Linguistically attested, nested constituency structures are preferentially learned across learning domains

Kaitlyn Harrigan (kharrigan@wm.edu)* **& Sadhwi Srinivas (ssrinivas@wm.edu)***
Program for Linguistics, William & Mary

Abstract

A well-known hypothesis for language learning is that learners are predisposed to learn patterns attested in human language, while eschewing alternative hypotheses that are linguistically unattested. Despite the popularity of this hypothesis, supporting experimental evidence has remained scarce. We aim to fill this gap by measuring the effectiveness with which adult learners can infer a linguistically attested pattern (a hierarchically nested constituency structure) in comparison to an unattested one (a constituency structure based on non-consecutive constituents) within a controlled artificial language learning task. We also investigate whether any observed learning biases extend to a non-linguistic, general puzzle-solving task. We find that our study participants are better at learning linguistically attested structures over unattested ones across the two types of learning domains. More broadly, our study serves as a methodological proof-of-concept extensible to testing the strength of other kinds of learning biases in linguistic and non-linguistic learning tasks.

Keywords: artificial language learning, domain-generality, hierarchical constituency, learning biases, syntax

Introduction

A prominent hypothesis for language learning claims that learners are predisposed to learn patterns attested in human language, while eschewing alternative generalizations that are linguistically unattested. Such predispositions are moreover alleged to be language-specific (Chomsky, 1975; Chomsky, 1980). Opposition to this framework questions on the one hand whether the learner brings any innate biases to the language-learning process at all, and on the other hand, whether any such biases insofar as they exist are truly language-specific (e.g., Tomasello, 2003; Bybee, 2006; Goldberg, 1995). The goal of the current work is to test: (i) whether predispositions towards inferring a specific linguistic property often claimed to be universal across languages can be robustly observed within a controlled language-learning task, and (ii) whether such a predisposition may also be observed in non-linguistic tasks.

The linguistic property we focus on here is that of hierarchical phrase structure, accepted as a universal tendency in human language.[1] Language learners have been said to bring a *hierarchical bias* to the task of language learning, wherein nested constituency structures are preferentially learned. Such a preference is claimed to extend even to learning contexts where the input does not uniquely determine a hierarchical structure (Crain & Nakayama, 1987). Despite the popularity of this hypothesis, empirical support from real-time language learning studies demonstrating such a preference has been quite limited (see also Futrell & Mahowald, 2025:13 who make a similar point).

The most direct piece of experimental evidence towards learners' preference for hierarchical structure-dependent rules comes from Smith et al. (1993), where a conlang "Epun" was used to examine the ease of acquisition of various rules by a language savant and four undergraduate controls. Epun instantiated an emphatic morpheme *nog* that was governed by a structure-independent computation rather than a hierarchical structure-based one. *Nog* always attached to the linearly third orthographic word in a sentence, regardless of its lexical category or other constituency structure. The rule for *nog* was found to be unlearnable.

More recently, Vender et al. (2020) report results from an Artificial Grammar Learning (AGL) study testing pattern-learning in children aged 8-11. The patterns were generated using either a grammar containing only linear regularities, or one containing both linear and hierarchical regularities. Using a reaction time measure, participants were found to be completely "sensitive" to the hierarchical pattern but only partially to the linear patterns. Some implicit neural evidence further supports the idea that hierarchical generalizations form a key aspect of language learning (e.g., Musso et al., 2003; Friederichi et al., 2006; Bahlmann et al., 2006), wherein only the language-attested patterns were found to produce activity in the known "language" areas of the brain.

Claims of a privileged status for hierarchical generalizations in language learning are nevertheless not uncontested. Ambridge et al. (2008) explicitly argue against Crain & Nakayama's influential finding by providing evidence that children do in fact make errorful inferences of the type claimed to be precluded by the presence of a hierarchical bias. It has also been claimed that the relevant input that children receive — which consists not only of the syntactic strings themselves but also form-to-meaning mappings in addition to general learning mechanisms such as

* Equal contribution.
[1] A reviewer points out Swiss German as a potential counter-example.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

abstract symbol manipulation/categorization and focusing of social and attentional cues — contains enough information to fully determine the hierarchical organization of natural language structures (e.g., Tomasello, 2005; Pullum & Scholz, 2002; Reali and Christiansen, 2005; van Valin, 1998; Goldberg & Suttle, 2010). In this view, hierarchical generalizations are only made by language learners because they are uniquely compatible with the input. As such, if the input were compatible instead with structure-independent rules, such rules would be as easily inferred. The recent remarkable successes of large language models have been touted as further support that innate learning biases are not necessary to the language learning process once sufficient input is ensured. Overall, a survey of the literature reveals mixed evidence towards the existence of a *hierarchical bias* in real-time language learning. Filling this gap using a controlled experimental design forms our first main aim.

A second aim of the current study seeks to further explore whether the purported *hierarchical bias* is specific to the language learning task, or if it is present domain-generally. Existing literature suggests that a preference for hierarchical generalizations is at play not only within the language domain, but in other areas of human cognition as well — including motor action, music, or mathematics (Coopmans, 2023; Dehaene et al., 2015). Fitch (2014) advances the "dendrophilia hypothesis," according to which "humans have a multi-domain capacity and proclivity to infer tree structures from strings, to a degree that is difficult or impossible for most non-human animal species" (cf. Trueswell, 2014). More generally, it has been argued that domain-general information processing constraints, and not language-specific constraints, lead to languages taking the form they do (e.g., Perfors et al., 2011; Futrell et al., 2015; Fedzechkina et al., 2018).

Note, however, that a domain-general bias for learning hierarchical structures in human learners does not rule out the possibility of an even more robust preference for hierarchical generalizations within an analogous linguistic task. Conversely, the propensity to disregard structure-independent hypotheses during pattern-learning may vary depending on whether or not the pattern being learned is known to be part of a natural language system. In the current study, we directly compare how well language-attested, hierarchical structures are learned within a language learning task and a more general (non-linguistic) puzzle-solving task. If a *hierarchical bias* is observed to the same degree across both tasks, this could be interpreted as evidence towards its truly domain-general nature. On the other hand, if we observe a difference between the linguistic and non-linguistic tasks, this can point to participants treating the exact same input 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 suggest domain-specific interactions with a potentially domain-general bias.

Pattern deduction task

Participants completed a pattern deduction task, in which they attempted to "escape from an escape room." During training phases, they were shown a series of sequences of words or shapes with evidence for the underlying constituency pattern. During the test phase, participants were tasked with recognizing whether novel strings were consistent with this pattern.

Participants

Participants were 141 undergraduates recruited from William & Mary's introductory Linguistics and Psychology courses. All materials were approved by W&M's Institutional Review Board. All participants gave informed consent and received course credit for their participation in the study.

Design

The study employed a 2x2 between subjects design. Participants' task was to "escape from a virtual escape room" by deducing one of two possible *constituency structures* (*nested* v. *crossed*) in one of two possible *contexts* (*language* v. *lockbox*). The goal was to test pattern learning abilities within and outside of language-learning contexts, for both linguistically attested and unattested grammars.

Contexts *Language context.* The *language context* was set up as an artificial language learning study. Participants were told that the guards of the escape room were speakers of a language, "Nog", who communicate by slipping notes under the door. To "make their escape", participants needed to trick the guards into thinking they also speak Nog by similarly slipping them notes. To prepare for this, participants were shown examples of Nog sentences to learn some basic words and patterns of the language. Specifically, participants were exposed to Nog sentences that were consistent with one of two possible patterns: one that is linguistically attested (*hierarchically nested constituency structure*), and one that is systematic but linguistically unattested (*crossed constituency structure*). The *language context* was made to emulate language learning in two ways: (1) the communicative component and (2) the language script. The language sequences were derived from the Georgian script, which has the clear appearance of being a language, but is a script with which participants are unlikely to be familiar. The words were derived from a small set of Georgian characters acting as "vowels" in the center of words, and each word having unique characters on the edges (Table 1). Throughout the study, participants were asked to use their knowledge of Nog based on the examples they saw during the exposure phases to choose between novel sequences in a forced choice task.

Lockbox context. The *lockbox context* was similar to the *language context*, except there was no communicative component, and the sequences were made up of more clearly non-linguistic shapes (Table 1). Instead of "learning a language", participants were told that to escape, they would need to solve a series of combination locks to unlock a series of doors and lockboxes. Participants were shown sequences consistent with one of the same two possible patterns as in the *language contex*t: one that is linguistically attested (*hierarchical nested constituency structure*), and one that is

systematic but linguistically unattested (*crossed constituency structure*). 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.

Constituency Structures *Nested constituency structure.* The *nested* constituency structure was a hierarchically organized "grammar" consisting of three phrases, each containing two or three words/shapes (Figure 1).

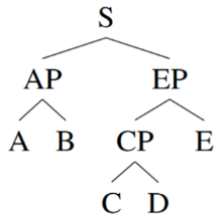


Figure 1: Constituents of the *nested* condition.

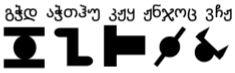
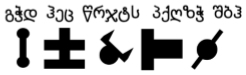
This grammar included hierarchical embedding, with one phrase (CP) nested inside another phrase (EP). It was designed to be a simplified version of natural language, following Takahashi & Lidz (2007), in which constituents are consecutive strings of words which may be hierarchically nested, and over which transformations occur. Basic, non-transformed sequences consisted of five positions, each of which had two possible words – a longer word and a shorter one chosen to mimic variable word lengths observed in natural languages – leading to 32 possible grammatical basic sequences (Table 1).

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

	A	B	C	D	E
language	კჟყ	შბპ	გჭდ	ჰეც	ვჩჟ
	პქლზჭ	ჟნჯოც	ჯფლცმ	აჭთჰუ	წრჯტს
lockbox					
					

Two operations occurred in the transformed sequences of this grammar: deletion and movement. In the deletion sequences, one of the three constituents was deleted. In the movement sequences, one constituent was moved to the beginning of the sequence. We did not include movement sequences for the constituent [AB], as this was already at the beginning of the utterance in the base form (Table 2).

Table 2: Transformed sequences of *nested* structure.

	constituent	sequence	examples
deletion	AB	CDE	გჭდ აჭთჰუ ვჩჟ
	CD	ABE	პქლზჭ შბპ ვჩჟ 
	CDE	AB	პქლზჭ შბპ 
movement	~~AB~~	~~ABCDE~~	
	CD	[CD]ABE	გჭდ აჭთჰუ კჟყ ჟნჯოც ვჩჟ 
	CDE	[CDE]AB	გჭდ ჰეც წრჯტს პქლზჭ შბპ 

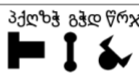
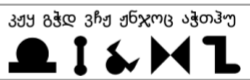
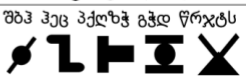
In this condition, the transitional probabilities between elements within constituents with two positions (AP, CP) is 1.0—A always predicts B, C always predicts D. They are adjacent in basic sequences, as well as in all transformed sequences. The transitional probability between the middle and last elements of the three-element constituent (D and E) is 0.7—they are adjacent in the basic sequences and move together when the constituent EP moves but are separated when the embedded constituent CP moves or is deleted. By contrast, non-constituents have transitional probabilities of either 0 or 0.33.

Crossed constituency structure. The linguistically unattested *crossed constituency structure* was a systematic pattern that also utilized transformations over "constituents." However, in this case the constituents were not consecutive strings but consisted of words in *linearly alternating* positions, leading to tangled or "crossed" branches of a syntax tree, violating the Non-Tangling Condition (see Partee et al., 1990, chapter 16.3). Basic, non-transformed sequences were identical to those in the hierarchical pattern (Table 1). The grammar consisted of two constituents: ACE, and BD (Figure 2).

Figure 2: Constituents of the *crossed* condition.

The same two operations occurred in the transformed sequences of this grammar: deletion and movement. In the deletion sequences, one of the two constituents was deleted. In the movement sequences, one of the two constituents would move to the beginning of the sequence (Table 3).

Table 3: Transformed sequences of *crossed* structure.

	constituent	sequence	examples
deletion	ACE	BD	
	BD	ACE	
movement	ACE	[ACE]BD	
	BD	[BD]ACE	

In this condition, the transitional probabilities between elements of constituents are between 0.44 to 0.67. This is due to the nature of the basic sequences—the elements making up "constituents" are linearly alternating and therefore are never adjacent to each other. In the transformed sequences, however, constituent elements are always adjacent to each other. Although transitional probabilities are never 1.0, transitional probabilities between elements of constituents are always higher than between any non-constituent pairings, which are always 0 or 0.33.

Procedure

Participants were randomly assigned to one of four possible conditions based on two possible factors: either the *nested* or the *crossed constituency structure* in either the *language* or *lockbox context*. The study was deployed via Qualtrics and guided participants through the multiple exposure and test phases.

Vocabulary Participants were exposed to all 32 possible basic combinations in an initial vocabulary exposure phase. They were then given eight vocabulary check questions, which asked them to choose which of two sequences was one they had been exposed to, testing their ability to distinguish grammatical from ungrammatical basic strings. The ungrammatical foils were scrambled sequences, where the words or shapes were not in the correct order. 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 16 total vocabulary check questions. No feedback about the correctness of the participants' chosen option was provided in either vocabulary check phase.

Transformations Participants then moved on to the transformation exposure phases. In the *language context*, they were told that the Nog-speaking guards had started to suspect that there was an imposter among them, so participants needed to increase the complexity of the notes they were passing in order to not get caught. Thus, during this phase they would see some harder sentences. In the *lockbox context*, they were told that as they moved through the escape-room they would need to solve more complex combination locks than before, and that they would see some more complex sequences to practice. During this phase, they were exposed to a representative set of 74 transformed sequences (Table 4). Because the deletion transformation generates a low number of logically possible sequences, some repetition was necessary in the training sequences in order to reserve novel deletion sequences for the test phase. The number of repeated examples in training were balanced across the *nested* and *crossed constituency structures*. No movement sequences were repeated in training.

Table 4: Distribution of transformed sentences.

				training sentences			test
	transformation	constituent	output of transformation	# unique	# repetitions	total	total
nested	deletion	AB	CDE	3	3	9	2
		CD	ABE	2	3	6	2
		CDE	AB	2	3	6	1
	movement	CD	CDABE	27	n/a	27	5
		CDE	CDEAB	26	n/a	26	6
crossed	deletion	ACE	BD	3	3	9	1
		BD	ACE	4	3	12	4
	movement	ACE	ACEBD	26	n/a	26	6
		BD	BDACE	27	n/a	27	5

Participants were then given eight test questions, in which they were asked to choose which of two novel sequences was more likely to be a grammatical sentence of the target language, or a valid lockbox combination. The ungrammatical sequences in both conditions consisted of a transformed sequence generated by deleting or moving a non-constituent string (AD or ADE). These strings had transitional probabilities of 0 under both the *nested* and the *crossed* constituency structures. Recall that the lower-level statistical cues for constituency were not identical across the two grammars—the highest transitional probability for constituents in the *nested* condition was 1.0 compared to 0.67 in the *crossed* condition. We chose transformed sequences with 0 transitional probabilities as foils for both conditions in order to mitigate this difference; in both cases they were choosing between highly likely sequences and sequences that categorically never occurred. The transformation exposure was then repeated, followed by eight new test questions. Thus, each participant saw a total of 148 training sequences and 16 test questions (5 deletion, 11 movement). Here again, participants did not receive any feedback about the correctness of their chosen options in the test phases.

Results

We excluded participants who chose the grammatical string at a rate less than 75% in the vocabulary check phase (18 in the *language context*, 3 in the *lockbox context*), leaving us with 120 participants for the final analysis—30 across each of the four conditions.

We find that participants successfully learned both the *nested* and *crossed* constituency structures above chance in both the *language* and the *lockbox contexts* (Figure 3).

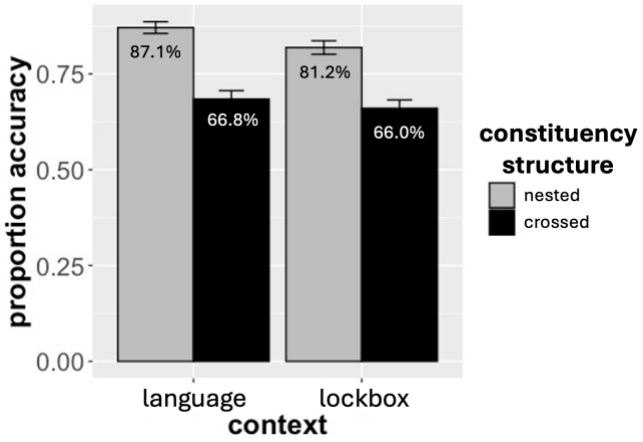


Figure 3: Results.

Table 5 shows accuracy by transformation across contexts and constituency structures.

Table 5: Accuracy by transformation.

		language	*lockbox*
nested	movement	.89	.84
	deletion	.83	.74
crossed	movement	.71	.71
	deletion	.63	.55

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, *constituency structure* (*nested* v. *crossed*) and *context* (*language* v. *lockbox*) as fixed effects, and a random intercept for *subject*. We further included *transitional probability* (*TP*) as a fixed effect to ensure that any effects from the manipulations persisted even after accounting for differences in *TP*s across the *nested* and *crossed* grammars. We find a significant effect of *TP* [$X^2_{(1)} = 26.05$, $p < 0.0001$]. Additionally, we find a main effect of *constituency structure* [$X^2_{(1)} = 33.20$, $p < 0.0001$] but no main effect of *context* [$X^2_{(1)} = 1.40$, $p = 0.24$] and no interaction between *constituency structure* and *context* [$X^2_{(1)} = 0.59$, $p = 0.44$]. The statistical model indicates that while the *TP* of individual constituents did influence accuracy, there is a difference in performance based on *constituency structure* impacting participants above and beyond the impact of these low-level statistical cues.

Comparison of performance on individual constituents within each grammar provide additional insight into the effect of low-level statistical cues on learning. Although the average transitional probability for constituents in the *crossed* condition is lower than that of the *nested* condition, not all individual constituents within a condition have identical transitional probabilities. For example, the constituent CDE in the *nested* condition has a comparable average transitional probability to that of the constituent BD in the *crossed* condition (~.7). If low level statistical cues were driving the observed differences, we would expect accuracy on items including these two constituents to be identical. Instead, we see a pattern that matches that of our overall finding: accuracy for CDE sequences in the *nested* condition is higher than accuracy for BD sequences in the *crossed* condition for both the *language* (92% v. 59%) and the *lockbox* (82% v. 61%) conditions.

Overall, we find that while participants successfully learned both *constituency structures* in both *contexts*, they learn the *nested constituency structure* more successfully than the *crossed constituency structure*, both in the *language* and the *lockbox contexts.*

Discussion

The current study was designed with two goals in mind. First, we sought to empirically verify the hypothesis that language learners bring a *hierarchical bias* to the language learning process, wherein they are better equipped to infer hierarchically nested constituency structures within strings as compared to language-unattested constituents. Second, we aimed to test whether such a bias is limited to language-learning contexts or whether it can be observed more generally across cognitive domains.

To address these goals, our study measured adult participants' ability to learn constituency structures and generalize transformations over two types of grammars: one containing hierarchically nested constituents of the kind that natural languages regularly instantiate, the other containing linguistically unattested, linearly crossed constituents. Each grammar was learned either as part of a language-learning task, or within a non-linguistic, puzzle-solving task. There were thus four types of pattern-learning situations overall, able to be fruitfully compared in order to answer our research questions.

First and foremost, we found evidence of successful learning by participants in all four learning situations, including the unattested, crossed constituency structure within the language learning task. Despite the learnability of the crossed constituency structure, it was nevertheless learned less successfully than the grammar containing nested constituents. Participants who were exposed to the latter grammar within the language-learning context enjoyed a significant learning advantage over their counterparts exposed to the former, suggesting the existence of a *hierarchical bias* in this context, confirming the long-held theoretical anticipation to this effect. Such a bias could drive hierarchical generalizations in situations where learners are provided ambiguous input (e.g., Crain & Nakayama, 1987).

We further find evidence that the *hierarchical bias* is domain-general, not limited to the language learning process, consistent with conclusions from other studies in the recent literature (Coopmans, 2023; Dehaene et al., 2015; Fitch, 2014). Not only did participants exhibit successful learning of both types of constituency structures within the non-linguistic learning context, but the nested constituency pattern was once again learned robustly better than the crossed constituency pattern. This preference for nested patterns was comparable across both learning contexts.

While our results are most consistent with the existence of a domain-general *hierarchical bias*, there exist a few, albeit unlikely, alternative possibilities. One might wonder whether the input provided to participants in the *lockbox* context was, despite the set-up of the task, nevertheless interpreted as linguistic input. The left-to-right ordering of symbols within the strings, along with the awareness that writing systems can vary widely in the symbols they use, could have led our undergraduate participants to implicitly treat this input as language, in spite of the top-down instructions they had received. To rule out this possibility, follow-up studies must test pattern learning in more clearly non-linguistic learning contexts (e.g., with colors or musical notes).

Another point worth considering is that the linguistic task in our study involved the learning of structures alone, and not structures mapped to meanings. This is not an unusual design for artificial language learning studies in the lab, but is quite unlike what is actually involved in the language-learning process outside of the laboratory. In a version of the current study where participants are taught form-to-meaning mappings instead of form alone, it is possible that language-specific biases would be observed more strongly.

The finding that participants in our study successfully learned linguistically unattested patterns is at odds with claims that learners never consider such patterns (e.g., Smith et al., 1993). It is possible that participants' observed inability to learn the structure-independent rule in that study was due to its juxtaposition on top of an underlying structure-dependent grammar, which may have contributed to difficulty in learning. By contrast, in the current study, participants were working with a language-unattested, truly structure-independent grammar.

This discrepancy could also be because the linearly-alternating pattern instantiated in our crossed grammar approximates one or more types of linguistically attested patterns more closely than we had intended. That is, while it is true that natural languages are unlikely to allow for crossed constituents, other types of non-adjacent, long-distance phenomena are known to occur. For example, long-distance filler-gap dependencies, assimilation processes such as vowel harmony, binding phenomena, to name a few. As such, it is possible participants identified the structures formed within our crossed grammar with such long-distance dependencies.

The acknowledgement of this possibility begs the larger question of what it would take to instantiate a pattern that we can be more fully certain is not linguistically attested. One candidate that comes to mind 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 recalling once again that in their study, such a rule was formulated on top of an essentially structure-dependent grammar. It is much harder to conceptualize a completely structure-independent grammar made up solely of rules along the lines of the one employed by Smith et al. Moreover, in the current study which aimed to directly compare the learnability of linguistically attested and unattested grammars within the same learning context, we required that the overall complexity of the two grammars be roughly matched, as well as their vocabulary sizes and lower-level statistics. We leave open the question of whether a more convincingly language-unattested pattern can be suitably instantiated within the current design.

Finally, though our findings indicate that the *hierarchical bias* is unlikely to be an instance of a language-specific bias, this does not rule out the possibility of other language-specific biases. It is our hope that further progress can be made on the question of whether there are any language-specific biases at all 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.

Conclusion

Our work provides evidence that while non-hierarchical, linguistically unattested patterns are learnable within a language context (contrary to previous claims as in Smith et al., 1993 or Yang, 2006), there is nonetheless a robust advantage for linguistically attested patterns, which also persists in non-linguistic tasks. Aside from the empirical findings, the current study also introduces a novel methodology to test the learnability of various types of patterns in language and non-language contexts.

References

Ambridge, B., Rowland, C.F. and Pine, J.M. (2008). Is Structure Dependence an Innate Constraint? New Experimental Evidence From Children's Complex-Question Production. *Cognitive Science*, 32, 222-255. https://doi.org/10.1080/03640210701703766

Bahlmann, J., Gunter, T. C., & Friederici, A. D. (2006). Hierarchical and linear sequence processing: An electrophysiological exploration of two different grammar types. *Journal of cognitive neuroscience*, *18*(11), 1829-1842.

Bybee, J. (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, N. (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, N. (1980). *Rules and Representations*. Oxford, England: Basil Blackwell.

Crain, S., & Nakayama, M. (1987). Structure dependence in grammar formation. *Language, 63*, 522–543.

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

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*(1), 2-19.

Fedzechkina, M., Chu, B., & Florian Jaeger, T. (2018). Human information processing shapes language change. *Psychological science*, *29*(1), 72-82.

Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, US adults, and native Amazonians. *Science Advances, 6*(26).

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

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America, 103*(7), 2458–2463. https://doi.org/10.1073/pnas.0509389103

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336-10341. https://doi.org/10.1073/pnas.1502134112

Futrell, R., & Mahowald, K. (2025). How Linguistics Learned to Stop Worrying and Love the Language Models. *arXiv preprint arXiv:2501.17047*.

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

Goldberg, A., & Suttle, L. (2010). Construction grammar. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(4), 468-477.

Liao, D. A., Brecht, K. F., Johnston, M., & Nieder, A. (2022). Recursive sequence generation in crows. *Science Advances, 8*(44). https://doi.org/10.1126/sciadv.abq3356

Musso, M., Moro, A., Glauche, V., Rijntjes, M., Reichenbach, J., Büchel, C., & Weiller, C. (2003). Broca's area and the language instinct. *Nature neuroscience*, *6*(7), 774-781.

Partee, Barbara BH, Alice G. Ter Meulen, and Robert Wall. (2012). *Mathematical methods in linguistics*. Vol. 30. Springer Science & Business Media.

Perfors, A., Tenenbaum, J. B., & Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118(3), 306-338. https://doi.org/10.1016/j.cognition.2010.11.001

Pullum, G. K., & Scholz, B. C. (2002). Empirical assessment of stimulus poverty arguments. *The linguistic review*, *19*(1-2), 9-50.

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/.

Reali, F., & Christiansen, M. H. (2005). Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence. *Cognitive Science, 29*(6), 1007–1028.

Smith, N. V., Tsimpli, I. M., & Ouhalla, J. (1993). Learning the impossible: The acquisition of possible and impossible languages by a polyglot savant. *Lingua, 91*(4), 279–347.

Takahashi, E., & Lidz, J. (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, M. (2005). Beyond formalities: The case of language acquisition. *The Linguistic Review, 22*(2–4), 183–197. https://doi.org/10.1515/tlir.2005.22.2-4.183

Truswell, R. (2017). Dendrophobia in bonobo comprehension of spoken English. Mind & Language, 32(4), 395-415. https://doi.org/10.1177/0956797617728726

van Valin, R. D., Jr. (1998). The acquisition of WH-questions and the mechanisms of language acquisition. In M. Tomasello (Ed.), *The new psychology of language: Cognitive and functional approaches to language structure* (pp. 221–249). Mahwah, NJ: Erlbaum.

Vender, M., Krivochen, D. G., Compostella, A., Phillips, B., Delfitto, D., & Saddy, D. (2020). Disentangling sequential from hierarchical learning in Artificial Grammar Learning: Evidence from a modified Simon Task. *Plos one*, *15*(5), e0232687.

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