Hierarchical Bias: Domain-General or Language-Specific?

Kaitlyn Harrigan, Sadhwi Srinivas, Aidan Burnham, and Nicholas Voivoda

1. Introduction

The "generativist view" of linguistics posits that humans bring certain innate constraints and biases to the task of learning a language. These constraints and biases are moreover alleged to be language-specific—they cannot be effectively reduced to those exhibited in other cognitive domains, in performing non-linguistic tasks (e.g., Chomsky 1975, Chomsky 1980, Yang 2006). Empiricist-skewing views oppose this framework (e.g., Bybee 2006, Foley & Van Valin 1984, Goldberg 1995, Halliday 1994, Lakoff 1977, Langacker 1987, Van Valin 1993, Tomasello 2003). In many cases, objections to the generativist viewpoint do not resist the claim that the learner may bring certain constraints or strategies to the process of language-learning; however, these biases are considered unlikely to be language-specific. They are thought instead to take the form of very general learning mechanisms that manifest across various cognitive domains and operate over the input in each domain.

The goal of the current work is to make a specific contribution to this important, long-standing debate by systematically testing: (i) whether predispositions towards inferring a specific, purportedly universal, linguistically attested pattern over an unattested one can be robustly observed within a controlled language-learning task, and (ii) whether such predispositions are specific to language learning tasks or if they are also similarly exhibited in analogous non-linguistic, puzzle-solving tasks. We focus on the specific linguistically attested pattern of hierarchical phrase structure, though more broadly, the current study is intended to serve as a methodological proof-of-concept easily extensible to testing the existence and strength of other kinds of inductive biases in linguistic and non-linguistic learning tasks. We chose hierarchical phrase structure due to the near-consensus with which it is accepted by linguists of all leanings to be "universal" across languages, as well as the relative ease with which it may be instantiated in a non-linguistic, general puzzle-solving task, in addition to a language learning one.[1]

*Kaitlyn Harrigan, William & Mary, kharrigan@wm.edu; Sadhwi Srinivas, William & Mary, ssrinivas@wm.edu; Aidan Burnham, William & Mary, aeburnham@wm.edu; Nicholas Voivoda, William & Mary, nwvoivoda@wm.edu.

[1] Some other candidates for language universals, and consequently for being a part of UG, include (but are not limited to): locality constraints in long-range dependency relations, recursion/center-embedding structures, predicate-internal subjects, lexical categories such as nouns and verbs, closed-class grammatical morphemes such as negation/case/modality etc. (Pinker 1994, cf. Tomasello 1995).

Proceedings of the 50th annual Boston University Conference on Language Development, ed. Romi Hill et al., 309-322. Somerville, MA: Cascadilla Press.

2. Background

2.1. Evidence for a hierarchical bias in language learning

A central claim within generative linguistics is that language learners consistently infer hierarchical rules, even when such rules are not uniquely determined by the input. A frequently discussed example of this *hierarchical bias* is the rule for polar yes/no questions in English, first noted by Chomsky as an illustration of the Poverty of the Stimulus (Chomsky 1975, 1980). Consider examples (1)-(2) below. The rule for generating the yes/no question in (1b) from the assertion in (1a) can be stated as a linear generalization ("move the first auxiliary to the front") or as a hierarchical one ("move the main clause auxiliary to the front"). However, (2) clarifies that only the hierarchical generalization is successful (2b); attempting to front the linearly precedent auxiliary results in an ungrammatical form (2c). English-learning children to seldom receive disambiguating data points such as (2) in their input, and yet apparently never produce ungrammatical utterances like (2c) (Crain & Nakayama 1987). This has led to the influential hypothesis that infants have an innate bias for hierarchical generalizations – that is, they assume hierarchical structure when deducing language rules.

(1) a. The man <u>is</u> a fool.
b. <u>Is</u> the man a fool?

(2) a. The man who is a fool <u>is</u> amusing.
b. <u>Is</u> the man who is a fool amusing?
c. *Is the man who a fool <u>is</u> amusing?

Empirical support from real-time language learning studies demonstrating the preference for hierarchical over non-hierarchical generalizations has been somewhat limited, however. This is most directly tested in Smith et al. (1993), where a conlang "Epun" was used to examine the learnability of various rules by a language savant, "Christopher," as well as four undergraduate controls. One of the rules in Epun was a "structure-independent" rule, in which an emphatic morpheme *nog* was governed by an "arithmetically-determined" computation rather than a hierarchical structurally-determined computation. Specifically, this emphatic element always attached to the orthographic word in the third position of any sentence, regardless of the lexical category of the word or the constituency structure of the sentence – as seen in (3).

(3) a. *Fa zaddil-in ha-bol-u-**nog** guv.*
The man-NOM PAST-go-3MS-EMPH yesterday
"The man did go yesterday." (Smith et al. 1993, ex. 42a)
b. *Lodon-in ha-bol-u guv-**nog**.*
Lodon-NOM PAST-go-3MS yesterday-EMPH
"Lodon did go yesterday." (Smith et al. 1993, ex. 42b)

Smith et al. report that the rule for *nog* was not successfully learned, either by Christopher or the four undergraduate controls, although no quantitative results are provided. While this result is potentially highly relevant to question of the generativist debate described above, we believe its strength is limited for several reasons. First, the low number of participants and lack of quantitative data make it difficult to gauge if this result truly generalizes across all contexts of language learning. Second, the structure-independent rule in this study was imposed on a grammar that was underlyingly hierarchically-structured, as most other rules in *Epun* were defined in a structure-dependent way. It is possible that participants were resistant to generalizing a non-hierarchical rule in the context of an otherwise hierarchically-structured grammar; perhaps they would be successful at generalizing such a rule within a language where all rules are similarly arithmetically-computed. A strong version of the hypothesis that human learners only entertain hierarchical structure-dependent rules within a language learning context would predict that structure-independent rules are more widely unlearnable, and not just as part of a system that otherwise contains hierarchical structure-dependent rules.

A different type of evidence for the robustness of hierarchical generalizations in language learning comes from Takahashi & Lidz (2007) and Takahashi (2009), who demonstrate using a series of artificial language learning tasks that adults and children are able to generalize beyond the training stimuli to novel structures based on evidence of constituency-based transformations within a hierarchical grammar. From this result, these authors conclude that learning based on input statistics cannot be the sole mechanism by which grammaticality of structures is deduced. Instead, statistical learning must interact with inherent expectations about the types of constituents that might exist (i.e., hierarchically structured ones) and the types of transformations they are able to undergo.

Overall, a survey of the literature reveals relatively weak experimental evidence for the existence of a hierarchical bias—a surprising state-of-affairs given how influential the Poverty of the Stimulus argument based on English auxiliary-fronting has been to the generativist program (see also Futrell & Mahowald 2025 who make a similar point). In particular, the inability of human language learners to pick up non-hierarchical, structure-independent rules— as claimed in Smith et al. (1993)— remains to be more directly and compellingly established. Filling this gap using a systematic, controlled experimental design forms the first main goal of the current study.

2.2. Evidence for a hierarchical bias *outside* of language learning

If it is indeed the case that language learners are predisposed to making hierarchical generalizations over non-hierarchical ones, the question arises if such a predisposition is limited to language-learning contexts or whether it is a more domain-general bias. The existence of language-specific inductive biases — of which the hierarchical bias may well be one — is key to the generative view. Nevertheless, a perusal of the existing literature reveals there may be reason to

believe that a preference for hierarchical structure-based generalizations is at play not only within the language-learning domain, but in other domains of human cognition as well — including motor action, music, or mathematics (Coopmans 2023, Dehaene et al. 2015). While such domain-generality of the hierarchical bias would seem to be consistent with cognitivist ideas that uphold a complete absence of language-specific innate learning biases, we contend that a conclusion along these lines is somewhat premature in the absence of results from tasks that are directly comparable across linguistic and non-linguistic domains. Behavioral evidence directly comparing pattern learning/processing across domains remains sparse.

Accordingly, a second goal of the current study is to directly compare the strength of a possible hierarchical bias within a language learning task and a more general (non-linguistic), puzzle-solving task. If the tendency for hierarchical generalization is observed to the same degree across both tasks, this could be interpreted as evidence towards the truly domain-general nature of the hierarchical bias. On the other hand, if we observe a difference between the linguistic and non-linguistic tasks in the strength of the observed hierarchical bias, this would point to participants treating the exact same evidence differently depending on whether they expected to be learning a language or to be solving a more general pattern-finding puzzle. Such a result would seem to suggest domain-specific interactions with a potentially domain-general bias, contributing significant nuance to the old generativist *vs.* functionalist/cognitivist debate in language acquisition.

2.3. The current study

To address our first question of interest, which seeks to clearly and directly establish the existence of a *hierarchical bias* in real-time language-learning, we make use of an Artificial Language Learning (ALL) task that builds closely on Takahashi & Lidz (2007), who exposed adult participants to constructed, hierarchically structured grammars over a series of learning sessions. In one of our study conditions, we employ a similar hierarchically structured, constructed grammar as a way to replicate a part of what these authors found—that hierarchical structures are indeed learnable within a language learning context. In addition to a hierarchical grammar, we also employ an additional non-hierarchical grammar in this task, where the rules are determined based on the sequential order of the words rather than a nested hierarchical structure. A strong interpretation of Smith et al. (1993) predicts that the hierarchically-structured grammar will be learned more successfully than the non-hierarchical one, which may not be learned at all. We further extend this ALL methodology to a non-linguistic, puzzle-solving context, allowing us to compare the strength of the potential hierarchical biases observed within a language-learning task *vs.* a non-linguistic task. Such a comparison is essential to determine if the preference to learn hierarchical patterns is truly domain-general, or if it varies from one learning domain to another.

3. Experiment: Pattern deduction in and outside of language contexts

Participants completed a pattern deduction task consisting of interleaved training and test phases. During training phases, they were exposed to a series of sequences of words or shapes. During the test phase, participants used this knowledge to identify whether novel strings were consistent with the pattern they had seen during the training.

3.1. Participants

127 undergraduates were recruited to participate in our study from William & Mary's introductory Linguistics and Psychology courses. All materials were approved by William & Mary's Institutional Review Board. All participants gave informed consent and received course credit for their participation in the study.

3.2. Design

The study employed a 2x2 between-subjects design. Participants were tasked with "escaping from a virtual escape room" by deducing one of two possible constituency *patterns* (*hierarchical* v. *linear*) in one of two possible learning *contexts* (*language* v. *lockbox*). The goal here was to test if there are any differences in pattern learning abilities for linguistically attested and unattested grammars, both within and outside of language-learning tasks.

Contexts. The *language context* was set up along the lines of a classic artificial language learning study. Participants were informed that the "escape room" was patrolled by guards who spoke a novel language named "Nog". The guards communicated with each other by slipping notes under the doors written in Nog. To "make their escape", participants were required to trick the guards into thinking they too are speakers of this language, by similarly slipping them notes containing Nog sentences. To prepare for this, participants were shown some examples of Nog sentences from which the basic vocabulary and sentential patterns of the language could be inferred. Specifically, participants were exposed to sentences that were consistent with either a linguistically attested, hierarchically structured constituency structure (i.e., a *hierarchical pattern*), or a systematic but linguistically unattested structure consisting of linearly alternating constituents (a *linear pattern*). Participants were asked to pay attention to the sequences being shown to them but were told to avoid any explicit memorization. Interleaved throughout the study were test phases, where participants were asked to use their inferred knowledge of the language from the examples seen during the exposure phases to choose between novel sequences in a forced-choice task. In particular, they were asked to choose the sequence that they thought would be a better note to pass under the door to a Nog-speaking guard.

The *lockbox context* was similar to the *language context*, except that instead of "learning a language" to communicate with guards, participants were informed that they would need to solve a set of combination locks to unlock a series of

doors and lockboxes in order to escape from the escape room. The training and test phases were set up identically as before. Participants were shown sequences that were consistent with one of the same two possible *patterns*. Throughout the study, participants were asked to use their knowledge of possible combinations based on the examples they saw during the exposure phases to choose between novel sequences in a forced choice task.

Patterns. The linguistically attested *hierarchical pattern* was a hierarchically organized "grammar" consisting of three phrasal constituents (Figure 1), based on the stimuli used in Takahashi & Lidz (2007). Each constituent was composed of either two or three consecutive words (in the *language context*) or shapes (in the *lockbox context*). This grammar included hierarchical embedding, with one phrase (CP) nested inside another phrase (EP), resembling a simplified version of natural language wherein constituents are often consecutive strings of words with internal hierarchical structure and over which transformations occur. Unlike natural language, however, the *hierarchical pattern* did not include any optional elements.

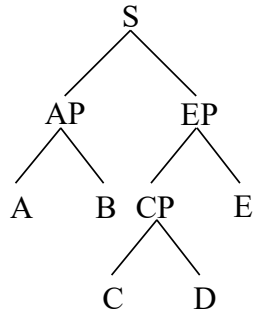

Figure 1. Constituents of the *hierarchical pattern*.

Thus, basic, non-transformed sequences generated by this grammar consisted necessarily of five positions (akin to lexical categories). Each position was mapped to two possible vocabulary items (words or shapes depending on the *context*), resulting in a total of 32 grammatical basic sequences (Table 1).

Table 1. Vocabulary items for *language* and *lockbox* contexts.

		A	B	C	D	E
language	option 1	ksėm	hmoswæd	prürris	qălhom	pyllŭq
	option 2	vdimrık	zlėr	vbek	rryn	kfas
shapes	option 1					
	option 2					

Further grammatical sequences could be generated by subjecting the basic sequences to two types of constituency-based transformations commonly observed in natural languages: (i) deletion (obtained by simply omitting a constituent), and (ii) movement (obtained by moving a constituent to the left or front of all other constituents). Table 2 shows examples of such transformed sequences. Note that transformed movement sequences for [AB] were excluded, as this constituent already occurs at the beginning of the basic sequences.

Table 2. Transformed sequences in the *hierarchical pattern*.

	constituent	sequence	examples
deletion	AB	CDE	prürris rryn kfas 
	CD	ABE	ksėm hmoswæd kfas 
	CDE	AB	ksėm zlėr 
movement	~~AB~~	~~ABCDE~~	
	CD	[CD]ABE	prürris rryn ksėm zlėr pyllŭq 
	CDE	[CDE]AB	vbek qălhom pyllŭq ksėm zlėr 

The alternative, linguistically unattested pattern (the *linear pattern*) was systematic and also utilized transformations over "constituents," except in this case, constituents contained non-consecutive, linearly alternating elements (ACE and BD; Figure 2).

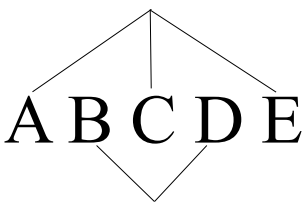

Figure 2. Constituents of the *linear pattern*

The basic sequences generated under this pattern were identical to those within the *hierarchical pattern*; however, constituency-based transformations resulted in a different set of grammatical sequences. Here as well, constituents underwent either deletion or movement transformations. In sequences obtained

through deletion, one of the two constituents was omitted. In sequences obtained through constituent movement, one of the two constituents was moved to the beginning of the sentence. Table 3 shows examples of the transformed sequences under the *linear pattern*.

Table 3. Transformed sequences in the *linear pattern*.

	constituent	sequence	examples
deletion	ACE	BD	zlėr rryn 
	BD	ACE	ksėm vbek pyllŭq 
movement	ACE	[ACE]BD	vdimrık vbek pyllŭq hmoswæd rryn 
	BD	[BD]ACE	zlėr rryn ksėm prürris kfas 

3.3. Procedure

Each participant was assigned to either the *hierarchical* or the *linear pattern* within either the *language* or the *lockbox context*. The study was deployed on a lab computer via Qualtrics and guided participants through multiple training and test phases. In particular, participants were first exposed to (and tested on their knowledge of) basic sequences, which identical across the *hierarchical* and *linear patterns*. This was followed by exposure to transformed sequences, which disambiguate the *hierarchical pattern* from the *linear pattern*.

Phase 1: Vocabulary. Participants were exposed to all 32 possible basic sequences in an initial exposure phase. They then completed eight vocabulary check questions, wherein they chose which of two sequences was one they had been exposed to, testing their ability to distinguish basic grammatical sequences from ungrammatical, scrambled foils. This vocabulary exposure was then repeated, once again followed by eight vocabulary check questions. Thus, participants saw a total of 64 example sequences (each possible combination twice) and responded to 16 total vocabulary check questions.

Phase 2: Transformations. Participants then moved on to the second type of exposure phase, where they were shown examples of transformed sequences. In the *language context*, they were informed that the Nog-speaking guards had begun to suspect the possibility of an imposter among them, and participants would need to pass more sophisticated messages in order to avoid getting caught. Thus, during this phase, they would be shown examples of more complex Nog sentences. In the *lockbox context*, participants were informed that as they moved through the escape room, they would need to solve more sophisticated combination locks than

before, and that they would therefore begin to see more complex sequences to practice. In this phase, participants were first exposed to a set of 74 representative transformed sentences (containing examples with both deleted and moved constituents). Participants then completed eight test questions, where they chose between *novel* grammatical transformed sequences and ungrammatical foils. In this phase, foils consisted of transformations over non-constituents. The transformation exposure was then repeated where participants saw the same 74 sequences as before, followed by eight new test questions. Thus, each participant saw a total of 148 training sequences (74 repeated twice) and responded to 16 forced-choice questions (5 deletion, 11 movement).

3.4. Results

We excluded participants who chose the grammatical string at a rate less than 75% in the vocabulary check phase (4 participants excluded in the *language context*, 3 in the *lockbox context*). This left a total of 120 participants for the final analysis—30 per condition. Overall, we find that participants learned both the *hierarchical* and the *linear patterns* above chance across the two learning contexts (*language* and the *lockbox*); see Figure 3.

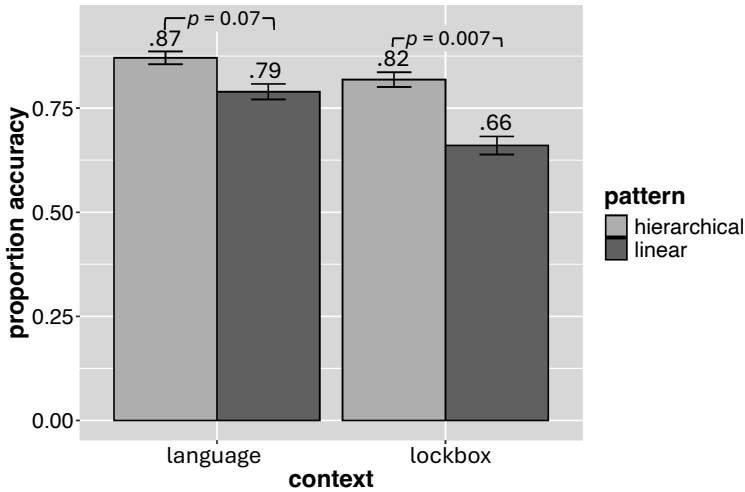

Figure 3. Accuracies across *hierarchical* and *linear patterns* in *language* and *lockbox contexts.*

The results were analyzed in R (R Core Team 2021, RStudio Team 2020) using a binary logistic regression model with *accuracy* as the dependent variable, *pattern* (*hierarchical* v. *linear*) and *context* (*language* v. *lockbox*) as fixed effects, and *subject* as a blocking variable. Results revealed a marginally significant main effect of *pattern* [$X^2_{(1)} = 3.13$, $p = 0.077$] but not *context* [$X^2_{(1)} = 1.70$, $p = 0.192$],

indicating an overall preference for learning the *hierarchical pattern* compared to the *linear pattern* across the *language* and *lockbox contexts*. Further pairwise comparisons reveal that such a preference was stronger in the *lockbox context* than the *language context* (Table 4).

Table 4. Pairwise comparisons for *language* and *lockbox contexts*.

context	***pattern***	**estimated marginal means**	***p*-value**
language	*hierarchical*	0.93	0.077
	linear	0.86	
lockbox	*hierarchical*	0.88	0.008
	linear	0.72	

We also looked at pairwise comparisons comparing the learning of each *pattern* across both *contexts* (Table 5). We find that participants are equally good at learning the *hierarchical pattern* regardless of *context*, but surprisingly, they are better at learning the *linear pattern* in the *language context* than in the *lockbox context*.

Table 5. Pairwise comparisons for *hierarchical* and *linear patterns*.

context	***pattern***	**estimated marginal means**	***p*-value**
hierarchical	*language*	0.93	0.192
	lockbox	0.86	
linear	*language*	0.88	0.028
	lockbox	0.72	

Overall, we find that participants learn both *patterns* in both *contexts*, although we find differences in the accuracy of their performance across different combinations of factors.

4. Discussion

4.1. Summary of findings

The current study was designed with two goals in mind. First, we sought to verify and strengthen the claim about the learnability of language rules made in Smith et al. (1993), that while human language learners are able to generalize structure-dependent (hierarchical) patterns, they struggle to do so with structure-independent rules (e.g., a rule based on the arithmetically-computed positions of elements in a sentence). Such a finding would strongly establish the presence of a hierarchical bias in human language learning. Our second research goal was to test whether the hierarchical bias, if one exists, is limited to language-learning contexts—as expected under the traditional generativist view, or whether it can be observed more generally across cognitive tasks and domains.

First, we found evidence of successful learning by participants in all four conditions, including the linear grammar in the language learning task. This

finding is unexpected in light of Smith et al. (1993) where adult language learners had reportedly failed to learn a structure-independent rule, and undermines the strong hypothesis that structure-independent rules are unattested in language because they are impossible to learn. It is possible that the observed unlearnability of the structure-independent rule in Smith et al.'s study was due to the rule being juxtaposed on top of an underlying structure-dependent grammar. By contrast, in the current study, participants were working with a language-unattested, truly structure-independent grammar, which may have contributed to their successful learning of this grammar.

Note, however, that the learnability of the language-unattested linear grammar does not necessarily entail the lack of a hierarchical bias in language learning. Such a bias may still be said to exist if the hierarchical structure-dependent rule is learned to a substantially higher degree than the structure-independent rule given the same amount of input within the same learning context. This is in fact what we observed. Participants exposed to the hierarchical grammar in the language-learning context enjoyed a learning advantage over their counterparts exposed to the linear grammar, indicating the presence of at least a weak hierarchical bias in this context.

We move on next to our second question of interest: whether the hierarchical bias is specific to language learning, or whether it appears more domain-generally. Our study participants once again exhibited successful learning of both types of patterns (hierarchical and linear) within the non-linguistic *lockbox context*. Furthermore, the *hierarchical pattern* was once again learned robustly better than the *linear pattern* — in fact, to an even stronger degree than in the language-learning task. Specifically, participants were worse at generalizing the *linear pattern* in the *lockbox context* when compared to the *language context*; the hierarchical pattern itself was learned comparably well across both contexts. We thus find evidence for the domain-generality of a hierarchical bias, in line with conclusions from other studies in the recent literature (Coopmans 2023, Dehaene et al. 2015, Fitch 2014).

4.2. Success in the *linear language* condition

The finding that a hierarchical bias is domain general is not surprising in light of previous theories about the role of hierarchical patterns and human cognition. It was unexpected, however, for this effect to be stronger in the *lockbox* than the *language context*. We find that this was driven not by an advantage to learning hierarchical patterns in the *lockbox context*, but rather by the learning of our non-linguistically attested pattern being unexpectedly *good* in the *language context*.

One possible explanation for this surprising finding is that of surface-level methodological differences between the two contexts. One specific contributor may have been the "readability" of the items in the *language context*. In order to reinforce the language status of the sequences in the *language context*, we created words using letters based on the Latin alphabet. Although we used some sequences not consistent with English phonotactics and introduced a handful of

non-English characters, many of the words could be approximately pronounced (see Table 1). This may have given our English-speaking participants a heuristic that made it easier to keep track of sequences, leading to an overall advantage for the both *patterns* within the *language context*. Participants may have rehearsed the language sequences via the phonological loop—an aspect of working memory that helps the listener or reader store acoustic information over a short period of time (Baddeley et al. 1998). No such strategy was available to participants in the *lockbox context*.

An ongoing follow-up study controls for the effect of the readability heuristic in our *language context*. In this new study, we use the same patterns and the same set-up as our current *language context*, but use an alternative non-Latin writing system instead of using words based on the Latin alphabet. Such a system should be recognizable by English speakers as language, but would not allow for the use of a heuristic based on phonological rehearsal. Preliminary results on this follow-up show a strong hierarchical bias (equivalent to the *lockbox condition* in the current study). This suggests that participants in the *language context* condition of the current study are benefiting from the readability of the sequences, and therefore the true hierarchical bias is being masked in the current *language* condition.

An additional factor could have contributed to participants' success at learning the linear grammar within the language learning context. It is possible that the type of non-adjacent dependency instantiated in our linear grammar is not, in fact, unlike anything that occurs in natural languages. It may be instead approximating one or more types of linguistically-attested patterns more closely than was intended. That is, while non-adjacent constituents are linguistically impossible (or more accurately, hereto unattested), other types of non-adjacent, long-distance phenomena are of course known to occur in the grammar. Examples of this include long-distance filler-gap dependencies, assimilation processes such as vowel harmony, binding phenomena, etc. Participants in the study may have identified the structures formed within our linear grammar with such long-distance dependencies, and moreover, that familiarity with such long-distance dependencies specifically within language made it possible for participants to learn the linear grammar more easily within the language task compared to the non-linguistic, lockbox-solving task.

Given this possibility, the question arises of what it would take to instantiate a pattern that we can be more fully certain is not linguistically attested. One such possibility is the arithmetically-computed rule in Smith et al. (1993), where a morpheme was always affixed to the linearly third word in a sentence. However, it is worth once again recalling here that in their study, such a rule was formulated on top of an otherwise hierarchically-structured grammar. It is more difficult to imagine a completely structure-independent grammar made up solely of linear rules of the sort employed by Smith et al. Moreover, the current study aimed to directly compare the learnability of linguistically-attested and unattested grammars within the same learning context, and thus required that the overall complexity of the two grammars as well as their vocabulary sizes and lower-level

statistics be roughly matched. It remains an open question whether a more convincingly language-unattested pattern can be feasibly instantiated within the current design.

Though our findings indicate that the hierarchical bias is unlikely to be an instance of a language-specific bias, this does not by any means rule out the possibility of other language-specific biases. It is our hope that further progress can be made toward identifying the existence of any possible language-specific biases by extending the methodology described in this article to testing other candidate phenomena contested (to varying degrees) to be language-specific—for instance, the learning of recursive, center-embedded structures, or locality constraints in long-range dependency relationships.

5. Conclusion

Our work provides evidence that there is an advantage for the structure-dependent patterns, although non-hierarchical, structure-independent patterns are learnable within a language context (contrary to previous claims as in Smith et al. 1993 or Yang 2006). This advantage is also observed in non-linguistic tasks to a stronger degree than in the language task, not due to a difference in how well the hierarchical pattern was learned across the learning contexts, but rather due to a discrepancy in the learning of the linear pattern. Specifically, we found that the hierarchical as well as the linear pattern was more effectively learned in the language task. Ongoing work explores the role of the readability of our *language* stimuli in this effect. The current study also introduces a novel methodology to test how well various types of patterns are learned in language and non-language contexts. Such comparisons are essential to settle a central debate in modern linguistics over whether or not there exist language-specific innate biases that learners bring to the task of language learning.

References

Baddeley, Alan, Susan Gathercole, & Costanza Papagno. 1998. The phonological loop as a language learning device. *Psychological Review, 105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language, 82*(4), 711–733. https://doi.org/10.1353/lan.2006.0186

Chomsky, Noam. 1975. Questions of form and interpretation. In R. Austerlitz (Ed.), *The Scope of American Linguistics: Papers of the First Golden Anniversary Symposium of the Linguistic Society of America, held at the University of Massachusetts, Amherst, on July 24 and 25, 1974* (pp. 159–196). De Gruyter Mouton. https://doi.org/10.1515/9783110857610-008

Chomsky, Noam. 1980. *Rules and Representations*. Oxford, England: Basil Blackwell.

Coopmans, Cas W. 2023. *Triangles in the brain: The role of hierarchical structure in language use*. PhD Thesis, Radboud University Nijmegen, Nijmegen.

Dehaene, Stanislas, Florent Meyniel, Catherine Wacongne, Liping Wang, & Christopher Pallier. 2015. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2-19.

Fitch, W. Tecumseh. 2014. Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews, 11*, 329–364.

Foley, William A., & Robert D. Van Valin. 1984. *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Futrell, Richard, & Kyle Mahowald. 2025. How Linguistics Learned to Stop Worrying and Love the Language Models. *arXiv preprint arXiv:2501.17047*.

Goldberg, Adele. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Halliday, Michael Alexander Kirkwood. 1994. *An introduction to functional grammar* (2nd ed.). London: Edward Arnold.

Lakoff, Robin. 1977. What you can do with words: Politeness, pragmatics and performatives. In *Proceedings of the Texas Conference on Performatives, Presuppositions and Implicatures* (pp. 79–106). Arlington, VA: Center for Applied Linguistics.

Langacker, Ronald W. 1987. *Foundations of cognitive grammar* (Vol. 1). Stanford, CA: Stanford University Press.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.r-project.org/

RStudio Team. 2020. RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Smith, Neil V., Ianthi-Maria Tsimpli, & Jamal Ouhalla. 1993. Learning the impossible: The acquisition of possible and impossible languages by a polyglot savant. *Lingua, 91*(4), 279–347.

Takahashi, Eri, & Jeffrey Lidz. 2007. Beyond statistical learning in syntax. In A. Gavarró & M. J. Freitas (Eds.), *Proceedings of GALA 2007: Language acquisition and development* (pp. 446–456). Cambridge Scholars Publishing.

Tomasello, Michael. 2003. *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Van Valin, Robert D., Jr. 1993. A synopsis of role and reference grammar. In R. D. Van Valin, Jr (ed.), Advances in Role and Reference Grammar (pp. 1–164). Amsterdam and Philadelphia: John Benjamins

Yang, Charles. 2006. *The infinite gift: How children learn and unlearn the languages of the world*. New York: Scribner (Simon & Schuster).

Proceedings of the 50th annual Boston University Conference on Language Development

edited by Romi Hill, Xinhan Jiang, Danutham Worapipat, and Aditya Yedetore

Cascadilla Press Somerville, MA 2026

ISSN 1080-692X
ISBN 978-1-57473-047-0 (2 volume set, paperback)

Ordering information

To order a copy of the proceedings or to place a standing order, contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, sales@cascadilla.com, www.cascadilla.com