Investigating stages of morphosyntax acquisition using artificial language learning

Calvin Engström, BA[1,2]; Heidi Getz, PhD[1,2]
[1] Georgetown University, Department of Linguistics; [2] Georgetown University, Department of Neurology

Research on distributional learning has shown that human learners are sensitive to distributional patterns in linguistic input, and they can use these patterns to segment words in speech streams [1], learn phrase boundaries [2], and form word categories [3]. Distributional learning is proposed to play a role in language acquisition, but the exact role it plays is debated. In some nativist theories, distributional learning complements innate abstract representations, providing a means for interpreting input and testing hypotheses about the linguistic structures being learned [4]. In terms of an acquisition trajectory, this perspective predicts continuity: learners begin with abstract representations, and these representations are gradually elaborated to encode knowledge of language-specific patterns. However, an alternative hypothesis suggests that abstract, hierarchical representations are themselves acquired gradually through distributional learning [5-6]. This theory predicts an initial stage in which learners represent concrete language-specific patterns and not an abstract grammar, and a later stage in which learners represent an abstract grammar. **Here we tested these two predicted acquisition trajectories using an artificial language learning paradigm.**

We exposed adults on Prolific (N=30) to a miniature artificial language in two 1-hour sessions on two separate days. The language contained three nouns, two lexical verbs, two adverbs, and one auxiliary desiderative verb ‘to want’ (Table 1). The structure of the language was based on the verb-second (V2) syntax of German [7]. Sentences were generated from an underlying Subject-Adverb-Object-Verb word order using two movement rules: (1) front the Subject, Adverb, or Object and (2) raise the Verb to the second position (Table 2).

We split participants into two conditions and tested their knowledge at different stages in the learning process using a 2AFC test. Participants in the Day 1 condition (N=15) were tested after the first exposure session, and those in the Day 2 condition (N=15) were tested after a second exposure session 24 hours later. In each test trial, participants chose between a grammatical sentence generated by the language’s grammar and an ungrammatical sentence containing a single error (Table 3). There were two categories of test items: *Patterns* items tested knowledge of verb placement and inflection, and *Rules* items tested knowledge of the relative order of words that were not affected by movement rules.

Preliminary results (see Figure 1) suggest a two-stage learning process, consistent with predictions of the distributional learning hypothesis. On Day 1, participants were above chance on the language’s Patterns (Wilcoxon sign-rank test: $Z = 13.58$, $p < 0.001$, one-sided), but they were at chance on Rules ($t(14) = 0.34$, $p = 0.37$, one-sided). This indicates knowledge of concrete patterns without abstract representations of the language’s underlying structure and movement rules. In contrast, on Day 2, participants were now above chance on Rules as well ($t(14) = 2.16$, $p = 0.024$, one-sided).

These early findings suggest that participants’ knowledge initially consists of superficial patterns like verb position and inflection and later develops into fuller representations of the language’s grammar. In ongoing work, we are testing learners after a third exposure session to see if this trend strengthens. If confirmed in additional research, especially work with children, these findings lend support to a theory of morphosyntax acquisition driven by distributional learning in which initial knowledge of concrete, language-specific patterns gradually develops into an abstract grammar.

Part of Speech	Word	English Gloss
Noun	flugit	flugit
	daffin	daffin
	mawg	mawg
Verb	zemper(ka)	hit
	nim(ka)	squish
(Auxiliary)	klidum(ka)	want
Adverb	lapal	slowly
	spad	twice

Table 1. Lexicon of the artificial V2 language. The artificial language contained 8 words: 3 Nouns, 3 Verbs, and 2 Adverbs. "ka" refers to a suffix attached to verbs in the 2nd position. Because the nouns referred to made-up characters participants saw during exposure, they have no separate English glosses.

Stage	Structure	Example
Underlying structure	S-A-O-V-Aux	flugit lapal mawg nim klidum
Fronting	O- -S-A-[ ]-V-Aux	mawg flugit lapal [ ] nim klidum
Verb Raising	O-Aux-S-A-[ ]-V-[ ]	mawg klidum-ka flugit lapal [ ] nim [ ]
Surface structure	**O-Aux-S-A-V**	**mawg klidum-ka flugit lapal nim**
	English gloss:	*Mawg wants flugit slowly to-hit*

Table 2. Sample sentences in the miniature V2 language. The artificial verb-second (V2) language had a transformational grammar based on the syntax of German, a natural V2 language. Surface structures were produced by applying two movement rules to the underlying SOV structure: (1) Fronting of the Subject, Adverb, or Object and (2) Verb Raising of the final verb to the V2 position. In all surface structures, a verb inflected with the morpheme "ka" appeared in the 2nd position in the sentence, and in sentences with the auxiliary verb "want" raised to the 2nd position, the lexical verb remained uninflected at the end of the sentence.

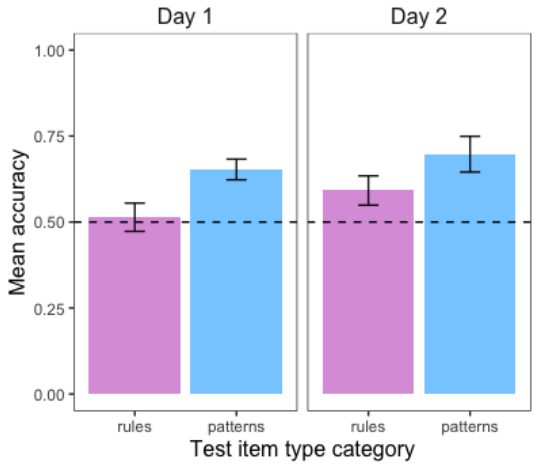

Figure 1. Mean accuracy across all participants on Rules and Patterns test items, with bars representing standard error of the mean. While participants tested on Day 1 and those tested on Day 2 (i.e., after 1 or 2 exposure sessions) performed well on Patterns test items, but only participants tested on Day 2 performed above chance on Rules test items. These results align with a learning trajectory in which early knowledge of surface patterns later allows learners to develop more complex, abstract knowledge of the grammar underlying those patterns

Contrast Type	Grammaticality	
	Foil	Target
V2 patterns (24)		
Verb position (12)	*Vka-S-A-O	S-Vka-A-O
	*Vka-O-S-A	O-Vka-S-A
	*S-Wka-V-A-O	S-Wka-A-O-V
	*A-Wka-V-S-O	A-Wka-S-O-V
Inflection (12)	*S-Vø-A-O	S-Vka-A-O
	*A-V-S-O	A-Vka-S-O
	*S-W-A-O-Vka	S-Wka-A-O-V
	*O-W-S-A-Vka	O-Wka-S-A-V
Rules (24)	*A-Wka-O-S-V	A-Wka-S-O-V
	*O-Vka-A-S	O-Vka-S-A
	*A-Vka-O-S	A-Vka-S-O
	*S-Wka-O-A-V	S-Wka-A-O-V
Embedded Order (8)	*S_m-Wka-[S_e-O-A-V]	S_m-Wka-[S_e-A-O-V]
	*S_m-Wka-[O-A-S_e-V]	S_m-Wka-[S_e-A-O-V]

Table 3. 2AFC test items. Numbers in parentheses indicate the number of trials of each type. On each trial, the two alternatives were identical except for a single error in the ungrammatical sentence.

References: **[1]** Saffran, Aslin, & Newport. (1996). Statistical learning by 8-month-old infants. *Science*. **[2]** Thompson, & Newport. (2007). Statistical learning of syntax: The role of transitional probability. *Language learning and development.* **[3]** Reeder, Newport, & Aslin. (2013). From shared contexts to syntactic categories: The role of distributional information in learning linguistic form-classes. *Cognitive psychology*. **[4]** Lidz, & Gagliardi. (2015). How nature meets nurture: Universal grammar and statistical learning. *Annu. Rev. Linguist.*. **[5]** Aslin, & Newport. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current directions in psychological science*. **[6]** Getz, H. R. (2018). *Sentence first, arguments after: Mechanisms of morphosyntax acquisition*. [Doctoral dissertation, Georgetown University]. **[7]** Holmberg, A. (2015). Verb second. *Syntax–theory and analysis: An international handbook*.