On what counts as domain-specific: a response to Wilcox et al. (2023)

Deniz Satık, Harvard University
deniz@g.harvard.edu

This short discussion reexamines Wilcox et al. (2023), drawing attention to the usage of the filler-gap dependency as the object of investigation in their study.[1] This squib points out that most purported innate properties, barring a few such as recursion and structure dependence–which enjoys universal acceptance among generative linguists–are untenable. This standpoint critically impacts the Poverty of Stimulus argument, as it aligns with the modern generative view that most linguistic attributes arise from domain-general, rather than language-specific, mechanisms. Under a more modern conception of generative grammar, the filler-gap dependency is reduced to principles of computational efficiency. The goal of this short discussion is to modernize the poverty of stimulus argument in light of Darwin's problem, and provide directions for future research involving language models to take.

Keywords: Language Acquisition, Generative Grammar, Innateness, Domain-General Learning, Grammaticality

1 Introduction

Wilcox et al. (2023) aims to examine the syntactic patterns that language models trained with data volumes equivalent to a child's exposure can discern, in light of the filler-gap dependency and their island constraints. By understanding what these data-driven algorithms can learn without clear biases towards specific domains, they set a baseline for learnability, and help identify what, if any, is a candidate for a domain-specific linguistic property. My goal in this squib is to show that the filler-gap dependency is reduced to the domain-general explanation of computational efficiency under more modern concepts of generative grammar, without needing to posit domain specificity. I propose alternate paths for research via language models, focusing on Merge and its more general consequences such as structure dependence.

2 Background

The seminal criticism by Chomsky (1959) towards Skinner (1957) sparked a central debate in linguistics: does an innate understanding of grammar drive human language capabilities, or is exposure alone sufficient? Today's linguistic scholars typically gravitate towards either Chomsky's or

[1] Chaves (2020) presents more specific objections to Wilcox et al. (2018, 2019) on what language models actually learn regarding the filler-gap dependency, to which I refer the reader for further details. My discussion here will not be centered on the experimental method.

Skinner's view on this matter, as detailed in (Dąbrowska, 2015). Those favoring Chomsky believe that certain abstract, domain-specific grammatical rules are integral to the universal language faculty of humans.

On the other hand, usage-based theories, which link linguistic phenomena to functional reasons, negate the existence of such domain-specific constructs. However, interpretations within these theories differ (Tomasello, 2009). Various domain-agnostic mechanisms like abstraction (Lieven, 2014), statistical learning (Christiansen and Chater, 2016), and generalization (Goldberg, 2019) are proposed as adequate underpinnings for the process of language acquisition.

Chomsky's *poverty of the stimulus* (PoS) argument (Chomsky, 1959, 1980) sits at the heart of this ongoing debate. The PoS contention emphasizes that specific linguistic insights cannot solely arise from environmental input, pointing towards innate cognitive structures. This argument is methodically broken down into four distinct parts, as proposed by (Pullum and Scholz, 2002).

The 'acquirendum' refers to a particular piece of syntactic understanding. The 'indispensability' part pinpoints the essential input for a learner to assimilate this 'acquirendum'. The 'inaccessibility' part highlights that such pivotal input is elusive to the learner. Finally, the 'acquisition' component illustrates that even amidst these limitations, syntactic comprehension manifests at an impressively young age. Assuming this holds, we come to the conclusion that learners are capable of grasping syntactic intricacies that aren't fully inferable from their surroundings, hinting at inherent learning boundaries.

The PoS contention has stirred considerable debate in the literature, with numerous researchers advocating for or challenging the notion of innate linguistic principles. However, the emergence of language models (Chen and Goodman, 1999) is purported to offer empirical grounds to evaluate these learnability claims (Elman et al., 1996). While critics might point out that these models don't replicate human language acquisition—given they process vastly greater volumes of data, as (Chomsky et al., 2023) point out—recent endeavors in deriving linguistic structures through statistical learning, such as the filler-gap dependency, even strive to replicate the linguistic exposure children receive, offering them an analogous quantum of stimulus (Wilcox et al., 2023).

As observed by Contreras Kallens et al. (2023), the responses generated by Large Language Models in particular are predominantly grammatically sound. The actual shortcomings predominantly appear in the realms of semantics, pragmatics, and facets tied to social engagement. Consequently, they argue that Large Language Models dispense with the necessity to postulate inherent, domain-specific linguistic properties.

With this background established, Wilcox et al. (2023) study the *filler-gap dependency*, and their *island constraints*, following Ross (1967), and whether they can be acquired by three types of language models.[2] The filler-gap dependency describes the relationship between a wh-word or

[2]The three models that Wilcox et al. (2023) use in their experiment, *n*-grams, Recurrent Neural Networks and Transformers, are both which are both domain general and weakly biased in the sense of Pearl and Sprouse (2013). The n-gram model is used as a baseline. They find that RNNs are likely to be better representations of human linguistic skills than Transformers, at least concerning this particular phenomenon.

There are, of course, independent concerns with the models that Wilcox et al. (2023) use. For instance, one wonders whether they would be able to acquire ungrammatical structures with the right set of data–which no human can do. This appears to be likely. Mitchell and Bowers (2020) explores the learning capabilities of domain-general and recurrent Long Short-Term Memory (LSTM) networks: in particular, their ability to learn unnatural linguistic structures involving number agreement, and they find that they do acquire ungrammatical structures. But for my purposes, my critique of their experiment is very general, so the in-depth performances of the models and their characteristics

phrase (known as a filler) and an unoccupied syntactic space (referred to as a gap). To illustrate, consider (1a). Here, the relationship between the filler and the gap stretches across two embedded clauses. But in a similar structure, like (1b), the sentence becomes ungrammatical if the gap is located inside a syntactic island, more specifically, a Complex Noun Phrase.

(1) a. I know what the guide said his friend saw the lion devour last night.
b. * I know what the guide saw the lion that devoured last night.

For a comprehensive understanding of the filler-gap dependency and its rules, two interconnected points must be addressed. First, regardless of variations between languages, similar structures consistently emerge as syntactic islands across different languages. Second, despite limited or even incorrect input from caregivers, children within a specific language group seem to consistently identify these syntactic islands. Wilcox et al. (2023) primarily explores this second point, delving into the aspect of learnability.

To examine the filler-gap dependency sentences, they employ a 2x2 interaction design, drawing inspiration from two grammar-based predictions. The first hypothesis is rooted in the understanding that gaps need fillers for valid syntactic formation. Consider (1a). If we alter the filler to a non-wh expression like *that*, the sentence becomes invalid, as in (1b). [3]

a. I know what the lion devoured _ yesterday.
b. * I know that the lion devoured _ yesterday.

To gauge if the models grasp this prediction, they analyze the surprise element in the adverbial phrase modifier and concluding punctuation following the primary verb, as in *yesterday*, from (1a)-(1b). If the models understand that gaps are only valid when preceded by a filler, then the sequence from *devoured* to *yesterday*, which misses a clear NP object, should be more expected when a filler is present and less so without it. In simpler terms, *yesterday* should seem more out of place in (1b) than in (1a). This is referred to as the *wh-effect* in the +GAP condition.

The second grammar-based hypothesis asserts that a filler at the beginning mandates a gap later in the sentence. When analyzing the correct sentence (1a) and replacing the *that* with a wh-expression, as in (1b), we find that the structure breaks.

a. I know that the lion devoured the gazelle yesterday.
b. * I know what the lion devoured the gazelle yesterday.

If models understand that an initial filler anticipates a subsequent gap, then upon encountering a filler, they should foresee an upcoming void. Consequently, the presence of *the gazelle* should be more unexpected in (1b) than in (1a).[4] This is referred to as the *wh-effect* in the −GAP condition. Integrating both predictions, they conceptualize a 2x2 contrast. To be more precise, they anticipate an interaction between the occurrences of fillers and gaps, expecting models to exhibit greater comprehension when both elements coexist compared to their singular presence.

From this experiment, they derive two key insights: Firstly, the outcomes show that models are adept at navigating "wh-expectations" within intricate syntactic structures. This ability enables us to ascertain that the diminishing wh-effects noted in the hierarchical study aren't merely

are irrelevant. I refer the reader to Wilcox et al. (2023) for further details.

[3]However, it's crucial to note that this only applies when the primary verb mandates an object: "I know that the lion ate yesterday" remains valid.

[4]This phenomenon, identified by Crain and Fodor (1985) and Stowe (1986), is termed the filled-gap effect.

a result of extended distance but are influenced by inherent structural relationships. Secondly, considering that these models probably encountered a minimal number of sentences with four or more embedding layers, their pronounced "wh-effects" under such circumstances suggest they are extrapolating knowledge beyond just their training input. In essence, the rationale behind the models' efficacy is rooted in their primary goal: to optimize the likelihood of data interpretation.

The experimental approach being utilized is undeniably innovative and offers a fresh perspective on this age-old debate. However, the real traction in this dialogue can only be achieved when the appropriate linguistic property is under the microscope. To my current understanding, the literature hasn't marked filler-gap dependencies as emerging from inborn, domain-specific characteristics. And, even granting that assumption, such a stance has been rendered untenable in the light of the Minimalist Program introduced by Chomsky in (1995).

By the dawn of the 1990s, Chomsky, among other linguists, had already initiated a critical examination of the scope and role of computational principles within UG. The most optimal solution would be to assume that UG reduces to the absolute simplest computational principles. This possibly includes nothing beyond the recursive and structure-generating operation known as Merge—a concept encapsulated by the Strong Minimalist Thesis (SMT). While the SMT has been articulated in many ways by Minimalists, at its core, it essentially postulates that the entirety of human linguistic syntax can be derived from three elements:

(2) a. the syntactic operation Merge
    b. interface conditions (involving semantic and phonetic restrictions)
    c. principles of "efficient computation"

In this conceptualization of the human linguistic capacity, Merge stands alone as the sole entity exclusive to language. As highlighted by Chomsky (2020), this conclusion appears counterintuitive at first: many properties of language, be it the sequential arrangement of words, filler-gap dependencies, or the act of copy deletion, aren't intrinsically linguistic in nature. Instead, they emerge from the interface conditions and principles that prioritize efficient computation, both realms that operate independently of language. Let's delve deeper into this notion.

3 Syntax as simple

Chomsky (2000) posits that language emerges as an optimal adaptation to legibility criteria. This perspective dovetails with Berwick and Chomsky's (2011) proposition that the generative engine prioritizes computational efficiency. Language is molded by Merge, an elementary recursive process adept at meeting interface demands while preserving computational efficacy. Merge is pivotal in syntactic construction, combining two syntactic entities into a unified syntactic set. Drawing a parallel, just as snowflakes are crafted by inherent natural laws, Berwick & Chomsky argue that language is sculpted by interface constraints and principles of efficient computation.

This perspective is termed the Strong Minimalist Thesis (SMT). Upon finalizing all syntactic functions in the derivation of a sentence, it is sent to both the phonological and semantic interfaces for further evaluation. The phonological interface, affiliated with the sensorimotor system, facilitates externalization, encompassing activities like speech production and parsing. This interface might also be tasked with deleting copies in a syntactic derivation, among other operations.

On the other hand, the semantic interface, associated with the conceptual-intentional system, oversees cognitive processes such as inference, strategic planning, and interpretation. Frameworks like the Case theory, binding theory, control theory, and the θ-Criterion might find explanations within this system. Crucially, it's essential to recognize that these systems are *external to language*, implying they aren't intrinsic parts of Universal Grammar (UG). The process of Merge operates on lexical entries, which are clusters of both semantic and phonological attributes found in the *lexicon*. Employing Merge, tree structures are constructed from the ground up, selecting lexical components from the lexicon.

The Strong Minimalist Thesis (SMT) is formally defined by Chomsky (2004) as follows. Let's consider that the faculty of language starts with an innate state, denoted as S_0. This initial state, S_0, stands for Universal Grammar (UG) and delineates all potential states in which a particular language, L, can exist. The primary objective of a Minimalist is to minimize the components within S_0. Observing this from a language acquisition perspective, our primary focus falls on the three elements (3a)-(3c):

(3) a. unexplained elements of S_0
    b. interface conditions (the principled part of S_0)
    c. principles of efficient computation

Chomsky (2004) defines the SMT as the proposition that there are no unexplained elements of S_0: (3a) is devoid of content. While Merge stands as a distinct linguistic operation, Chomsky considers it an inherent, rather than inexplicable, element. This is attributed to its characterization as the most rudimentary operation able to elucidate recursion in human language.

Chomsky proposes that numerous syntactic operations and theories, including Case, agreement, binding theory, filler-gap dependencies (pertinent to our current discussion), the elimination of Copies, and many others, can essentially be distilled down to either (3b) or (3c). Such a perspective offers an elegant resolution to Darwin's puzzle on the evolution of language. As per the SMT, the foundational structure for human language was pre-existing and evolved due to a single genetic alteration, bestowing upon humans the capability of recursion.

The SMT is not universally accepted. For instance, regarding (3b), Chomsky and fellow Minimalists acknowledge their current limitation in formally defining parts of the SMT. As (Progovac, 2019) points out, this makes it unfalsifiable. On the other hand, Satık (2022) suggests that certain crosslinguistic parallels, derived from a myriad of languages, defy a simple reduction to the SMT's framework of interface conditions or principles of computational efficiency. As highlighted by Wilcox et al. (2023), these crosslinguistic similarities might bolster the case for linguistic innateness.

But no matter where one stands on the SMT debate, there's unanimous agreement among the scholars mentioned that Darwin's dilemma cannot be overlooked when crafting hypotheses. Instituting innate systems or structures might be viable under certain specific conditions, but such assertions must be rigorously substantiated, especially when the SMT doesn't suffice in explaining observable phenomena. In terms of language acquisition, it's crucial to provide adequate reasons to counteract the potential explanations offered by (3b) or (3c).

Under any interpretation of Minimalism, one property that is clearly and uncontroversially amenable to reduction is the filler-gap dependency. According to Chomsky (2023), the filler-gap dependency is a bona fide example of an instance where computational efficiency is at odds with communicative effectiveness, in which the latter is usually sacrificed. As he notes, an inherent

characteristic of computational processes is that symbols retain their meaning throughout the operation, a principle which Chomsky calls Preservation. Total annihilation of syntactic objects is not undertaken by the Merge operation.[5]

Consequently, when generating a phrase such as "which books did you read", given that trees in the Minimalist Program are built bottom up, it initially forms as "which books did you read which books", translating directly to "for which books x, did you read the books x". What our cognition processes is the repeated phrase "which books did you read which books", whereas what is perceived audibly and registers in our consciousness is the truncated "which books did you read?". This process of externalization discards all repetitions, save for one that serves as an indicator that the operation has taken place. This is a syntactic economy condition, significantly minimizing both cognitive processing and the motions involved in articulation.

These operations aim to enhance computational efficiency, albeit giving rise to what is termed a filler-gap problem for syntactic analysis. Here, the parser encounters a wh-phrase such as "which books" and endeavors to identify the corresponding gap that warrants fulfillment for interpretation. This emerges as a significant complexity in regard to communicative efficiency, posing one of the cardinal challenges in syntactic analysis. This complexity could be substantially diminished if the gap was occupied. However, nature in its design of language opted for the more elegant approach, undeterred by the complexities it introduces, even as sentences grow more intricate. In other words, the filler-gap dependency has nothing to do with domain-specific linguistic properties. It is merely reduced to (3c)–principles of communicative efficiency–and is not technically a part of syntax, and thus irrelevant to the poverty of stimulus argument.

With the bar set higher now for proving truly domain-specific linguistic ability in generative grammar than it was previously, this necessitates a reconsideration of the poverty of stimulus argument.[6] In order for one to make genuine progress on the poverty of stimulus argument, one has to look at the sole domain-specific entity for language: Merge. One natural consequence of Merge is the property of structure dependence in syntax.[7] Consider (4) from Chomsky (2023):

(4) The man who fixed the car carefully packed his tools.

This sentence is subject to two interpretations: either the act of fixing the car was done carefully, or the packing of the tools was carried out with care. However, when we reposition the adverb at the beginning, like in (5), the ambiguity vanishes, leaving only the interpretation that the packing of the tools was done carefully.

(5) Carefully, the man who fixed the car packed his tools.

Though the initial position of the adverb would logically prompt it to modify the nearest verb phrase, which is "fixed the car", this is bypassed instinctively for the more distant verb phrase

[5] See Merge as defined by Chomsky (1995).

[6] Alternatively, as Wilcox et al. (2023) mention, another potential method of supporting linguistic nativism is to identify similarities or differences across cross-linguistic data. As Satık (2022) indicates, if these properties resist a reduction to interface conditions or principles of computational efficiency in the style of the SMT, then they provide evidence for domain-specific properties beyond just recursion.

[7] In addition to structure dependence, it would be useful to look at how language models acquire recursion, which is a more general consequence of Merge. But it would be useful to develop a fuller understanding of how it is acquired in children, as well. Though several studies regarding the acquisition of recursion in children have been conducted, many of these involve recursion in possessive NP-constructions (Pérez-Leroux et al., 2012; Terunuma et al., 2017; Shi and Zhou, 2018; Giblin et al., 2019).

"packed his tools". This choice becomes apparent when analyzing the sentence's structure: the "packed his tools" phrase actually maintains closer proximity to the adverb in the underlying structure, which guides our interpretation even though it's not overtly presented in the linear arrangement of words. Structure dependence such as this is ubiquitous in syntax.

In fact, there is already discussion in the literature whether models are capable of acquiring this property. Reali and Christiansen (2005) constructed three models to explore the acquisition of yes-no questions: a bigram statistical model, a trigram statistical model and a simple recurrent neural network model. They conclude that statistical information is sufficient for acquiring auxiliary fronting in polar interrogatives. Kam et al. (2008) and Berwick and Chomsky (2011) provide a response to Reali and Christiansen (2005), claiming that the results for each model are far from compelling. Both note that the findings of the experiment are significantly limited, and given that other, more robust statistical models have only managed to excel in the same restricted set of circumstances as the bigram model, it is still unclear whether the concept of stimulus richness can be more broadly validated. But it remains to be seen whether the methods of Wilcox et al. could help push this debate forward.

4 Concluding Remarks

To recap, the primary objective of this short discussion has been to highlight that any hypothesis attributing linguistic ability to innate properties is subject to Darwin's conundrum: how did these innate features evolve? Upon examination, it becomes increasingly improbable for most supposed innate properties to exist, apart from structure dependence and recursion, which are universally acknowledged as innate by generative linguists, and potentially a few others.

This has a significant and direct effect on the PoS argument, as according to Chomsky, the majority of linguistic properties are not language-specific but are the outcomes of domain-general mechanisms. As such, although the methods and tools used in Wilcox et al. (2023) are a significant step forward, instead of focusing on phenomena like filler-gap dependencies, the debate regarding the truth of linguistic nativism would be further advanced via future research on the acquisition of structure dependency or recursion–whether by children or by language models.

References

Berwick, Robert C., and Noam Chomsky. 2011. Language evolution and the minimalist program: The origins of syntax. In *The Biolinguistics Enterprise: New perspectives on the evolution and nature of the human language faculty*, ed. C. Di Sciullo, A. M. Boeckx, 461–491. Oxford: Oxford University Press.

Chaves, Rui P. 2020. What don't rnn language models learn about filler-gap dependencies? 3.

Chen, S. F., and J. Goodman. 1999. An Empirical Study of Smoothing Techniques for Language Modeling. *Computer Speech Language* 13:359–394.

Chomsky, Noam. 1959. Review of *Verbal Behavior* by B.F. Skinner. *Language* 35:26.

Chomsky, Noam. 1980. *Rules and representations*. New York: Columbia University Press.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, Massachusetts: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–156. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In *Structures and beyond: The cartography of syntactic structures*, ed. Adriana Belletti, volume 3, 104–131. Oxford: Oxford University Press.

Chomsky, Noam. 2020. The UCLA Lectures. URL `https://ling.auf.net/lingbuzz/005485`.

Chomsky, Noam. 2023. Genuine Explanation and the Strong Minimalist Thesis. *Cognitive Semantics* 8:347–365.

Chomsky, Noam, Ian Roberts, and Jeffrey Watumull. 2023. The False Promise of ChatGPT. *The New York Times* URL `https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html`.

Christiansen, M. H., and N. Chater. 2016. *Creating Language: Integrating Evolution, Acquisition, and Processing*. Cambridge, MA: MIT Press.

Contreras Kallens, Pablo, Ross Deans Kristensen-McLachlan, and Morten H. Christiansen. 2023. Large Language Models Demonstrate the Potential of Statistical Learning in Language. *Cognitive Science* 47.

Crain, Stephen, and Janet Dean Fodor. 1985. *How can grammars help parsers?*, 94128. Studies in Natural Language Processing. Cambridge University Press.

Dąbrowska, Ewa. 2015. What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology* 6.

Elman, J. L., E. A. Bates, M. H. Johnson, A. Karmiloff-Smith, D. Parisi, and K. Plunkett. 1996. *Rethinking Innateness: A Connectionist Perspective on Development*. Cambridge, MA: MIT Press.

Giblin, Iain, Peng Zhou, Cory Bill, Jiawei Shi, and Stephen Crain. 2019. The Spontaneous eMERGEnce of Recursion in Child Language. In *Proceedings of the 43rd Boston University Conference on Language Development*, ed. Megan M. Brown and Brady Dailey, 270–285. Somerville, MA: Cascadilla Press.

Goldberg, A. 2019. *Explain me this*. Princeton, NJ: Princeton University Press.

Kam, Xuân-Nga Cao, Iglika Stoyneshka, Lidiya Tornyova, Janet D. Fodor, and William G. Sakas. 2008. Bigrams and the richness of the stimulus. *Cognitive Science* 32:771–787.

Lieven, Elena. 2014. First language development: a usage-based perspective on past and current research. *Journal of Child Language* 41:48–63.

Mitchell, Jeff, and Jeffrey Bowers. 2020. Priorless recurrent networks learn curiously. 5147–5158. Barcelona, Spain (Online): International Committee on Computational Linguistics. URL `https://aclanthology.org/2020.coling-main.451`.

Pearl, Lisa, and Jon Sprouse. 2013. *Computational models of acquisition for islands*, 109131. Cambridge: Cambridge University Press.

Pérez-Leroux, Ana Teresa, Anny P. Castilla-Earls, Susana Bejar, and Diane Massam. 2012. Elmo's sister's ball: The problem of acquiring nominal recursion. *Language Acquisition* 19:301–311.

Progovac, Ljiljana. 2019. *A Critical Introduction to Language Evolution: Current Controversies and Future Prospects*. Cham: Springer.

Pullum, G., and B. Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19:9–50.

Reali, Florencia, and Morten H. Christiansen. 2005. Uncovering the richness of the stimulus: Structure dependence and indirect statistical evidence 29:1007–1028.

Ross, John. 1967. Constraints on variables in syntax. Doctoral Dissertation, Massachusetts Institute of Technology.

Satık, Deniz. 2022. The Strong Minimalist Thesis is too strong: syntax is more than just Merge. *Biolinguistics* 16.

Shi, J., and P. Zhou. 2018. How possessive relations are mapped onto child language: A view from mandarin chinese. *Journal of Psycholinguistic Research* 47:1321–1341.

Skinner, B. F. 1957. *Verbal Behavior*. Princeton, NJ: Prentice-Hall.

Stowe, Laurie A. 1986. Parsing WH-constructions: Evidence for online gap location. *Language and Cognitive Processes* 1:227–245.

Terunuma, Akiko, Miwa Isobe, Motoki Nakajima, Reiko Okabe, Shunichiro Inada, Sakumi Inokuma, and Terue Nakato. 2017. Acquisition of Recursive Possessives and Locatives within DPs in Japanese. In *Proceedings of the 41st Annual Boston University Conference on Language Development*, ed. Maria LaMendola and Jennifer Scott, 626–636. Somerville, MA: Cascadilla Press.

Tomasello, Michael. 2009. The usage-based theory of language acquisition. 69–87. Cambridge, MA: Cambridge University Press.

Wilcox, Ethan, Roger Levy, Takashi Morita, and Richard Futrell. 2018. What do RNN language models learn about filler–gap dependencies? 211–221. Brussels, Belgium: Association for Computational Linguistics. URL `https://aclanthology.org/W18-5423`.

Wilcox, Ethan Gotlieb, Richard Futrell, and Roger Levy. 2023. Using Computational Models to Test Syntactic Learnability. *Linguistic Inquiry* 1–44.

Wilcox, Ethan Gotlieb, Roger Levy, and Richard Futrell. 2019. What syntactic structures block dependencies in RNN language models?