Universität Stuttgart

Foundational Questions Regarding LLMS

https://www.ims.uni-stuttgart.de

Winter Semester 2024-2025
Lecturer: Prof. Dr. Jonas Kuhn

Institut für **Maschinelle Sprachverarbeitung**

Bridging LLMs and Linguistics: Towards a Unified Framework for Language Understanding - *A review of recent work*

Yassir El Attar
3751369
st191841@stud.uni-stuttgart.de

Abstract

The recent success of Large Language Models (LLMs) has not only thrived in specific NLP tasks, but also reached various areas inspiring more innovative research and study. However, one of the most related fields that should have always been integrated with these models is linguistics since their core is the same: human language. Yet, the study of LLMs and linguistics has been, to an extent, treated as separate disciplines. In this paper, I try to shed light on the intersection of the two and how the advancements in artificial intelligence, in general, and LLMs, in particular, have imposed a challenge to such separation and to traditional linguistic approaches. Due to the limit of this work, we try to focus mainly on two approaches in linguistics; Generative Grammar and Construction Grammar as a starting point, which was inspired by the debate that emerged following the work of Piantadosi (2024), which proposes a brutal dismissal and refutation of Chomsky's approach due to the advances and success of LLMs. But since the goal here is to find an approach that can integrate both rather than one dismissing the other, we also look at works that showed success in such integration in both Generative Linguistics and CxG (Construction Grammar). First, drawing from Piantadosi (2024) and Chesi (2024), we explain how LLMs have not only surpassed many cornerstone claims by traditional linguistic theories, but also provide empirical evidence of such claims. Second, we look at the possibilities of the integration of such fields and how it would be a more fruitful path forward to understanding human language with a framework that bridges computational and theoretical linguistics, allowing for a mutual refinement of both fields. The two papers mentioned earlier, Piantadosi (2024) and Chesi (2024) focus on the end, in a way, of Generative Grammar, mainly PoS (Poverty of Stimulus) hypothesis, UG (Universal Grammar), and Minimalism, but a similar work on CxG does not, to my knowledge, exist. However, there have been some interesting and remarkable studies of the integration of both (CxG and LLMs/ artificial neural nets), such as Weissweiler et al. (2023) and Madabushi et al. (2024). Both provide new insights on how such integration would produce unprecedented success for the fields.

1 Introduction

Chomsky's approach to language has been one of the most dominant linguistic theories for decades. Since its emergence in the mid-20th century, core concepts of the Chomskyan framework, such as Universal Grammar (UG), have gained remarkable attraction and have been widely adopted by linguists across the globe. His theory has had a transformative impact not only on linguistics but also on other key domains such as psychology and cognitive science. The contributions of his theory are undeniably influential in understanding and modeling human language. Chomsky challenged earlier theories like behaviorism (e.g., those of B.F. Skinner) in language acquisition and paved the way for a novel model grounded in ideas like innate grammar, Government and Binding, Transformational Grammar, the Poverty of the Stimulus (PoS) hypothesis, Minimalism, and other foundational subtypes of Generative Grammar.

On one hand, Generative Linguistics, in general, seeks to provide —or rather discover— a 'generative' model that accounts for how language is understood, produced, and acquired. Such a model uses specific rules and algorithms to generate a new language, such as the Merge theory which was nicely described by Stabler (1997). The system must not generate sentences that violate these rules. In other words, the grammar of a language L should produce an infinite number of sentences judged as grammatically correct in L, while preventing ungrammatical ones. For example, using a simple phrase-structure grammar in Figure 1 provided by Portelance and Jasbi (2024), the sentence "*The old cat wore a hat*" is considered grammatical, whereas "*the wore cat hat the old*" cannot be derived using the same grammar and is thus ungrammatical. However, this paper does not aim to explain Generative Grammar in full details, but rather explore how it may be integrated with modern Large Language Models (LLMs) and how such rules and models could benefit both disciplines.

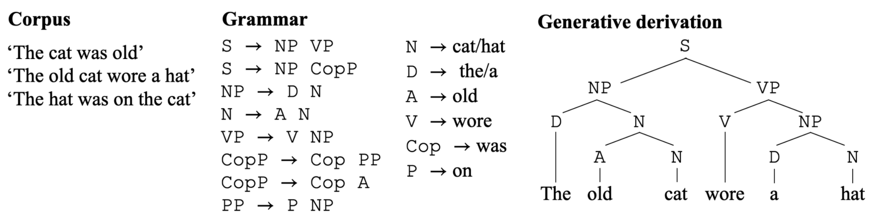

Figure 1: Example of a simple phrase-structure grammar

On the other hand, Construction Grammar (CxG) offers a distinct theoretical approach, often viewed as separate from Generative Grammar. CxG sees language as a collection of 'constructions': form-meaning pairings. While the Chomskyan approach treats syntax, semantics, and pragmatics as separate components, CxG emphasizes that constructions, ranging from single words to entire sentence patterns, are integrated units of linguistic patterns involving both form and meaning. Examples include words (“*apple*”), idioms (”*piece of cake*”), morphological forms (“-*er*”, “-*ment*”, ”*un*-”), and syntactic patterns (”*X give Y Z*”). According to CxG, if we cannot predict ’form’ or ’meaning’ of a construction from its components and cannot use it to generate other known constructions, then it is stored as a unit in memory. CxG argues that there is no rigid division between lexicon and syntax, viewing both as constructions that are learned similarly through exposure to language. This approach emphasizes frequency, analogy, and pattern recognition over an innate Universal Grammar. From this perspective, we can observe key parallels between CxG and how LLMs learn and generate language. This intersection has already become the focus of numerous studies, such as those by Weissweiler et al. (2023) and Weissweiler et al. (2024).

The success of LLMs has become a central focus in computational linguistics in recent years. Leveraging their ability to learn from massive datasets has led to major advances across many scientific disciplines. LLMs like GPT-4 are deep neural networks that do not rely on explicit grammar or semantics. Instead, they learn statistical regularities in text. Their pre-training typically involves next-word prediction (e.g., in GPT models) or masked-word prediction (e.g., in BERT), followed by fine-tuning for downstream tasks. A notable property of LLMs is their emergent abilities (Wei et al., 2022), which enable them to perform increasingly complex linguistic tasks without being explicitly programmed. These include acquiring syntactic fluency (e.g., handling long-range dependencies), semantic understanding (e.g., disambiguating meaning), pragmatic reasoning (e.g., following conversational norms, adapting style), and generalization (e.g., applying knowledge to novel scenarios). These capabilities arise from the scale and architecture of the models, not from hand-coded rules.

Piantadosi (2024) presents compelling arguments that modern language models challenge Chomsky’s foundational claims. Chesi (2024) builds on this by suggesting that we may be witnessing a turning point for generative linguistics. Unlike Piantadosi, who directly argues that LLMs undermine core generative grammar principles, Chesi offers a more cautious analysis, urging generative linguists to embrace LLMs - whether as tools or as new linguistic theories. The adoption of deep learning models (or neural networks in general) in other scientific fields has yielded transformative results, such as David Baker’s work on protein design and Demis Hassabis and John Jumper’s on protein structure prediction, awarding them a Nobel prize in Chemistry. Why shouldn’t theoretical linguistics also benefit from such innovations?

Both Piantadosi (2024) and Chesi (2024) argue that generative linguists have largely resisted language models — a stance that may now be counterproductive. Piantadosi highlights Chomsky’s long-standing skepticism, citing a 2012 interview in which Chomsky dismissed Bayesian models as ’useless’, and his 2023

New York Times op-ed(Chomsky et al., 2023) with Ian Roberts and Jeffrey Watumull, where they argue that language models differ fundamentally from human cognition. While their concerns about over-reliance are not entirely unfounded, it is clear that LLMs have already facilitated theoretical and empirical progress in multiple domains. In linguistics, recent studies by Weissweiler et al. (2023), Madabushi et al. (2024), and Zhang et al. (2024) show promising results in modeling constructions, linguistic representations, and beyond. Yet, responses to Piantadosi's work from generative linguists have often been defensive rather than collaborative. This can be seen in the reactions of Katzir (2023) and Kodner et al. (2023), which reflect a resistance to the opportunities presented by LLMs.

To answer the question, "Would integrating neural language models into the study of human language help advance linguistic theory?", this paper will focus on four key areas. First, Section 2 discusses the idea of LLMs as valid "theories of language", as proposed by Piantadosi (2024) and supported by Chesi (2024). Second, Section 3 explores core concepts of generative grammar, particularly Innate UG, the PoS argument, and Minimalism and whether they could be adapted for LLMs or vice-versa. Third, in section 4, we examine how the integration of Construction Grammar (CxG) into LLMs has already shown remarkable results and how this approach might be extended to generative linguistics. Finally, Section 5 looks beyond grammar, exploring whether syntax and semantics are distinct or intertwined — comparing insights from Generative Grammar, CxG, and LLMs.

2 Are LLMs Truly 'Good Theories' of the Science of Language?

Piantadosi (2024) presents a range of arguments in support of treating large language models (LLMs) as legitimate scientific theories, grounded in their adherence to traditional criteria for scientific theorizing. He begins by referencing prominent critiques from generative linguists—most notably Noam Chomsky—and subsequently uses these critiques to underscore the ways in which LLMs have, in some respects, outperformed generative grammar as a theoretical framework. In this section, we uncover the key claims Piantadosi advances in defense of this position, followed by a discussion of counterarguments raised by generative linguists in response. Finally, we examine whether these LLMs can be considered as good 'theories' not only in the scope of lingsuitics, but in broader fields.

2.1 Large language models as scientific theories

Similar to other scientific theories, LLMs have demonstrated remarkable success in adhering to the traditional theoretical approaches used to study specific phenomena. Piantadosi (2024) draws a comparison between how tuning a model's parameters offers insight into formalizing and comparing theories and how similar processes are used in other sciences, such as modeling hurricanes or pandemics. These models begin with a set of assumptions that allow them to make predictions; based on those predictions, their parameters are refined to improve accuracy. This mirrors the process of scientific theory development. Over time, models have evolved, from Bayesian frameworks and n-gram models to transformer architectures based on attention mechanisms, each representing a different approach to testing hypotheses about how the human mind might work. By employing these methods, researchers can identify the most effective hypotheses (i.e., models) to serve as predictors of linguistic behavior. These state-of-the-art models now approximate human language performance more closely than any previous (or any existent) alternatives.

Piantadosi (2024) further highlights the precision and formalization of LLMs, which can be implemented in actual computational systems - unlike most subfields of generative linguistics (Pullum, 1989). Because these models are implemented, they allow us to generate predictions that can be evaluated against psychological measures (Hoover et al., 2023; Shain et al., 2024). He also notes that transformer-based models "predict nearly 100% of explainable variance in neural responses to sentences" (Schrimpf et al., 2021). More broadly, Piantadosi argues that neural models can be viewed as hypotheses about cognition, which can be scientifically tested through performance evaluation, analysis of attention mechanisms, and ablation studies (Warstadt and Bowman, 2024)).

Unlike generative linguistics, which often relies on abstract and informal principles, large language models (LLMs) constitute precise, computable theories that are subject to empirical evaluation. Their internal architectures, such as attention mechanisms, have been shown to align with known linguistic structures (Pullum, 1989) and to predict neural responses observed in the brain (Manning et al., 2020; Schrimpf et al., 2021). In contrast to theories that are rarely operationalized, LLMs exhibit internal consistency and yield falsifiable predictions across domains including syntactic competence and semantic representation.

LLMs have demonstrated superior empirical performance compared to traditional generative approaches in areas such as syntax, semantics, and acceptability judgments (Warstadt et al., 2019; Gauthier et al., 2020). On standardized tests like SyntaxGym, state-of-the-art models have achieved nearly 90% accuracy, whereas generative theories offer no comparable empirical benchmarks. Furthermore, syntactic phenomena such as filler-gap dependencies and island constraints are effectively captured by LLMs (Wilcox et al., 2024), undermining critiques like those found in Chomsky et al. (2023).

In a recent study by Beguš et al. (2025), the results exhibited by the models show promising signs of forming a genuine scientific theory of language. Beguš et al. (2025) argues that LLMs demonstrate clear *metalinguistic abilities*—not merely using language, but reasoning about it, much like human linguists do. The tested models showed high accuracy in tasks such as drawing syntactic trees, identifying recursion and ambiguous structures, and even writing phonological rules for previously unseen languages—tasks typically reserved for expert linguists. Additionally, a particular model (OpenAI's o1) exhibits a strong *chain-of-thought mechanism*, which is said to mirror the kind of reasoning humans use in complex cognitive tasks. A similar feature is also observed in DeepSeek's DeepThink (R1). While this may not constitute definitive proof of their status as scientific theories, it certainly marks a meaningful step in that direction.

Chesi (2024) also clearly supports the status of large language models as linguistic theories. As he claims that while vLLMs (very Large Language Models) are overrated as linguistic theories, the methodologies and training paradigms proposed by Wilcox et al. (2024) and Warstadt et al. (2023) explain how these vLLMs are really the best theories in existence now and that they are observationally more adequate than any MG (Minimalist Grammar).

2.2 Beyond Prediction: Why LLMS fall short as scientific theories

Katzir (2023), like many other generative linguists, views LLMs as effective engineering tools but denies that they contribute to theorizing about human cognition, as claimed by Piantadosi (2024). A central argument made by generative linguists is that LLMs are not trained on data comparable in size or quality to that received by a child. Even when trained on similarly sized or much larger datasets, these models still fail to generalize in a human-like manner. Katzir (2023) refers to this as a ”necessary condition of adequacy”—a bar that LLMs, in his view, do not meet. To support this point, he tests ChatGPT on Ross's coordinate structure constraint, a well-known syntactic constraint understood by all English-speaking children, and finds that the model fails. Another significant critique he raises concerns LLMs' inability to distinguish between grammaticality and frequency: while humans can recognize the difference, LLMs rely solely on statistical likelihood from training data, equating the frequent with the grammatical.

Complementary arguments are provided by Kodner et al. (2023), who emphasize that what LLMs produce is a mere simulation of human language, not a duplication. As Guest and Martin (2023) also argues, similar outputs do not imply similar internal processes. Both human cognition and LLMs are effectively black boxes, and similar observable behavior does not imply shared underlying mechanisms. Kodner et al. conclude that while LLMs may approximate human language output, their lack of mechanistic similarity and explanatory power disqualifies them from being considered scientific theories of language. They argue that Piantadosi (2024) confuses ”prediction” with ”explanation” and ”tools” with "theory." As Popper (1959) reminds us, the role of a scientific theory is not merely to predict but to *elucidate* and *explain*—and LLMs do neither.

Additionally, the lack of interpretability in LLMs further weakens the claim that they constitute scientific theories. Although various studies have attempted to investigate the inner workings of these models (Belinkov and Glass, 2019; Tenney et al., 2019; Rogers et al., 2020; Linzen and Baroni, 2021), success has been limited due to their overwhelming complexity. Even Piantadosi (2024) concedes that "we don't deeply understand how the representations these models create work."

2.3 Summary of the section

Piantadosi (2024) argues that large language models (LLMs) should be regarded as genuine scientific theories of language. He maintains that, like models used in other scientific fields, LLMs formalize assumptions, generate predictions, and are subject to empirical testing. Their internal mechanisms, such as attention and layers, can be evaluated against human linguistic behavior, neural data, and psychological measures. Piantadosi emphasizes that these models go beyond engineering tools by offering computationally implemented, falsifiable hypotheses about how language works. He also cites their ability to capture grammatical structures without explicitly programmed rules as evidence of their theoretical relevance. Another remarkable point that he makes is regarding how linguistic corpora can be seen as a low-dimensional projection of syntax and thought, and therefore, by studying such low-dimensional projection, we can recover at least some aspects of the cognitive systems in a similar way to embedding theorems in mathematics.

However, Katzir (2023) and Kodner et al. (2023) reject the notion that LLMs qualify as scientific theories. They argue that while LLMs may simulate human-like language behavior, they do not duplicate the cognitive processes underlying it. Drawing on the concept of multiple realizability, they stress that similar outputs do not imply similar mechanisms. Furthermore, LLMs rely on massive datasets and biologically implausible learning algorithms like backpropagation, which differ from how humans (children) acquire language. Kodner et al. (2023) also highlight that most state-of-the-art LLMs are non-transparent, violating core principles of scientific transparency and reproducibility. They argue that Piantadosi confuses prediction with explanation, citing Popper (1959)'s view that scientific theories must elucidate mechanisms, not just fit data. Ultimately, they see LLMs as powerful tools, not explanatory frameworks.

3 Generative Grammar vs. Large Language Models

The debate sparked by Piantadosi's publication (Piantadosi, 2024) shook the foundations of generative linguistics. In this context, Cristiano Chesi's work (Chesi, 2024), titled "Is it the end of (generative) linguistics as we know it?", further intensifies the discussion. While the question may provoke even more skepticism among linguists, Chesi's aim is not to declare an 'end', but to issue a wake-up call: generative linguists should move beyond their skepticism and acknowledge the success of LLMs integration. In the following sub-sections, we highlight some key components of generative linguistics and explore how LLMs challenge these foundational concepts. Between Piantadosi's critiques and Chesi's warnings, we examine why a refinement of generative linguistics may be necessary in light of the rise of large (neural) language models.

3.1 To what extent is the Innateness of Grammar necessary to learn language?

If you ask ChatGPT-4o, for instance, about the definition of Universal Grammar under the scope of Generative Linguistics, you get a response as below:

> "Universal Grammar (UG), in the generative grammar tradition, refers to the idea that humans are born with a built-in set of linguistic principles that guide language acquisition. Proposed by Chomsky, UG is thought to explain how children learn complex grammar rapidly and uniformly across cultures, despite limited input (the "Poverty of the Stimulus"). It posits that core grammatical structures are hardwired into the brain, enabling the rapid development of language in early childhood."

which aligns very well with the general definition of UG. However, LLMs learn such complex grammar without any pre-built-in rules or any innate syntactic modules as claimed by Generative Grammar approach. LLMs learn hierarchical structures, movement, recursion, and many other rules without the need of any prior knowledge, but from large-scale text data alone as shown in the work of Manning et al. (2020), Warstadt and Bowman (2020), Linzen and Baroni (2021), and many others.

Nevertheless, Piantadosi (2024) does not agree with such a definition and 'refutes' specific aspects of Universal Grammar claimed to be innate (Chomsky, 1965), including aux-inversion, wh-movement, and other syntactic structures. Piantadosi argues that LLMs acquire such structures without any need for predefined rules, relying solely on training data. For instance, unlike the UG principle which states that an interrogative sentence is derived from a declarative one, LLMs do not use such an assumption. Instead, they learn both sentence types directly from data and are able to acquire hierarchical structures. In the same line of argument, Chesi (2024) demonstrates that vLLMs (very Large Language Models) are more robust and descriptively adequate than any MG (Minimalist Grammar), as they are capable of modeling re-analysis strategies to recover from ill-formed inputs; A task MG fails to handle adequately.

The hierarchical structure of language, often emphasized by generative linguists as a principle built into the human mind, is a crucial concept in this discussion. Yet, despite the criticisms from Chomsky and many other generativists who argue that language models cannot learn such hierarchy, recent research suggests otherwise. Alongside Piantadosi (2024)'s arguments, Portelance and Jasbi (2024) discusses how not only modern models (e.g., LLMs), but also early models (e.g., Bayesian learners), can learn hierarchical patterns even from relatively small datasets.

Supporting this claim, Beguš et al. (2025) provide empirical evidence, particularly through chain-of-thought reasoning, which challenges prior claims such as those by Şahin et al. (2020), who argue that LLMs lack the ability for iterative reasoning, and Katzir (2023), who states that pre-prompting does not improve performance and that "further time and resources are of no use to ChatGPT."

Beguš et al. (2025)'s empirical studies demonstrate impressive capabilities—extending beyond linguistics to cognitive reasoning—such as generating syntactic trees for ambiguous sentences, handling recursion, syntactic movement, and even deriving phonological rules. Their work tests various LLMs (versions of GPT and LLaMA models), and the results are shown in Figure 2, which illustrates performance across four different systems. (I cited their results here for more clarity on the empirical evidence.)

3.2 Poverty of Stimulus: Is it still a valid hypothesis in the light of LLMs?

From the concepts of Innate and Universal Grammar, we move to one of the central cornerstones of generative linguistics: the Poverty of the Stimulus (PoS) hypothesis, which aims to explain how children, under normal conditions, are able to acquire adult-like linguistic competence. The term "poverty" refers to both the limited quantity and quality of linguistic data available to the child. The idea is that, despite this limited input, children still manage to acquire complex grammatical knowledge.

Chesi (2024) challenges this hypothesis by distinguishing between two interpretations: a stronger and a weaker one. The stronger version claims that adult linguistic competence cannot be achieved regardless of the amount of input; the weaker version suggests that adult-like competence is only possible after an "unreasonable" amount of exposure. Chesi argues that many generative linguists now lean toward the weaker interpretation. His main point is that when children are exposed only to positive evidence (i.e., grammatical sentences), which is typically the case, only trivial finite grammars are technically learnable—an idea rooted in classic learnability theory. However, modern machine learning models

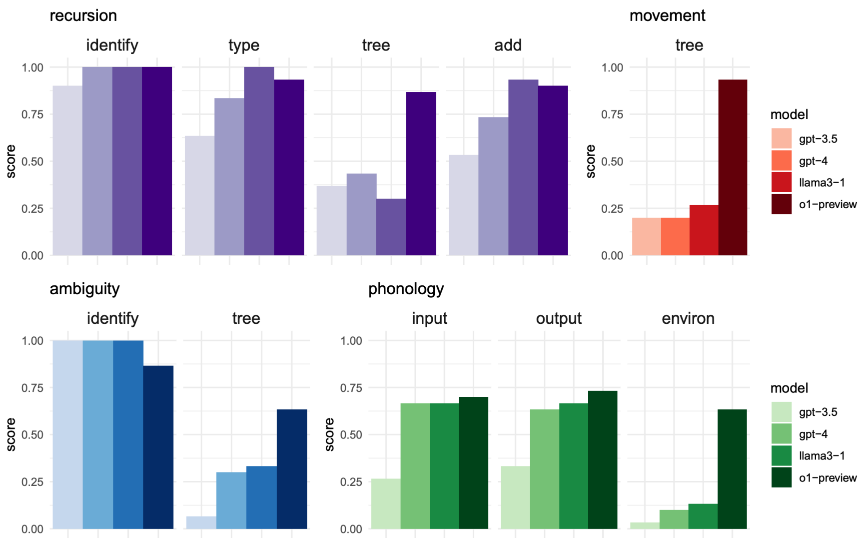

Figure 2: Performance of the four models (gpt-3.5-turbo-0125, gpt-4-0314, meta.llama3-1-405b-instruct-v1, o1-preview-2024-09-12) on four tasks (ten subtasks in total). In the bottom left-hand quarter, the ambiguity task: (i) is the sentence ambiguous? and (ii) draw syntactic trees illustrating the ambiguity. In the upper left-hand quarter, the recursion task: (i) is the sentence recursive?, (ii) what type of recursion is it?, (iii) draw a syntactic tree representing recursion, and (iv) add a layer of recursive structure of the same type. In the upper right-hand quarter, the movement task: illustrate movement with a syntactic tree. In the bottom right-hand quarter. the phonology task: identify the phonological rule's (i) input, (ii) output, and (iii) environment.

(Beguš et al., 2025)

employ similar input settings via self-supervised learning, rather than needing explicit ungrammatical input. Chesi explains that through self-supervision, models like Elman's SRNs and modern Transformers provide a solution that is (a) easy to implement, (b) psycholinguistically and cognitively plausible, (c) effective in capturing non-local dependencies, and (d) capable of making both gradual and predictive judgments (Chesi, 2024). This approach provides a learning mechanism that directly addresses and challenges the assumptions of the PoS hypothesis.

Piantadosi (2024) extends this critique by arguing that the PoS hypothesis overstates the need for innate constraints. While generative linguistics claims that language cannot be learned without them, Piantadosi counters that "learning without constraints is not only possible, it has been well-understood and even *predicted*." He further supports this claim by showing that LLMs acquire complex syntactic structures like auxiliary inversion and recursion purely from data, without the need for any hardwired/ programmed grammatical rules.

Going back to the recent empirical work by Beguš et al. (2025), it also confirms that modern LLMs display human-like preferences and generalizations in tasks involving hierarchical syntax and ambiguity resolution—despite being trained solely on positive evidence. These findings further weaken the PoS hypothesis by demonstrating that exposure to surface-level linguistic data, when processed through powerful learning architectures, may indeed be sufficient.

3.3 Summary of the section

The emergence of Large Language Models (LLMs) has sparked significant debate within generative linguistics, particularly regarding foundational concepts like Universal Grammar (UG) and the Poverty of the Stimulus (PoS) hypothesis. Piantadosi (Piantadosi, 2024) and Chesi (Chesi, 2024) argue that LLMs challenge the necessity of innate grammatical structures. Unlike the generative view—which posits that features like aux-inversion and wh-movement are hardwired—LLMs acquire such structures directly from large-scale textual input, without built-in syntactic modules. Chesi further notes that LLMs, through reanalysis and error recovery, outperform Minimalist Grammars in descriptive adequacy.

The hierarchical nature of language, often considered an innate human capacity, has also been learned by LLMs through data alone, as demonstrated in works by Manning et al. (2020), Portelance and Jasbi (2024), and Beguš et al. (2025). These models exhibit complex reasoning abilities, including handling recursion, ambiguity, and phonological rule derivation. Moreover, PoS is increasingly questioned: if language cannot be acquired from positive evidence alone, how do LLMs succeed? Chesi and Piantadosi highlight how self-supervised learning replicates core learning mechanisms without negative evidence. This suggests that key generative assumptions—like the unlearnability of syntax without innate biases—may be overstated.

The challenges disccusses in this review span more than just the innateness of grammar and PoS hypothesis. Chesi (2024) demonstrates a well-detailed critique of core ideas of language theory such as explanatory and descriptive adequacy, computational efficiency, competence and performance, and formalization according to Generative Grammar and provides an alternative approach through using LLMs and intagrating them.

Together, these findings imply that generative principles such as modularity, rule-based derivations, minimalisim, and input poverty may need revision or reinterpretation in light of empirical evidence from modern AI systems.

4 Construction Grammar and Neural Language Models

Construction Grammar (CxG) conceptualizes grammar as a network of learned form-meaning pairings (constructions) that range from individual morphemes to complex syntactic patterns (Goldberg, 1995; Weissweiler et al., 2023, 2024). This approach contrasts with traditional generative frameworks by emphasizing the role of usage, semantics, and schematicity in language understanding. With the rise of large pre-trained language models (LLMs) like BERT and GPT, researchers have begun to explore whether and how these models capture constructional knowledge. Studies have shown that LLMs can distinguish between different constructions and often encode constructional structure implicitly, particularly in tasks involving argument structure or idiomatic expressions (Madabushi et al., 2024). Other work suggests that probing LLMs through a constructional corpus reveals both their potential and their current limitations in modeling the semantic richness of constructions (Weissweiler et al., 2023). These findings highlight a growing cooperation between CxG and computational linguistics, offering new methods for enhancing natural language understanding by grounding it in constructional representations (Xu et al., 2023).

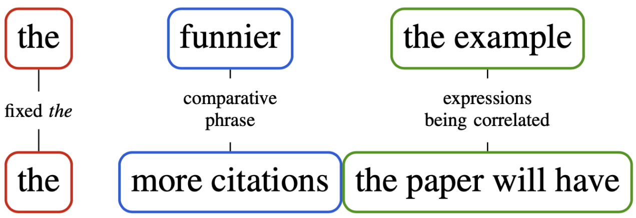

Figure 3: An example illustrating the complexity of a construction. It is an instance of the English Comparative Correlative (CC), with its syntactic features highlighted above the text and paraphrases illustrating its meaning below.

(Weissweiler et al., 2023)

4.1 How beneficial is it to make use of CxG theories to develop Neural language models' ability to generate text?

Weissweiler et al. (2023) asks a trivial question in this context: Do "PLMs (Pretrained Language Models) model constructions as gestalts in both form and meaning". He also compares the performance on probing tasks and the lingsuitic competency of these models in both GG (Genertive Grammar) and CxG (Construction Grammar). This would open up intersting discussions as the models might be used ad cognitive models if we could show that LMs conform better to CxG rather than GG. In their work, they discuss and evaluate previous research that dealt with this topic and draw on them new insights toward more understanding of language through integration of CxG in LMs tasks. Madabushi et al. (2020) find that if models were trained on sentences that were sorted into documents based on their constructions do not reliably perform better than those trained on typical data (original-unsorted). From a different approach, the work by Li et al. (2022) provides insights into how PLMs prefer sorting by construction as the size of their training data increases in a similar way to the human data, which is sorted by the English proficiency. Li et al. (2022) probe their model based on the idea that if construction information is picked up by the model, the contextualized embedding of the 'word' (e.g., verb) should acquire some constructional meaning, which would bring it closer to the corresponding prototypical 'word' meaning than the rest. However, Weissweiler et al. (2023) claim that constructions could be intrinsically difficult for language models because they include non-compositional meaning that is not attached to a token. Hence, they suggest that the only place where constructional information could be stored is the models' weights, but these are very difficult to examine.

Other work focused on improving the language representation and models performance with constructional information for NLU (Natural Language Understaning) and other NLP (Natural Language Processing) tasks. We dive into the work presented by Xu et al. (2023) here to see how such constructional information could help in enhancing the language representations for different models. They perform extensive experiments which demonstrate HyCxG's (hypergraph network of construction grammar—*their framework*) strong performance on NLU tasks, with multilingual results confirming the value of constructional information across languages. The model's constructional representations also support building a construction network, offering insights that advance usage-based CxG. HyCxG (Xu et al., 2023) enhances language representation by outperforming baselines on ABSA and GLUE tasks, including key gains on CoLA, RTE, and MRPC. It proves effective even with reduced model complexity and shows strong multilingual performance. Its Cond-MC strategy improves construction selection, and its construction network offers insights for usage-based CxG. Figure 4 is retrieved from the same paper and it displays interesting distinction in different constructions representations between BERT and their model HyCxG.

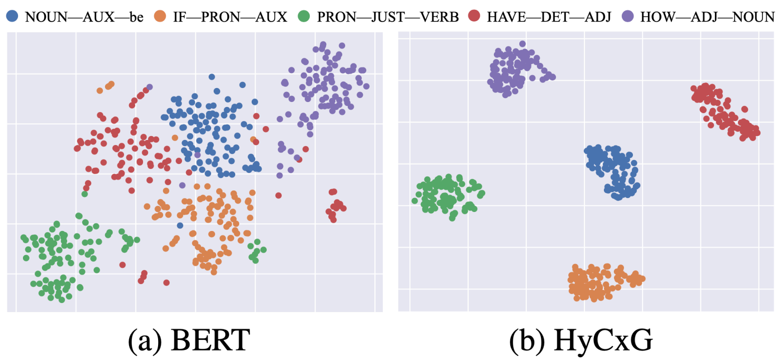

Figure 4: 2D t-SNE plot of construction representations.

(Xu et al., 2023)

Now, moving from how constructional information enhance language representations, we also unveil another work (Weissweiler et al., 2024) that uses a hybrid corpus (human and LLMs) of constructions and evaluate the LLMs on rare phenomena in linguistics. First, they collected their evaluation data using a hybrid pipeline combining automated tools and human annotation, which allowed for efficient large-scale annotations for rare linguistic phenomena. The steps involved here are very interesting and, even if not considered a strong point for the objective of this paper, it shows a valid use of the integration of LLMs in linguistic studies. They following the following strategy:

1. **Dependency parsing** was first applied to a large Reddit corpus to identify syntactic patterns associated with the caused-motion construction (CMC).
2. **GPT-3.5** was then used for **few-shot classification** to further filter likely CMC instances.
3. **Human annotators verified the filtered results**, resulting in a high-quality hand-annotated dataset of 765 sentences.
4. They then extrapolated this to a semi-automatically labeled corpus of 127,955 sentences using recurring 4-tuples of key sentence components (verb, direct object, preposition, prepositional object).

The evaluation showed that all tested LLMs struggled to understand the caused-motion construction (CMC), particularly in non-prototypical cases. Even the best-performing model, Mixtral 8x7B (instruction-tuned), had an error rate exceeding 30%. Surprisingly, GPT-4 underperformed compared to smaller open models like Mistral and LLaMA2. These results reveal significant limitations in LLMs' ability to capture the semantic nuances of rare and complex constructions (Weissweiler et al., 2024). However, this integration of studying such phenomena in linguistics is truly an inspiring path for more research.

A recent work by (Madabushi et al., 2024) reviews multiple work in the scope of Construction Grammar and Language Models. While there is improvements to focus on, they provide interesting results of such interaction. One key finding is that was revealed by Madabushi et al. (2020) is that certain pre-trained large language models, like ERT and RoBERTa, show evidence of capturing constructional information despite not being explicitly trained for it. The experiments conducted (Madabushi et al., 2024, 2020) explains that these models can distinguish between sentences that share the same construction and those that don't, with high accuracy even when trained on very few examples. They also urge to make use of previous work (Bergen and Chang, 2003; Janda et al., 2020; Feldman, 2022) in CxG for a successful integration of LLMs with CxG for mutual benefits: LLMs help uncover and analyze constructions at scale, while CxG insights guide the interpretation and evaluation of what LLMs "know" about language. Additionally, they also showcase how such interaction and collaboration supports usage-based theories of language acquisition, showing how language patterns can emerge from data exposure— mirroring human learning.

4.2 Summary of the section

CxG studies in cooperation with LLMs is a growing field of interest for many linguistics and language models researchers alike. CxG, which treats grammar as learned form-meaning pairings, aligns well with LLMs' data-driven learning from large corpora. Xu et al. (2023) demonstrate that injecting constructional information into models (HyCxG) significantly improves performance on NLU tasks across languages, especially in tasks like sentiment analysis and linguistic acceptability. Other works confirm that LLMs can be systematically evaluated on rare constructions, revealing that even top-performing models like GPT-4 still struggle with nuanced, non-prototypical uses, underscoring gaps in constructional understanding.Weissweiler et al. (2023) show that LLMs can implicitly learn constructional patterns and advocate for using CxG to better probe and interpret LLMs. Similarly, Dunn (in several works (Dunn, 2017, 2022)) argues that LLMs can model construction networks and support usage-based grammar, offering a scalable way to study how constructions emerge and function in real data and how to extract them. Putting everything together, we can clearly see that bridging LLMs and Linguistics (CxG in this case) is beneficial for both areas of research. They only require enthusiastic and motivated scholars to develop even more advanced models and to build a theory that is close to understanding both human language and mind.

5 Separationg of Syntax and Semantics: In light of LLMs, CxG, and Generative Grammar

The classical distinction between syntax and semantics has long shaped linguistic theory, with generative grammar emphasizing an autonomous syntactic module, largely independent of meaning. However, emerging perspectives—particularly from usage-based models like Construction Grammar (CxG) and recent advances in large language models (LLMs)—challenge this separation. Piantadosi (2024) argues that modern LLMs, despite lacking explicit symbolic syntactic representations, exhibit linguistic behavior that undermines core claims of generative grammar, such as the necessity of innate syntactic structures. These models suggest that syntax and semantics may be more deeply intertwined, learned through usage and probabilistic pattern recognition rather than innate rules. This section explores how LLMs and CxG jointly call into question the generative separation of syntax from semantics, offering a more integrated and data-driven view of language competence.

5.1 Autonomy of Syntax in Generative Grammar vs. Form-Meaning pairings in Construction Grammar

On one hand, Generative Grammar, particularly in Noam Chomsky's framework, treats syntax as an independent system that operates separately from semantics (Chomsky, 1965). In this view, syntax generates structurally correct sentences, while semantics acts as a secondary system that assigns meaning to those forms. A core idea is the "syntax-first" approach: meaning is derived from structure, not the other way around (Chomsky, 1993). The theory also emphasizes modularity, treating syntax, semantics, and phonology as distinct components of language. Underlying it all is Universal Grammar (UG), the idea that syntax follows innate, biologically grounded rules (Chomsky, 2015). As Chomsky famously argued, syntactic structure shapes meaning, not the reverse—meaning doesn't dictate how sentences are formed.

On the other hand, Construction Grammar (CxG) flips the script on traditional syntax by arguing that language is built from constructions—learned pairings of form and meaning Goldberg (1995). Unlike modular theories, CxG sees syntax and semantics as inseparable; you can't have one without the other. These constructions aren't abstract rules but emerge from real language use, making grammar a product of how we actually speak and write (Goldberg, 2005). Even seemingly rigid grammatical patterns carry meaning—nothing is purely structural (Fillmore et al., 1988). As Goldberg puts it, constructions are the building blocks of language, not just combinations of smaller, meaningless pieces (Goldberg, 2005).

5.2 LLMs representations of semantics-syntax and human's cognition similarity

As Piantadosi (2024) states in his work: "*... a key feature of these models is that they integrate semantics and syn- tax. The internal representations of words in these models are stored in a vector space, and the locations of these words include not just some aspects of mean- ing, but properties that determine how words can occur in sequence (e.g. syntax)*". This claim is not supported by his own research alone, but others have revealed similar assumptions and tested them. Other work such as Tjuatja et al. (2023) tested how well language models like BLOOM, GPT-2, and GPT-3 understand 'agentivity' (who does what in a sentence) while mainting a semantic-syntax correlation. GPT-3 (davinci-003) performed best—even matching human judgment—but model size alone didn't explain its success. Its mistakes were often on tricky cases that confuse people too, suggesting LMs could help discover new linguistic patterns that embody both structure and meaning.

However, another interesting work Huang et al. (2021), while not on the exact representations of semantics in LMs, approached the question at hand by creating a model fo semantic sentence embedding (ParaBART). They create a model that learns to ”disentangle semantics and syntax in sentence embeddings from pre-trained language models. Their experiments show that their semantic sentence embeddings yield strong performance on unsupervised semantic similarity tasks” (Huang et al., 2021).

Similarity in human and LMs' representations of meaning, LMs' have become a central point for research and study. While the both can be approached differently, many argue that they should be studied in parallel. For instance, Cai et al. (2024) discuss the large language models (LLMs) like ChatGPT and Vicuna that predict text word-by-word based on vast training data (Brown et al., 2020; Ouyang et al., 2022). Surprisingly, their experiments showed they often mimic human language processing—ChatGPT matched human patterns in 10/12 psycholinguistic tasks, Vicuna in 7/12. Both models linked word forms to meanings (e.g., sound-symbolism), adapted to recent context (priming effects), and tracked dialogue partners, suggesting they build flexible, context-sensitive representations (Ethayarajh, 2019). However, neither model shortened predictable words (e.g., ”math” vs. ”mathematics”) or resolved syntactic ambiguities using context, unlike humans. ChatGPT's edge over Vicuna likely stems from its larger size and human feedback training. While LLMs can approximate human-like language use, their mistakes—like overlooking implausible sentences—reveal fundamental differences. They're powerful tools for generating hypotheses or modeling linguistic effects, but their outputs lack intent or true understanding (Bender and Koller, 2020). Yet, this encourages more refined research to push the field forward for more success. Moreover, the study of semantics, while should be incorprated with syntax, it is still a challenge even in human cognition. For instance, the study of semantic abstraction can be modelled differently for various types of semantic representations (i.e., hypermyny, concrete vs. abstract concepts...) in different modalities (im Walde and Frassinelli, 2022).

6 Conclusion

The rise of large language models has sparked a crucial debate in linguistics: Can traditional theories like Generative Grammar coexist with—or even benefit from—modern AI? This paper explored that question by examining how LLMs challenge long-standing assumptions, particularly in syntax-semantics separation, innate grammar, and the Poverty of Stimulus hypothesis. While Generative Grammar insists on modular, rule-based systems, LLMs and Construction Grammar suggest a more integrated, usage-driven view of language—one where form and meaning are learned together from data.

In spite of the differences in how LLMs or AI models and humans learn language, these models show surprising similarities to how humans use language. They can understand complex sentences, pick up on subtle meanings, and even make some of the same mistakes people do. This suggests that maybe language isn't as separate from general learning as some theories claimed. Construction Grammar, which sees language as learned patterns of form and meaning, seems to fit well with how these models actually work.

What's most interesting is that we don't have to choose one approach over the other. Instead of seeing AI as a threat to linguistic theories, we can use it as a powerful tool or a theory (depending on where you stand). These models give us new ways to test old ideas and discover new patterns in language. They can help us ask better questions: How much of language really needs to be "pre-programmed" in the brain? Or do we actually need such pre-programmed rules? How do form and meaning actually connect? What makes human language unique? Could it be extend it to study more variants of communcation systems beyond human's?

The future of linguistics might involve working with these AI systems, not against them. By combining the deep theoretical knowledge from linguistics with the data-crunching power of language models, we could make real progress in understanding both how language works and how we learn it. This partnership could lead to better language education, improved AI systems, and ultimately, a clearer picture of what makes human language so special and powerful. It could even lead to better understanding of human cognition and brain functionalities.

Rather than arguing about which theory is "right," the smartest path may be finding ways that different approaches can learn from each other. The path forward isn't dismissal but collaboration. By integrating computational power with linguistic insights, we might finally bridge the gap between how machines process language and how humans experience it—ultimately enriching both fields.

References

Beguš, G., Dąbkowski, M., and Rhodes, R. (2025). Large linguistic models: Investigating llms' metalinguistic abilities.

Belinkov, Y. and Glass, J. (2019). Analysis methods in neural language processing: A survey. *Transactions of the Association for Computational Linguistics*, 7:49–72.

Bender, E. M. and Koller, A. (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. In Jurafsky, D., Chai, J., Schluter, N., and Tetreault, J., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.

Bergen, B. and Chang, N. (2003). Embodied construction grammar in simulation-based language understanding. *Construction grammars: Cognitive grounding and theoretical extensions*.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language models are few-shot learners.

Cai, Z. G., Duan, X., Haslett, D. A., Wang, S., and Pickering, M. J. (2024). Do large language models resemble humans in language use?

Chesi, C. (2024). Is it the end of (generative) linguistics as we know it?

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA.

Chomsky, N. (1993). *Lectures on Government and Binding: The Pisa Lectures*. De Gruyter Mouton, Berlin, New York.

Chomsky, N. (2015). *The Minimalist Program*. The MIT Press, 20th edition. Accessed: 2025-04-04.

Chomsky, N., Roberts, I., and Watumull, J. (2023). The false promise of ChatGPT. *The New York Times*. Opinion piece.

Dunn, J. (2017). Computational learning of construction grammars. *Language and Cognition*, -1:1–39.

Dunn, J. (2022). Replication data for: Exposure and emergence in usage-based grammar: Computational experiments in 35 languages.

Ethayarajh, K. (2019). How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings. In Inui, K., Jiang, J., Ng, V., and Wan, X., editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, Hong Kong, China. Association for Computational Linguistics.

Feldman, J. (2022). *Advances in Embodied Construction Grammar*, pages 147–167.

Fillmore, C. J., Kay, P., and O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The case of let alone. *Language*, 64(3):501–538. Accessed: 2025-04-04.

Gauthier, J., Hu, J., Wilcox, E., Qian, P., and Levy, R. (2020). SyntaxGym: An online platform for targeted evaluation of language models. In Celikyilmaz, A. and Wen, T.-H., editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 70–76, Online. Association for Computational Linguistics.

Goldberg, A. (2005). *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press, Chicago, IL.

Guest, O. and Martin, A. E. (2023). On logical inference over brains, behaviour, and artificial neural networks. *Computational Brain Behavior*, 6(2):213–227.

Hoover, J. L., Sonderegger, M., Piantadosi, S. T., and O'Donnell, T. J. (2023). The plausibility of sampling as an algorithmic theory of sentence processing. *Open Mind*, 7:350–391.

Huang, J. Y., Huang, K.-H., and Chang, K.-W. (2021). Disentangling semantics and syntax in sentence embeddings with pre-trained language models. In Toutanova, K., Rumshisky, A., Zettlemoyer, L., Hakkani-Tur, D., Beltagy, I., Bethard, S., Cotterell, R., Chakraborty, T., and Zhou, Y., editors, *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1372–1379, Online. Association for Computational Linguistics.

im Walde, S. S. and Frassinelli, D. (2022). Distributional measures of semantic abstraction. *Frontiers in Artificial Intelligence*, 5.

Janda, L., Endresen, A., Zhukova, V., Mordashova, D., and Rakhilina, E. (2020). How to build a constructicon in five years: The russian example. *Belgian Journal of Linguistics*, 34:162–175.

Katzir, R. (2023). Why large language models are poor theories of human linguistic cognition: A reply to piantadosi. *Biolinguistics*, 17:e13153.

Kodner, J., Payne, S., and Heinz, J. (2023). Why linguistics will thrive in the 21st century: A reply to piantadosi (2023).

Li, B., Zhu, Z., Thomas, G., Rudzicz, F., and Xu, Y. (2022). Neural reality of argument structure constructions. In Muresan, S., Nakov, P., and Villavicencio, A., editors, *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 7410–7423, Dublin, Ireland. Association for Computational Linguistics.

Linzen, T. and Baroni, M. (2021). Syntactic structure from deep learning. *Annual Review of Linguistics*, 7(1):195–212.

Madabushi, H. T., Romain, L., Divjak, D., and Milin, P. (2020). Cxgbert: Bert meets construction grammar.

Madabushi, H. T., Romain, L., Milin, P., and Divjak, D. (2024). Construction grammar and language models.

Manning, C. D., Clark, K., Hewitt, J., Khandelwal, U., and Levy, O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences*, 117(48):30046–30054.

OpenAI (2024a). GPT-4o Technical Report. https://openai.com/index/gpt-4o. Accessed: 2025-04-03.

OpenAI (2024b). o1 Model Overview. https://openai.com/index/o1. Accessed: 2025-04-03.

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., and Lowe, R. (2022). Training language models to follow instructions with human feedback.

Piantadosi, S. T. (2024). Modern language models refute Chomsky's approach to language. In Gibson, E. and Poliak, M., editors, *From fieldwork to linguistic theory: A tribute to Dan Everett*, volume 15 of *Empirically Oriented Theoretical Morphology and Syntax*, pages 353–414. Language Science Press. Updated with a postscript on replies to the original draft. Reference: lingbuzz/007180.

Popper, K. (1959). *The Logic of Scientific Discovery*. Routledge, Abingdon-on-Thames, United Kingdom.

Portelance, E. and Jasbi, M. (2024). "on the goals of linguistic theory": Revisiting chomskyan theories in the era of ai.

Pullum, G. K. (1989). Formal linguistics meets the boojum. *Natural Language Linguistic Theory*, 7(1):137–143.

Rogers, A., Kovaleva, O., and Rumshisky, A. (2020). A primer in bertology: What we know about how bert works.

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., and Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45):e2105646118.

Shain, C., Meister, C., Pimentel, T., Cotterell, R., and Levy, R. (2024). Large-scale evidence for logarithmic effects of word predictability on reading time. *Proceedings of the National Academy of Sciences*, 121(10):e2307876121.

Stabler, E. P. (1997). Derivational minimalism. In Retoré, C., editor, *Logical Aspects of Computational Linguistics*, volume 1328 of *Lecture Notes in Computer Science*, pages 68–95. Springer Berlin Heidelberg.

Tenney, I., Xia, P., Chen, B., Wang, A., Poliak, A., McCoy, R. T., Kim, N., Durme, B. V., Bowman, S. R., Das, D., and Pavlick, E. (2019). What do you learn from context? probing for sentence structure in contextualized word representations.

Tjuatja, L., Liu, E., Levin, L., and Neubig, G. (2023). Syntax and semantics meet in the "middle": Probing the syntax-semantics interface of LMs through agentivity. In Palmer, A. and Camacho-collados, J., editors, *Proceedings of the 12th Joint Conference on Lexical and Computational Semantics (*SEM 2023)*, pages 149–164, Toronto, Canada. Association for Computational Linguistics.

Warstadt, A. and Bowman, S. R. (2020). Can neural networks acquire a structural bias from raw linguistic data?

Warstadt, A. and Bowman, S. R. (2024). What artificial neural networks can tell us about human language acquisition.

Warstadt, A., Mueller, A., Choshen, L., Wilcox, E., Zhuang, C., Ciro, J., Mosquera, R., Paranjabe, B., Williams, A., Linzen, T., and Cotterell, R. (2023). Findings of the BabyLM challenge: Sample-efficient pretraining on developmentally plausible corpora. In Warstadt, A., Mueller, A., Choshen, L., Wilcox, E., Zhuang, C., Ciro, J., Mosquera, R., Paranjabe, B., Williams, A., Linzen, T., and Cotterell, R., editors, *Proceedings of the BabyLM Challenge at the 27th Conference on Computational Natural Language Learning*, pages 1–34, Singapore. Association for Computational Linguistics.

Warstadt, A., Singh, A., and Bowman, S. R. (2019). Neural network acceptability judgments. *Transactions of the Association for Computational Linguistics*, 7:625–641.

Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., and Fedus, W. (2022). Emergent abilities of large language models.

Weissweiler, L., He, T., Otani, N., Mortensen, D. R., Levin, L., and Schütze, H. (2023). Construction grammar provides unique insight into neural language models.

Weissweiler, L., Köksal, A., and Schütze, H. (2024). Hybrid human-llm corpus construction and llm evaluation for rare linguistic phenomena.

Wilcox, E. G., Futrell, R., and Levy, R. (2024). Using computational models to test syntactic learnability. *Linguistic Inquiry*, 55(4):805–848.

Xu, L., Wu, J., Peng, J., Gong, Z., Cai, M., and Wang, T. (2023). Enhancing language representation with constructional information for natural language understanding. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, page 4685–4705. Association for Computational Linguistics.

Zhang, Z., Zhao, J., Zhang, Q., Gui, T., and Huang, X. (2024). Unveiling linguistic regions in large language models. In Ku, L.-W., Martins, A., and Srikumar, V., editors, *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 6228–6247, Bangkok, Thailand. Association for Computational Linguistics.

Şahin, G. G., Kementchedjhieva, Y., Rust, P., and Gurevych, I. (2020). Puzzling machines: A challenge on learning from small data.