proportion of incorrect "acceptable" grammaticality judgments for each verb subcategory. We fit a mixed effects model with proportion VAP productions as the outcome variable, and verb subcategory, learning condition, and false alarm rate as fixed effects. We also included participant as a random intercept, as the model with by-participants random slopes did not converge. This model showed that false alarm rates were significantly related to the likelihood of producing the VAP structure ($t(132.5) = 3.65$, $p < .001$), Participants consistently produced fewer VAP structures for verbs in the VPA-ka subcategory ($t(137.3) = -5.145$, $p < .001$), and participants with overall higher false alarm rates produced more VAP forms ($t(132.5) = 3.65$, $p < .001$). However, there was no effect of learning condition in this analysis ($t(23.3) = 0.652$, $p = .521$).

Discussion

This study demonstrated that a relationship between verb meaning and syntax facilitates artificial language learning by adults. While participants in both learning conditions were able to comprehend the language they had learned with a high degree of accuracy, those in the semantics condition were better able to produce grammatical utterances in the language. Additionally, adult learners can use such a relationship to guide their use of novel verbs. Learners in the semantics condition used the relationship between verb meaning and syntax to judge the grammatical acceptability of novel verb uses with a higher degree of accuracy and to produce novel verbs in appropriate structures, as compared to participants in the statistics-only condition who had no basis for determining how novel verbs should be used, as each structure was equally likely in their experience. These findings support an account by which statistics and semantics operate in tandem during learning. Meaning-based cues to structure are not used only when statistics are unavailable. Rather, meaning cues support learning overall and improve learners' ability to extend novel verbs to unattested structures.

These findings are consistent with an account of language learning in which verb meaning and syntax are entwined. This idea underlies both the Semantic Bootstrapping (Pinker, 1989) and Syntactic Bootstrapping (Gleitman, 1990) hypotheses regarding natural language learning. Children are known to use the relationship between verb meaning and verb subcategory to make decisions about novel words. Infants use factors such as number of noun phrases to guess the meanings of novel verbs (Arunachalam & Waxman, 2010; Yuan & Fisher, 2009) and older children use verb meaning to determine the limits of that verb's use in particular structures (Ambridge, Pine, & Rowland, 2012; Gropen et al., 1989). However, previous artificial language research has focused almost exclusively on the use of distributional cues in learning with researchers hypothesizing that distributional learning precedes the incorporation of meaning-based cues (e.g., Wonnacott et al., 2008). We have shown here that adults with less than 3 total hours of exposure to a miniature artificial language can use a form-meaning relationship to constrain their uses of novel verbs and to improve language learning overall. This demonstrates that adults will use both meaning and distributional cues in their earliest stages of learning a new language. Learners do not need an extended period of distributional learning before they can incorporate meaning cues.

These findings complement previous work on the learning of structure meanings. Perek and Goldberg (2015, 2017) and Thothathiri and Rattinger (2016) investigated how the meaning of a structure could be learned and extended to new verbs. In those studies, participants were able to learn that a construction could encode the strength of an effect or the interpretation of a third noun without selecting specifically for verb meaning. Constructions affected the interpretation of the verb, but were not restricted in terms of which verbs they could contain. The study presented here indicates that learners can also detect a relationship between form and meaning where particular structures select for semantic subclasses of verbs. While English-learning children have been shown to be sensitive to semantic subclasses in such familiar alternations as the dative (Gropen et al., 1989) and the figure/ground locative (Ambridge, Pine, & Rowland, 2012), participants in those studies had years of experience with the target language, making it impossible to determine whether they had learned a distributional pattern first and then applied meaning information to it or whether distribution and meaning were learned in tandem. By using an artificial language paradigm, we have demonstrated that, in principle, learners can use both cues to verb subcategorization at the earliest stages of learning a language.

In this particular study, the relationship between syntax and verb meaning was somewhat arbitrary, as both structures and all three semantic types described simple agent-patient interactions. Syntactic distinctions between these meaning classes do exist in English (e.g., change-of-posture verbs tend to be used in verb-particle constructions), but those distinctions are not consistent with those made in this study. In this study, the specific relationship between meaning and structure was counterbalanced across participants, so there was no risk of accidentally capturing an unknown natural regularity. This means that participants were not

drawing on an innate relationship between structure and meaning, nor were they relying on a relationship that is familiar to them from English. They learned both the structure and the meaning relationship during the study (see Casenhiser & Goldberg, 2005, for similar findings in children).

These findings do not eliminate the possibility that statistics play a significant role in early language learning. Indeed, participants in the statistics-only condition were able to understand the language just as well as the participants in the semantics condition. They performed above chance in terms of judging the grammaticality of sentences with known verbs and were able to use familiar verbs in an appropriate structure much of the time, as were the participants in Wonnacott et al. (2008). Adults were able to learn about artificial language alternations in the absence of a semantic cue; their learning improved with the addition of the semantic cue.

The biggest differences between conditions surfaced with generalization to new items. Participants in the statistics-only conditions were unable to predict the arbitrary structural subcategories of new verbs, which was expected. They had no grounds for determining what would or would not be allowed. They heard an equal number of instances of each structure and the same number of verb types in each structure. However, the addition of a consistent relationship between verb meaning and structural subclass boosted participants' ability to judge grammaticality and to produce sentences in the language. Unlike their counterparts in the statistics-only condition, participants in the semantics condition were able to make informed decisions about the behavior of unfamiliar words in spite of a relatively brief period of exposure. The presence of the semantic cue provided learners with a basis for generalization. Whether another kind of cue, such as a phonological cue, could also support such generalization is an open question. One might expect that, because natural languages do contain relationships between verb meaning and alternations (Levin, 1993), semantic cues would be preferentially used by learners, as they mirror natural properties of language, whereas relationships between phonological properties of verbs and their argument structures are less widely reported. Empirical research on this issue is warranted.

Another open question is how younger learners would perform in this task and whether the implications of these findings can be extended to language learning by infants and toddlers. All of the tasks that adult learners completed in this study involved overt responses, and the comprehension and grammatical acceptability tasks required that they make a decision between two choices, which may have led participants to try to "figure out" the rules of the languages more explicitly than a young language learner would. The results presented here could, therefore, reflect a level of metalinguistic reasoning that young language learners may not engage in. Indeed, recent research using an artificial noun-particle relationship demonstrated that both adult and child participants showed the highest degree of generalization when they were also explicitly aware of the semantic properties of the relationship (Brown et al., 2022). Children in that study showed both lower awareness of the semantic patterns in the language and reduced levels of generalization. Future work examining the role of metalinguistic awareness in the learning of grammatical alternations would improve our understanding of this factor.

It is also possible that younger learners will differ from adults due to their more limited information processing capacities, which could make some forms of information more usable to learners. Previous work has suggested that learners might use distributional information before semantic information because that distributional information is more accessible at earlier stages of learning (Culbertson et al., 2017; Gagliardi & Lidz, 2014). Developmentally, the cognitive burden of tracking two cues instead of one may limit children's ability to use semantic cues to verb subcategories alongside distributional information. Scaled-down versions of artificial language studies examining verb properties can be conducted with children (e.g., Casenhiser & Goldberg, 2005; Wonnacott, 2011), and future work examining the use of meaning cues by child learners in such studies would be informative regarding whether children can use distribution and meaning as adults have in the present study or whether they rely more heavily on distributional cues. This would indicate whether the results here can be extended to hypotheses about the acquisition of natural languages.

Finally, it is important to consider how meaning and distribution might work together in language learning, particularly with regard to generalization. In the present study, participants in the statistics-only condition had no basis for deciding which forms a novel verb could appear in because they had experienced an even distribution of both structures and verb types. In natural languages, verb types do not occur equally across experiences, nor do alternating structures appear with equal frequency. Prior work on distributional learning of artificial sentence structures has demonstrated a benefit for skewed verb distributions where one verb type is highly frequent relative to other verb types (Casenhiser & Goldberg, 2005; Wonnacott et al., 2012, 2017). Research on learning artificial noun classes has shown that meaning cues to noun class can support generalization, but that type frequency of those classes does not impact generalization (Brown et al., 2022). Our findings show that semantic information boosts generalization of novel verbs to appropriate sentence structures with an even distribution of verb types and sentence structures. How the effect of meaning on generalization would be affected by a skewed distribution of either verb types or sentence structures would be an interesting direction for future research.

Adult learners in an artificial language paradigm learned verb subcategories more successfully when verb meaning predicted structural subclass than they did when statistical information was the only cue to verb subcategorization. They could also extend that meaning-structure relationship to make judgments about permitted uses of novel verbs. These findings suggest that form-meaning contingencies can be detected in the early stages of learning a language and that they complement statistical information for learning about verb argument alternations. This finding is an important first step in mapping out how learners integrate the various cues available during language learning and provides evidence that meaning plays an important role in both in learning specific items and in generalization.

Acknowledgments

This research was funded by NIH T32 MH067556-05, NSF-BCS 0921012 to JS, and NIH P20 GM103505 (Pilot funding to EC, Mark McCourt, PI). The content is solely the responsibility of the authors and does not necessarily reflect the views of the NSF or NIH. We thank the members of the Harvard Laboratory for Developmental Studies for comments on the design and Chelsea Sargent, Garrett Bowen, Matthew Kramer, Rachel Helgeson, Shaun Anderson, and Kathryn Dockter for assistance creating stimuli, running participants and organizing data. We also thank three anonymous reviewers for their constructive comments on a previous version of the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the National Science Foundation, Division of Behavioral and Cognitive Sciences [0921012]; National Institutes of Health [P20 GM103505].

References

Ambridge, B., Pine, J. M., & Rowland, C. F. (2012). Semantics versus statistics in the retreat from locative over-generalization errors. *Cognition*, *123*(2), 260–279. https://doi.org/10.1016/j.cognition.2012.01.002

Ambridge, B., Pine, J. M., Rowland, C. F., & Chang, F. (2012). The roles of verb semantics, entrenchment and morphophonology in the retreat from dative argument-structure overgeneralization errors. *Language*, *88*(1), 45–81. https://doi.org/10.1353/lan.2012.0000

Ambridge, B., Pine, J. M., Rowland, C. F., Freudenthal, D., & Chang, F. (2014). Avoiding dative overgeneralization errors: Semantics, statistics or both? *Language and Cognitive Processes*, *29*(2), 218–243. https://doi.org/10.1080/01690965.2012.738300

Anderson, J. L., Morgan, J. L., & White, K. S. (2004). A statistical basis for speech sound discrimination. *Language and Speech*, *46*(2–3), 155–182. https://doi.org/10.1177/00238309030460020601

Arunachalam, S., & Waxman, S. R. (2010). Meaning from syntax: Evidence from 2-year-olds. *Cognition*, *114*(3), 442–446. https://doi.org/10.1016/j.cognition.2009.10.015

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

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). *Fitting linear mixed-effects models using lme4*. ArXiv e-print, retrieved from http://arxiv.org/abs/1406.5823

Brown, H., Smith, K., Samara, A., & Wonnacott, E. (2022). Semantic cues in language learning: An artificial language study with adult and child learners. *Language, Cognition and Neuroscience*, *37*(4), 509–531. https://doi.org/10.1080/23273798.2021.1995612

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500–508. https://doi.org/10.1111/j.1467-7687.2005.00441.x

Culbertson, J., Gagliardi, A., & Smith, K. (2017). Competition between phonological and semantic cues in noun class learning. *Journal of Memory and Language*, *92*, 343–358. https://doi.org/10.1016/j.jml.2016.08.001

Culbertson, J., & Wilson, C. (2013). Artificial grammar learning of shaped-based noun classification. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual meeting of the cognitive science Society*. Paper presented at the 35th annual meeting of the cognitive science society, Berlin (pp. 2118–2123). Cognitive Science Society.

Gagliardi, A., & Lidz, J. (2014). Statistical insensitivity in the acquisition of Tsez noun classes. *Language*, *90*(1), 58–89. https://doi.org/10.1353/lan.2014.0013

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. https://doi.org/10.1207/s15327817la0101_2

Gomez, R. L., & Gerken, L. (1999). Artificial grammar learning by 1-year-olds leads to specific and abstract knowledge. *Cognition*, *70*, 109–135. https://doi.org/10.1016/S0010-0277(99)00003-7

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy*, *7*(2), 183–206. https://doi.org/10.1207/s15327078in0702_4

Gropen, J., Pinker, S., Hollander, M., & Goldberg, R. (1991). Affectedness and direct objects: The role of lexical semantics in the acquisition of verb argument structure. *Cognition*, *41*(1–3), 153–195. https://doi.org/10.1016/0010-0277(91)90035-3

Gropen, J., Pinker, S., Hollander, M., Goldberg, R., & Wilson, R. (1989). The learnability and acquisition of the dative alternation in English. *Language*, *65*(2), 203–257. https://doi.org/10.2307/415332

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. University of Chicago Press.

Marcus, G. F., Vijayan, S., Bandi Rao, S., & Vishton, P. M. (1999). Rule learning in seven-month-old infants. *Science*, *283*(5398), 77–80. https://doi.org/10.1126/science.283.5398.77

Moeser, S. D., & Bregman, A. S. (1972). The role of reference in the acquisition of a miniature artificial language. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 759–769. https://doi.org/10.1016/S0022-5371(72)80010-0

Morgan, J. L., & Newport, E. L. (1981). The role of constituent structure in the induction of an artificial language. *Journal of Verbal Learning and Verbal Behavior*, *20*(1), 67–85. https://doi.org/10.1016/S0022-5371(81)90312-1

Perek, F., & Goldberg, A. E. (2015). Generalizing beyond the input: The functions of the constructions matter. *Journal of Memory and Language*, *84*, 108–127. https://doi.org/10.1016/j.jml.2015.04.006

Perek, F., & Goldberg, A. E. (2017). Linguistic generalization on the basis of function and constraints on the basis of statistical preemption. *Cognition*, *168*, 276–293. https://doi.org/10.1016/j.cognition.2017.06.019

Pinker, S. (1989). *Learnability and cognition: The acquisition of verb argument structure*. Harvard University Press.

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistics form-classes. *Cognition Psychology*, *66*, 30–54. https://doi.org/10.1016/j.cogpsych.2012.09.001

Roland, D., Dick, F., & Elman, J. L. (2007). Frequency of basic English grammatical structures: A corpus analysis. *Journal of Memory and Language*, *57*(3), 348–379. https://doi.org/10.1016/j.jml.2007.03.002

Saffran, J. R. (2001). Words in a sea of sounds: The output of infant statistical learning. *Cognition*, *81*(2), 149–169. https://doi.org/10.1016/S0010-0277(01)00132-9

Saffran, J. R. (2002). Constraints on statistical language learning. *Journal of Memory and Language*, *47*(1), 172–196. https://doi.org/10.1006/jmla.2001.2839

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Stefanowitsch, A. (2008). Negative evidence and preemption: A constructional approach to ungrammaticality. *Cognitive Linguistics*, *19*(3), 513–531. https://doi.org/10.1515/COGL.2008.020

Thothathiri, M., & Rattinger, M. G. (2016). Acquiring and producing sentences: Whether learners use verb-specific or verb-general information depends on cue validity. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00404

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. *Cognition*, *107*(1), 238–265. https://doi.org/10.1016/j.cognition.2007.11.012

Wonnacott, E. (2011). Balancing generalization and lexical conservatism: An artificial language study with child learners. *Journal of Memory and Language*, *65*(1), 1–14. https://doi.org/10.1016/j.jml.2011.03.001

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

Wonnacott, E., Brown, H., & Nation, K. (2017). Skewing the evidence: The effect of input structure on child and adults learning of lexically based patterns in an artificial language. *Journal of Memory and Language*, *95*, 36–48. https://doi.org/10.1016/j.jml.2017.01.005

Wonnacott, E., Newport, E. L., & Tanenhaus, M. K. (2008). Acquiring and processing verb argument structure: Distributional learning in a miniature language. *Cognitive Psychology*, *56*(3), 165–209. https://doi.org/10.1016/j.cogpsych.2007.04.002

Yuan, S., & Fisher, C. (2009). "Really? She Blicked the Baby?". *Psychological Science*, *20*(5), 619–626. https://doi.org/10.1111/j.1467-9280.2009.02341.x