DOI: 10.26346/1120-2726-250

No

Edgar Onea,[a] Filipe Hisao Kobayashi,[b] Susi Wurmbrand[c]

[a] University of Graz, Austria <edgar.onea-gaspar@uni-graz.at> [b,c] University of Salzburg, Austria <filipe.kobayashi@plus.ac.at> <susanne.wurmbrand@plus.ac.at>

This reply makes two main points. First, it lays out why very large language models [vLLMs], although useful as tools in linguistics, cannot be compared to linguistic theories. Linguistics as a science is necessary to understand the workings of human language, and to gain insights into the cognitive properties of human knowledge pertaining to language. Second, the reply clarifies the spectrum of generative grammar and points to major achievements in the field of theoretical linguistics.

KEYWORDS: linguistic theories, generative grammar, formal generative typology.

1. Introduction

No. This is our answer to the attention-grabbing title-question of the paper, *Is it the end of (generative) linguistics as we know it?*, which we reply to here. From a scientific perspective, the end of generative linguistic research, which has as its goal a theory of human cognition, specifically the part we may call the human language faculty, is nowhere closer to the end than theories of physics, biology, or psychology.

Criticism of theories of generative linguistics, as perhaps most fundamental theories in other fields, has a long tradition from Sampson (1997) to Ibbotson & Tomasello (2016), including substantial foundational discussions of theoretical assumptions and empirical methods as well as shallow polemic claims. Yet, generative research is still very much alive and, in our view, blossoming. The current paper lines up with a new wave of criticism initiated by the provocative views in Piantadosi (2024), suggesting that the very large language models [vLLMs] currently dominating the natural language processing field are in some sense 'better' theories of language than what generative linguistics has to offer. While Chesi's paper is more cautious in claiming a victory of vLLMs over theoretical linguistics, it still seems to paint a dire future for generative linguistics in view of impressive advances of vLLMs, suggesting that vLLMs may be competitors for generative linguistics. One main take-home message of our reply here is that vLLMs and generative grammar are in no way competitors, but rather two very dif-

 (received December 2024)

ferent models which exist side-by-side due to their very different goals, areas, and scope of applications.

We believe that much of the apparent debate stems from comparisons of not-likes, in particular regarding what the goals of the different models dealing with language are, different views of what linguistic 'theories' are, and a lack of engagement with actual linguistic theories beyond the highly restricted set of works in narrow Minimalism that are cited in the original paper. Paired with a certain populist trend,[1] a very important and exciting field – generative grammar – has come under attack, which, once the goals and accomplishments of the field are correctly represented, is unwarranted. The article by Chesi brings out many interesting questions and conclusions, some of which we agree with, others we don't.[2] The criticism on the lack of a unified formalization in generative linguistics is well taken, albeit unspecific: there hardly exists a scientific field studying complex empirical phenomena with a unified formalism

In this reply, we wish to make two general points. First, vLLMs are very different from linguistic theories and do not address what we take to be the main goal of theoretical linguistics, that is, to develop a theory of what it means to 'know' a language – i.e. native speakers' (usually implicit) knowledge about the rules and restrictions of their languages. Although vLLMs come very close to generating outputs that include all, and possibly also only, the grammatical expressions of a language, these systems can by no means be seen as linguistic theories. Second, the achievements of linguistic research are severely misrepresented in Chesi's article. It is correct that after 30 years of research within the Minimalist Program, we are still far from having a complete theory of human language. However, this is very different from saying that nothing has been achieved, and we believe that it is exactly this linguistic research, especially its numerous empirical discoveries and advancements, that any understanding of language (whether from a generative or vLLM perspective) needs to build on.

2. vLLMs are not theories

Chesi's position on the question of whether vLLMs are theories of language, i.e., linguistic theories, and let alone robust theories, is somewhat unclear in certain parts of the paper, but ultimately it seems fair to say that he takes vLLMs to be at least comparable to such theories, if not the most successful one. The following two quotes illustrate this. While the first quote explicitly claims that vLLMs are the best linguistic

theories, the second presupposes that they are state-of-the-art theories of language, demoting linguistics to – if anything – contributing interesting new pieces of data to be ultimately captured by the vLLMs.

> [Quote 1] While vLLMs are arguably overrated as linguistic theories, the methodology proposed by Wilcox and colleagues (Wilcox *et al.* 2023) represents an appropriate approach to testing them. […] In this respect, these vLLMs are, in fact, really the best theories on the market, i.e. observationally more adequate than any MG. (Chesi *this issue*: 39)

> [Quote 2] Ultimately, the most significant contribution that a generative linguist can provide is a linguistic minimal contrast challenging a specific theoretical assumption or the performance of a vLLM. Successfully incorporating this new contrast into a shared dataset, which any (r)evolutionary explicit formalism must confront, would represent quite a considerable accomplishment in my opinion. (Chesi *this issue*: 40)

But in what precise sense are vLLMs theories of language, one may ask? The first and most natural answer would be to say that they solve, or are close to solving, the language problem as formulated in (1).

(1) Language Problem (Chesi: 7)
Is theory X capable of generating and recognizing all and only the sentences Ss belonging to language L?

vLLMs of various kinds are remarkably good at generating grammatical sentences of English and to a certain degree they are also able to distinguish between grammatical and ungrammatical sentences of English if properly prompted. These behaviors can be seen as predictions of a fully formalized and in fact computationally implemented theory. However, a theory is usually understood as some explication of human understanding based on principles and rules and not just as a prediction machine (cf. also Kodner *et al.* 2023). If the underlying principles are not clear to a human, we cannot say that they serve understanding despite their success at the purely predictive level. Consider the following analogy: a box that always correctly predicts the weather for the next day. The box cannot be opened without destroying it, and its workings are entirely unknown by humanity (maybe it was gifted to humans by some alien civilization). It is hard to imagine that anybody would seriously claim that this box is in any interesting sense a theory of meteorology. The crucial missing bit is human understanding. In this sense, vLLMs seem to be nothing more than statistical models with their

output as their predictions. But even if one were to accept that statistical models are theories, Collins (2024) discusses a further mathematical reason why one would not want to consider vLLMs theories. Specifically, they are unbounded in the sense that vLLMs are able to approximate any mathematical function whatsoever. Thus, their success in a domain simply shows that success is possible, a point already shown by the very existence of humans who speak language.

But to the best of our understanding, it seems that Chesi, in agreement with Piantadosi, has something different in mind "As argued by Baroni 2022, […] language models should be treated as bona fide linguistic theories." (Piantadosi 2024: 360). Here is the core of Baroni's argument:

> It is more appropriate […] to look at deep nets as linguistic theories, encoding non-trivial structural priors facilitating language acquisition and processing. More precisely, we can think of a deep net architecture, before any language specific training, as a general theory defining a space of possible grammars, and of the same network trained on data from a specific language as a grammar, that is, a computational system that, given an input utterance in a language, can predict whether the sequence is acceptable to an idealized speaker of the language (e.g., Chomsky, 1986; Müller, 2020; Sag *et al.*, 2003).
> It is undoubtedly easier to inspect the inner workings of a symbolic linguistic theory than those of a trained deep net, and indeed a classic objection against artificial neural networks as cognitive theories is that they are unopenable black boxes (e.g., McCloskey 1991). However, going hand in hand with the development of more complex models, the field has also made extensive progress in the development of methods to analyse their states and behaviors (Belinkov and Glass, 2019), providing strong methodological support for a systematic analysis of deep nets. (Baroni 2022: 7)

In other words, under such a view, a vLLM is a theory that has a specific architecture and thus this architectural component, that can be well understood, is to be considered a general theory of language, and the specific states of those architectures reached by training can be seen as specific instantiations of grammars. Before evaluating this claim, it is worth being more precise about what this sort of architecture actually means. vLLMs, broadly speaking and glossing over differences, are essentially probabilistic functions from inputs to outputs with a complex computational algorithm typically based on neural network architectures. This complex algorithm is best understood as an architecture of various modules of complex networks that contain fixed operations and variable

parameters. The parameters are adjusted during training to optimize output. Training thereby is usually understood as an optimization of predicting the next token given an input context-window. There are three crucial components of the architecture of vLLMs that need to be mentioned explicitly. Firstly, the input to a vLLM is usually not understood as tokens of language but as 'language embeddings', i.e., linguistic expressions are represented by vector spaces that are derived from the analysis of the whole training corpus (Pennington *et al.* 2014). In other words, a word, syllable, or even letter (a.k.a. *subword tokenization*) is – in the very first step – transformed into a vector that stores something akin to information about its distribution but that is in fact optimized for the tasks of the vLLM. Secondly, vLLMs encompass so-called 'self-attention' operations that take the input sequence and transform it using learned linear projections. They compute pairwise interactions between all elements of the sequence, resulting in a set of dynamic weights (Vaswani *et al.* 2017). These weights are used to create a weighted sum of the input values. This process can be repeated multiple times in subsequent layers, enabling the model to iteratively refine its understanding of the relationships between different parts of the input, which allows the model to 'decide' which parts of the input to focus on and to what degree in the computational process (Brown *et al.* 2020). These two elements are crucial for two reasons: (i) language embeddings are specific probabilistic computational implementations of the notion of equivalence classes, where, however, syntax is not the sole criterion, but semantics, pragmatics and any other distributional aspect are also implicitly included; and (ii) self-attention is specifically designed to handle dependencies, and very specifically long-range dependencies, in language, along with the role of context for weighing the importance of linguistic information. The third component involves what many refer to as a black box, i.e., extremely complex neural network layers that essentially perform matrix-operations that cannot be easily comprehended by humans in a conceptual way, even though they are mathematically well-defined.

With this background, we suggest three reasons for why we must reject the idea that vLLMs are in any way, shape, or form linguistic theories. Firstly, to our knowledge, all existing vLLM models include an entirely opaque layer that transitions between their theoretically better understood activities (vectorization, self-attention) and their output in ways that are indiscernible for humans. If translated into generative mathematical terminology, this would amount to the following type of theory: (i) we explicate the numeration in some specific way; (ii) we apply some well understood syntactic operations leading us to some syntactic representation; (iii) we apply an unknown function F to that

syntactic representation; (iv) we get correct results in each and every case. As long as we do not understand the function F, the fact that such a function exists has only limited theoretical relevance. Specifically, the existence of such a function shows that the problem formulated in (1) can be solved. What it does not show, however, is that it is possible for humans to understand F. Should it turn out that F is similarly hard to understand as the human brain, the result would not be very exciting. Hence, speaking about such unknown functions would make it hard to judge the merits of the model irrespective of its empirical glory.

Secondly, by the same reasoning as the one Baroni implies, one can easily claim that a human who speaks English is a theory of English. After all, a human comes with a prior architecture (a human's brain has structure), and learning a language is a parameter fitting of the human brain in the huge space of possible theories, selected by what works best, e.g., in terms of producing the best responses. The point here is not that there is a difference between humans and machines, but that even if there were no differences at all, this would not make a vLLM more a theory of language than a speaker of the language. A theory would be the results of the 'study' of a vLLM. If we had a set of explanatory rules and generalizations, rules about how a vLLM makes its predictions, what it can and cannot predict in general, when it makes mistakes and when it does not, etc., that knowledge would yield a theory of language (as instantiated by a vLLM, and not necessarily about human language). But the vLLM itself is in no way a theory.

So, we may ask whether we want to study vLLMs instead of humans, which may appear cheaper and more ethical (although the impact on climate caused by vLLMs should also not be ignored). But, and this is our third point, one can actually be quite certain that vLLMs operate quite differently from humans and thus studying them will likely not get us very far in the endeavor of studying the human language faculty. There are several reasons to assume that vLLMs are fundamentally different in comparison to humans in terms of language faculty. One is that vLLMs can have quite radically different architecture, size, and training methods and yet achieve excellent (and thus comparable) results (see, e.g., Naveed *et al.* 2024 for a recent comprehensive overview). Another one is simply the amount of data including powerful pre-analysis of the data (e.g., vectorization) that these models are trained on; amounts of data that would seem impossible for a human to be exposed to. Yet another argument is that vLLMs currently do not integrate other cognitive faculties such as vision and other sensory input. Arguably, if vLLMs manage to successfully combine a more holistic picture of reality with linguistic data,[3] one can expect that vLLMs

could become much more similar to humans. For now, however, this just highlights that currently, they are very different. Finally and more importantly, since our knowledge about the human brain's workings is still limited, we simply would not be able to tell if some particular vLLM does or does not mirror the way human cognition works, and thus it could not be established whether or not such research is successful.

In conclusion, vLLMs are not theories and they are not about humans, which makes them, albeit useful in many ways, not relevant for learning much, if anything, about the human language faculty. Indeed, we would change the perspective and recommendations expressed by Chesi. It is not the task of generative linguistics to provide data for vLLMs, but instead generative linguists can use vLLMs to optimize their work process and possibly even perform small-scale experiments. vLLMs are tools for linguists, and not the other way around. This is because science is different from engineering.

3. What generative linguists really do

In the previous section, we pointed to fundamental differences between vLLMs and linguistic theories, which effectively make a comparison between the two not very informative. In this section, we concentrate on what generative linguistics (really) is, summarize areas which have led to great advancements, and clarify points of the debate which we felt are either missing or misrepresented in the article by Chesi.

The goal of a theoretical linguist can be summarized as in the quote below from Marantz (2019):

> But what should be clear to anyone reading these attacks on linguists is that computationalists are not engaged in the same scientific enterprise as linguists. The linguistic enterprise is about the knowledge of language that underlies everything that a speaker does with his/her language, including not only writing those web pages that serve as data for computational linguistics, but also understanding and making judgments about sentences that are carefully constructed by linguists as test cases to decide between competing theories. (Marantz 2019: 10)

In our understanding, all approaches that have as a goal the modeling of the human language capacity would fall under the umbrella 'generative grammar'. Specifically, restricting ourselves to syntax here, this would include, in addition to the narrow Minimalism used by Chesi, other developments of the Government and Binding [GB]/Principles

and Parameters [P&P] theories, Relational Grammar [RG], Lexical Functional Grammar [LFG], Tree Adjoining Grammar, Head-Driven Phrase Structure Grammar, and others. The reason why this is important to keep in mind is that there has been immense progress in these approaches, both empirically and theoretically (see below). While it is correct that the variety of frameworks comes with a not always unified variety of terminology and formalisms, we do not share Chesi's concern that this is necessarily a major hurdle for approaching the main goal of theoretical linguistics, which is to understand the knowledge of language that humans have. On the contrary, as pointed out in Marantz (2019), theory-internal predictions are often exactly what pushes the field forwards and leads to new discoveries. Moreover, there have also been major successes where predictions and generalizations gained through one theoretical lens have informed and altered thinking in other theoretical approaches (see, for instance, the RG/LFG discovery of unaccusativity and other grammatical function phenomena, which have become standard wisdom in GB/P&P/Minimalist approaches). Lastly, as we concluded in the previous section, it is not the task of theoretical linguistics to prepare data and unified concepts for further processing by vLLMs, but the main task is, and has always been, to model grammatical properties and dependencies (including those types on which vLLMs still perform poorly; see, e.g., Katzir 2023, Charchidi 2024) in a way that reflects the knowledge of language that humans have.

Generally, the article grossly misrepresents which theories fall under the label 'generative grammar', what phenomena have been investigated (the decades of generative research have gone far beyond word order), the range of generalizations that have been found (see D'Alessandro 2019 for an extensive summary), and the theoretical methods available in these approaches. We are not necessarily objecting to the critique against the corner of Minimalism offered by Chesi, but the extension from that critique to a condemnation of the entire generative enterprise, to us, is entirely invalid. We particularly object to the impression given in the article that no or only very little progress has been made in the field over the last decades. This impression does not reflect reality, and exactly such dismissiveness and ignorance have already been criticized in Marantz (2019):

> Linguists predict data they don't have, the body of empirical generalizations uncovered by the methodology grows year by year, and alternative accounts of phenomena are in fact pitted against each other, with the losers no longer viable. Progress in linguistics is transparently displayed in our major journals; nevertheless, some scientists and engi-

> neers that deal with language still question the legitimacy of the generative linguistic enterprise. (Marantz 2019: 9)

Chesi somewhat acknowledges this in the section titled *Generative Parameters and Word Order Variation*, referring to approaches in the 1980s and 1990s where the goal was a "comprehensive list of parameters and organizing them into a coherent hierarchy" (p. 36). The entire section, however, is just two pages long and thematically rather one-sided, which is quite puzzling, since Chesi himself states that "In this domain, I perceive the most significant advancements within generative linguistics" (p. 36).

Yet except for this short section, the rest of the article presents current linguistic research solely within the generative tradition concerned with the key Minimalist question "How close does language come to optimal design?", which, in practice, characterizes only one, arguably not even the most widespread, area of research in the field (see also Bobaljik & Wurmbrand 2008). As in any scientific enterprise, parsimony and theoretical simplicity still play a key role in linguistic theorizing, but the goals of many practicing linguists seem to be, instead, to understand the extent and regularities of cross-linguistic variation. We may agree with Chesi that some of the findings are still fragmentary and not always consistent with each other, but that is a natural consequence of the complexity of the variation.

We do want to highlight, however, that in many areas, significant progress has been made by exactly such research on universals and variation. A fruitful framework in the post-GB/P&P tradition, for instance, is 'formal generative typology' [FGT] (Baker 2009, Baker & McCloskey 2007) where findings from typological research are combined with the formal methods and concepts from the generative tradition. Among the most important of these findings, are so-called 'implicational hierarchies/universals', which Corbett (2010) has described as "the most powerful theoretical tools available to the typologist" as they "allow us to make specific and restrictive claims about possible human languages" (p. 1). Implicational universals are generalizations of the form: *if* a language has property A, it also has property B (but not vice versa). In other words, they are claims about non-existing languages – those that have A but not B. The goal of many works within FGT (e.g., Bobaljik 2012, Lohninger & Wurmbrand 2025) is to precisely explain these generalizations as a consequence of the very nature of the human language faculty. FGT approaches, therefore go beyond the Language Problem, as they typically aim for generalizations that are at the heart of the 'Poverty of Stimulus' problem. Similarly, Pesetsky (2024) makes an argument based on universals by considering the property of the *that*-trace effect:

> The fact that languages in general behave this way teaches us that someone's knowledge of the *that*-trace effect is unlikely to arise from that individual's exposure to data… precisely because it is a general fact about languages across the globe, not a fact particular to one speech community, much less one individual exposed to one particular corpus. (Pesetsky 2024: 38)

Findings from universals and the points of variation show that generative research goes far beyond testing whether the grammar of a given language L identifies all and only the sentences of L. It is unclear how vLLMs could ever reach such findings on their own (given also the difficulties of vLLMs in forming generalizations as noted in Katzir 2023), let alone explain them. It is generative linguistics, by its very nature, understanding individual languages as part of a broader shared capacity, that sets out (and answers) the kind of research program that can uncover generalizations that span all languages. Despite their successes in mimicking English (and possibly some other languages), vLLMs seem ill-suited to being able to probe questions of linguistic universals and cross-linguistic generalizations, especially considering that the vast majority of the world's languages are unlikely to ever have corpora of the size needed for vLLMs to even get off the ground.

4. Conclusion

Chesi's article makes many important points and helpful suggestions for both computational and generative linguists. In our reply we highlighted two points where we see the world rather differently: the view that vLLMs are linguistic theories and the assessment of the achievements and importance of generative linguistics. There are two more general take-home messages we would like to offer here: one specifically for generative linguists, one for anyone who thinks they would want to say something about generative linguistics.

Chesi's article has voiced a problem with the perception of what generative linguistics is (it is a diverse field), what its goals are (our goal is not to build a speaking machine mirroring human language use), and what tools and methods are available (e.g., universals, Formal Generative Typology). Given that the overall tone of Chesi's article seems to be to question the importance and chances of survival of the entire field of generative grammar, concentrating on a tiny corner of the field and leaving the vast majority of other parts completely aside is not only unsound from the perspective of the argumentation, but also

creates a skewed picture of the situation, leading, in particular to the untrained eye, to dangerous disinformation.

As for the generative linguists, our recommendation would be to continue all the exciting research programs, continue asking the important questions about the human capacity for language, but also engage more with research from other fields, via making our research more accessible and incorporating questions and results from different perspectives (see, for instance, the collaborative Austrian research project *Language between Redundancy and Deficiency* which the authors of this reply are part of and where we take seriously the role of the stochastic cognitive environment in which language as a generative rule-based system is embedded; or the ERC synergy project *Realizing Leibniz's Dream: Child Languages as a Mirror of the Mind* led by Alexiadou, Guasti, and Sauerland). Much of the debate appears to us to be caused by misunderstandings (at various levels), and our hope is that by mutual engagement and understanding of the different goals and methods, these could be avoided and the benefits of different approaches to language could not only inform each other but also strengthen the different frameworks.

Abbreviations

GB = Government and Binding; FGT = Formal Generative Typology; P&P = Principles and Parameters; vLLM = very Large Language Models.

Note

[1] Statements such as "Piantadosi's dismissal of Chomsky's approach is ruthless, but generative linguists deserve it", "A spectre was haunting generative linguistics", or "Exotic theoretical puzzles with funny names and acronyms" may be catchy, but have no place in a scientific article.

[2] Since it is not always clear to us what the exact conclusions are that Chesi reaches, some of our points may not address his article directly, but rather the works his paper is building on.

[3] This may be surprisingly difficult given that, as opposed to raw linguistic data, we do have not so much linguistic data that are actually mapped to the other sensory data in a useful way (such as, e.g., picture descriptions).

Bibliographical References

See the unified list at the end of this issue.