DOI: 10.26346/1120-2726-244

Is it the end of Generative linguistics as we know it?

Gillian Catriona Ramchand

Faculty of Linguistics, Phonetics and Philology, University of Oxford, UK
<gillian.ramchand@ling-phil.ox.ac.uk>
Department of Language and Culture, UiT the Arctic University of Norway, Norway
<gillian.ramchand@uit.no>

Are there certain core beliefs of generative grammar that are fatally undermined by the recent successes of Large Language Models and the unsupervised learning that trains them? Do Large Language Models then constitute a rival (and superior) 'theory' that can and should take over now from (all) previous theories in pushing the science forward? In this short article, I argue that the answer to both these questions is 'no'. On the positive side, I make an urgent case for maintaining theory at the centre of the new era of linguistic science, and for generative grammar to expand its energies into theorizing the link between competence and various aspects of performance in order to shore up its claims to explanatory adequacy.

KEYWORDS: Large Language Models, linguistics, poverty of the stimulus.

1. Introduction

The recent rise of Large Language Models (LLMs) as foundation models for a wide variety of artificial intelligence applications has led many to predict the end of the world 'as we know it'. Many of our old certainties are set to change in the wake of these new advances: the value and marketability of different kinds of labour and expertise (with its knock-on effects for education); improved pattern discovery tools for scientific and medical applications; threats to the reliability of our information ecosystems, with the global instability that can give rise to; the exacerbation of global inequalities in an era where a small number of stakeholders control access to the new technology and profit from it, while a large number will suffer from its energy requirements. To this dystopian list, we apparently need to add the demise of theoretical linguistics (or 'generative' linguistics as Chesi puts it) 'as we know it'. As a generative linguist myself, I must confess I am much less worried for my field than I am about the planet more generally, but I welcome the opportunity to revisit the deep and fascinating questions that inform the scientific study of language systems and their natural embedding in human cognition in the light of these recent developments in natural language processing models. The title of Chesi's (*this issue*) contribution that I am responding to alludes to the 'the end of generative grammar as we know it', and so in my own response I feel obliged to address the specific elements of that research program that have been claimed to have been refuted. For this reason, I will also respond to the particular claims in Piantadosi (2024) as part of my response to Chesi's own assessment. In the second half of this article, I will return to address Chesi's own main focus as I see it, which is the diagnosis of what generative grammar needs to do to forestall its alleged 'irrelevance' in the modern era.

So, what characterizes generative grammar as a field of inquiry, and what would it mean for it to be over? Chesi (*this issue*) does not define what he means by 'generative grammar' directly, calling it only 'a prototypical theoretical perspective on language'. The general thrust of his critique is that generative grammar needs to get its act together in terms of developing more precise and rigorous formulations of its model, which in turn would allow it to participate in the joint enterprise of evaluating its descriptive adequacy against a standardized dataset, comparing it with the LLMs that Piantadosi (2024) claims also count as 'genuine theories of language'. Piantadosi's own position seems to be more foundationally critical, asserting directly that LLMs 'refute' certain core principles of generative grammar (or rather, Chomsky's general approach to grammar). While Chesi is more circumspect here, the general thrust of his critique is that generative grammar needs to clean up its act as a theory, or risk becoming obsolete.

There are two issues that I want to address separately in this response article: (i) Are there certain core beliefs of generative grammar that are fatally undermined by the recent successes of LLMs and the unsupervised learning that trains them? (ii) Do LLMs then constitute a rival (and superior) 'theory' that can and should take over now from (all) previous theories in pushing the science forward?

2. What is 'Chomsky's approach', and how will I know when it has been refuted?

I first remind the reader that 'Chomsky's approach' to grammar and 'Generative Linguistics' should not be considered the same thing. Chesi criticizes minimalism in particular for not being sufficiently precisely and consistently formalized to generate the kinds of predictions required for a theory (with the exception of Stabler 1997, Collins & Stabler 2016). But other branches of generative grammar such as LFG and HPSG have a long history of being tightly controlled and formalized and have always gone hand in hand with work in computational linguistics. So criticizing minimalism should not be considered a criticism of generative linguistics as a whole. Maybe Chesi, like Piantadosi (2024), actually just means 'Chomskian approaches' here. In the next section, I take a look at a number of the ideas that might be considered to be criterial of 'Chomskian' approaches in this sense, to assess whether or not they have been refuted by the success of LLMs.

2.1. Innateness

What then defines the generative linguistics school of scientific endeavour? <thoughtco.com> says it is the idea that all humans are born with an innate capacity for language. Nativism therefore seems to be a good candidate for a principle that characterizes generative grammar in the minds of those who consider it to have been debunked. In Piantadosi (2024), the list of 'refutations' contributed by the success of LLMs includes the idea 'Hierarchical structure need not be innate'.

In fact, it is widely acknowledged that the concept of innateness is far more complex and subtle than this simple characterization suggests, and in ways that undercut the usefulness of trying to use it as a deciding ideological commitment to separate scientific schools of thought. Firstly, acknowledging the reality of an innate contribution is compatible with many different frameworks and positions. In *Rethinking Innateness*, Bates *et al.* (1996) point out that no scientist working on questions involving nature/nurture balance thinks that there is a simple relationship between genes and phenotype, and so it is increasingly difficult to draw the line between the kind of information contributed by the genome and that contributed by the developmental environment. Secondly, the kind of information that could be considered part of genetic endowment can also be quite abstract – in addition to being genetically hardwired to know certain things or perform certain behaviours, we could also possess genetic information that gives rise to the timing of certain developmental stages affecting rate and strategies for learning itself, the gross anatomy of certain brain regions and how they interact developmentally etc. Given what we currently know about brain plasticity and learning in humans, it is highly implausible that the kind of 'representational innateness' for language-specific principles that people may first think of when hearing the terms 'Universal Grammar' (UG) or 'the Language Acquisition Device' could be the case. Moreover, Chomsky himself, possibly in response to science's improving understanding of brain, genes and development, moved away from a crude kind of representational innateness in the old fashioned sense, to one that is much more minimal (Hauser *et al.* 2002). The more modern claim is that human minds have an innate capacity for recursion and that this allows hierarchical syntactic structures to emerge as a solution to the problem of language acquisition.

This is a far cry from the idea of UG as a templatic blueprint of abstract structures and rules given innately, in advance of any linguistic experience. I do not personally know of (many) working generative syntacticians who believe in that early simplistic version of UG. In short, the idea that hierarchical representations do not have to be innately given as 'representations' contradicts nothing in minimalist theorizing, and nor does it contradict the idea that there is some aspect of our genetic endowment that makes language possible in humans, an idea conceded already by the connectionist tradition in Bates *et al.* (1996).

2.2. A note on poverty of the stimulus

To what extent does language data in principle furnish enough information for the learner to infer the system that generates it? The lesson from LLMs and the transformer architecture seems to be that if a large enough data set is provided to the learner, then it appears that the distributional properties of language tokens provide enough implicit information to allow the model to infer the correct hidden layers of structure to perform correctly on a text generating task. Of course, a huge amount of data is required to achieve this feat, so the second relevant question here is whether this fact about LLMs at all defuses the poverty of the stimulus argument as it applies to actual human children. Children learn grammar from far less input than this, and also in a rather different kind of context, and motivated by a different kind of holistic 'task'. We know that humans do engage in predictive processing, but the driving force behind language learning for the child is not plausibly a game to guess what word is coming next, or indeed to figure out which sentences are grammatical or not. The child's 'task' is more likely to be the drive to understand the content and emotional value of what is being said to them, to predict the macro behaviour of other humans, and to acquire the toolbox to communicate in return. And for that children have a additional information coming from the context and shared experience of the world with their interlocutor. They also bring some things to the task including independent domain general facts about their own shared cognition with the speakers of the language they are learning, as well as possibly certain analytic and learning biases built in to their developing cognitive systems. So, is there enough information coming from the incoming signal to acquire the system the child is faced with, or do we need to take account of cognitive and learning biases? To my mind it is a fascinating and still open question understanding how a child achieves this feat. But it is equally clear to me that the current evidence from the behaviour of LLMs does not advance our understanding of the question either way. Concerning the training algorithms themselves, it is widely acknowledged that the back propagation training algorithms that LLMs use are very different from human learning in rather deep ways (Hinton 2022, Evanson *et al.* 2023), and so cannot be considered good models for human learning, even independently of the huge disparities in data size required to achieve the same results.

2.3. The relation between syntax and semantics

Another place where Piantadosi (2024) claims that LLMs refute a basic principle of Chomsky-inspired grammar lies in the relationship between syntax and semantics. The claim is that in the training data (and in the model that is built in response to it), syntactic and semantic information are 'integrated' and cannot meaningfully be separated. This supposedly contradicts Chomsky's view on the autonomy of syntax. I must confess that I am not sure I understand the point that is being made here. Nobody would deny that the language that the child is faced with is a combination of syntactic and semantic properties; the data that an LLM is trained on is the same, because it is just language, the same as for the child. Nevertheless, the model learns syntactic generalizations from this input, where nobody has given it the meta-information concerning how to separate the syntax from the semantics in principle. But this once again is the same for the child. The 'integration' of syntax and semantics which forms the 'refutation' here must lie in the implicit analysis or model that the LLM ends up embodying. While it is well known that inspection of the detailed representations of these models is quite difficult in principle (because of the 'black box' nature of the system), Piantadosi admits (and even elsewhere makes a virtue of the fact that) these models do seem to end up inferring syntactic generalizations, representing sentences hierarchically, implicitly characterizing word class membership, and tracking long distance dependencies. In other words, LLMs do end up representing syntactic information as a result of their training (Manning *et al.* 2020, Futrell *et al.* 2019, Linzen & Baroni 2021), even though they were not told to look for syntactic generalizations in advance, and even though the information is hopelessly entangled with semantics in the form of the distributional properties of linguistic tokens. In particular, they also seem to organize information in a way that is similar to tree structures (Manning *et al.* 2020), and the extent to which this is true even predicts the model's performance on generalization (Murty *et al.* 2022). The models do well, in addition, on function words (Kim *et al.* 2019, and filler-gap dependencies (Wilcox *et al.* 2018).

I do in fact think there is something interesting and remarkable about what these models achieve. Recall that the task that the LLMs are set in training is to 'predict' the next token, and they are constantly given positive and negative feedback on that task and undergo mind bogglingly gargantuan amounts of training on it. They are allowed to use any information they can to help them succeed at the task, and they come to the conclusion that they can predict better if they start to classify words into classes, and build a syntax around them, in addition to the more fine grained lexical distributional statistics. So LLMs figure out for themselves in some sense that Chomsky is right – general statistics between words is not enough, they must also infer and build in hidden syntactic structure to do a good job on this prediction task!

Piantadosi thinks that it is obvious that syntax and semantics are not separated in the model's analysis, but no-one has convincingly shown that they can simply and reliably inspect what is in these models' analysis (which is why e.g. BERTology is its own distinct industry and object of inquiry – Rogers *et al.* 2021). In seeming contradiction with the above point, when discussing the problem of the (inhumanly) huge amounts of training data required to achieve appropriate linguistic behaviours, Piantadosi informs us that the syntactic part of LLM competence is reached with much less data, with semantics and real world knowledge being the data-guzzling culprits, suggesting that "syntactic knowledge requires a small number of bits of information, especially when compared to semantics (Mollica & Piantadosi 2019)". It sounds to me as though the different components of the model's knowledge are being (at least implicitly) separated here by Piantadosi at the acquisitional and implementational level.

In this discussion, Piantadosi also makes some non-standard assumptions about what semantics is. For him, 'semantics' is being proxied by the distributional properties of individual words and tokens of the language. But if there is one thing we know about LLMs it is that they have no mapping between the tokens of language and anything at all that exists 'outside language'. They, in other words, have no concept of denotation, reference or truth in the mapping to a world outside of the language-internal system of dependencies and relations. This mapping to an external reality is what most semanticists would define as semantics, and this is in fact the conception of semantics that Chomsky was most keen to excise from syntax in his autonomy of syntax thesis (Chomsky

1995b). Now, for LLMs, it turns out that not having semantics in the referential sense (or embodied cognition, or shared attention, or communicative urges) is absolutely no impediment at all to acquiring the ability to produce grammatically acceptable and appropriate sentences! So once again, it seems to me that the lesson of learning in LLMs shows us that Semantics (with a capital S, in the semanticists' sense) is not necessary in order to acquire syntax, thus confirming Chomsky's point rather than refuting it.

3. Do LLMs constitute a theory of grammar?

Piantadosi claims that LLMs are a theory of grammar, and Chesi implicitly agrees with this position at least to the degree that he thinks that minimalist proposals should be assessed side by side with them on commonly agreed benchmarks. Ambridge & Blything (2024) claim LLMs do better than theoreticians on all the jobs that grammatical theory was supposed to do. Müller (2024) on the other hand disagrees, arguing that LLMs are not theories in the same sense at all. That paper argues that an LLM is a successful piece of engineering that matches the patterns in the corpus of textual data it is fed, but is not a theory. Specifically, it argues that LLMs are not just a different theory, or a wrong theory, they are not theories at all. Why the disagreement?

Here it is instructive to look at Piantadosi's own justification of the status of LLMs as theory in the form of his own illustrative analogy. How does the model create a theory by setting parameters? Piantadosi asks us to imagine a situation where physicists might for example be uncertain about whether gravitational force falls off as an inverse function of distance r, or of the square of r. We could imagine them constructing a super equation for gravitational force F which has a constant α whose value between zero and 1 represents the effect of the two different characterizations of the situation as in (1). If α is zero then the equation reduces to a function where 1/r2 is the correct determinant, whereas when α is 1, the equation reduces to one where only 1/r is the determinant.

(1) $F(r, \alpha) = \alpha . 1/r + (1 - \alpha) . 1/r2$

So now the job of the model is to inspect the data and select the value of the parameter α that maximizes the likelihood of getting the correct (i.e. matching with reality) answer. So in this case, inferring a parameter in this equation is tantamount to evaluating distinct theories against the data and coming up with the preferred one. So is this what LLMs are doing as well when they are setting their billions of parameters in response to the data? Almost. Piantadosi admits that in this case, we do not give the model a specific super equation in advance like (1), which embodies distinct theoretical proposals. Instead, there are some 'natural bases' or starting points, for which you can set parameters that will allow you to approximate essentially ANY COMPUTATIONAL THEORY. As Piantadosi puts it "Parameter fitting in these models is effectively searching over a huge space of possible theories to see which one works best, in a well-defined, quantitative sense."

And here is the crux of the matter. Because of these universal natural bases, you do not need to have a theory, or even a hunch, or a specific question to ask the oracle when you set one of these neural nets loose on the data. It is a well known proof about these kinds of neural nets that they are capable of approximating to any degree of precision, any function that is in principle computable no matter what it is (Cybenko 1989). So if the training algorithm is sound and the data contains the right information, then the neural net will end up mimicking the input data to an arbitrary degree of precision. And here is the other thing. You cannot at that point go back into the model and reconstitute which particular equation or hypothesis was being piecemeal approximated by its elaborate parameter settings. It would be as if your machine was capable of correctly predicting the value for gravitational force when fed with distance information, but you would have no way of figuring out post hoc what the equation was! I quote from Piantadosi again lest I be accused of unwarranted negativity "In fact, we don't deeply understand how the representations these models create work (see Rogers *et al.* 2021). It is a nontrivial scientific program to discover how their internal states relate to each other and to successful prediction."

Collins (2024) is the only response on the topic that I have seen that makes this point most clearly and trenchantly. He argues that LLMs are not theories because they can represent any complex relationship, and they represent them all in essentially the same kind of way. So it is no good inspecting the representations arrived at by these LLMs for theoretical insight into the workings of language (even if that were easy to do), because the LLM representations work by piecemeal approximation and flatten out any computational specificities inherent in the thing being approximated. If we want the physicists' equivalent of the equation for gravitational force, we are going to have to come up with it ourselves via human scientific theorizing and explicit hypothesis testing. It is possible that what Piantadosi and others are claiming here is that LLMs prove that humans themselves could also be just 'universal approximation machines', in which case the human child is just a supreme pattern matcher who will learn whatever language they are exposed to. This seems unlikely for a number of reasons. Firstly, it underplays the fact that children (and humans in general) are quite a bit worse at the sorts of computations that LLMs seem to excel at especially when it comes to seeing patterns in extremely large amounts of complex data. Secondly, in concentrating on the text prediction task, it misses the fact that human minds 'created' language systems (in all locations where humans can be found) in the first place, with a world-language relationship in mind. We would get no explanation of this phenomenon simply by asserting that the human brain is a massive approximating machine capable of imitating patterns it is exposed to in the form of disembodied language. In Müller (2024), another of the points raised is that studying LLMs trained on particular languages is unlikely to give us any purchase on crosslinguistic similarities and variation that exist. This is because comparative information of this kind is simply not extractable in a way that we as scientists can make sense of or interpret at a higher level. It is unclear therefore, what we gain from our marvellous engineering successes, other than a monetizable object for capitalism to chomp on. In terms of the kind of tangible, symbolically expressed theory that humans need in order to extend and generalize understanding into other domains, we do not seem to be able to extract something of that level from the LLMs we have created. The existence of SyntaxGym[1] provides a useful set of cross-model benchmarks. But it does not mean that the things being compared are all 'theories' in the same sense.

There are of course differences among different LLMs and how well they perform, but in general it appears that the differences between different neural networks come from the interplay between the nature of the training algorithms and the data they are being fed (Collins 2024). We could make a study of those systems and algorithms, but it since we already know there are deep differences between us and the neural nets in both the learning strategies and data exposure, it does not buy us anything to study the LLM instead of the human directly.

4. Where next for linguistic theory?

In this response so far, I have mostly concentrated on the claim that LLMs and their successes have seriously undermined specifically Chomskian approaches to grammar, that they are a refutation of and alternative to those theories. I have argued firstly that nothing criterial to the enterprise has been refuted, and secondly that what has been produced is not qualitatively the same thing as a 'theory of grammar'.

To my mind, what characterizes generative grammar, and the Chomskian approach in particular, was the radical reconception of the object of inquiry when it came to language science. Chomsky reframed the scientific question away from the cataloguing and analysis of linguistic behaviour, towards the psychological questions concerning the nature of linguistic knowledge in human minds that allows them to produce these behaviours. We want to understand the system that generates linguistic behaviour, not simply analyze the produced patterns that are measurable and recordable as outputs of that system. In that sense, the school of thought that considers the LLM model itself to be the 'theory', the desired endpoint of scientific endeavour, is basically a return to the crudest kind of behaviourism where the model is evaluated by how well it succeeds in mimicking the externally observed data, and not in how it helps us to understand human minds.

So indeed, I do think we are seeing a shift that threatens to strike at the heart of the Chomskian enterprise, but not in the way Piantadosi imagines. It is a shift that reifies the patterns of external data as an object of inquiry in its own right, where the goal is to produce generative systems that will demonstrably reproduce the fine detailed patterning of that data. Engineering success is defined by matching output behaviour, not by achieving an understanding of how this happens within the engineered device (let alone how it happens within human minds which are quite different). Dataism is the real existential threat, and one which should be resisted, if we have the goal of understanding the role of language within human cognition. Dataism left unchecked can give rise to a kind of theoretical nihilism, which will lead to dead ends as soon as solutions need to be extended, or generalized over, or metatheorized.

If we are threatened with the demise of generative grammar, it is at this level, as part of a general distrust of symbolic theories and an enthusiasm for bottom up, theory-free, engineering solutions. I think that this is not progress. I think that humans and human scientists have made great strides in understanding the world by using the cognitive 'gadgets' of language and symbolic theorizing. My hunch is that representing information symbolically is the gadget that allows us to generalize explicitly and metacognize in increasingly sophisticated ways. In this new era of big data and artificial intelligence aids to pattern-discovery, we need to maintain a pivotal role for human scientific expertise and theorizing. Generative grammar (or theoretical linguistics more generally) is a natural constituency for where that expertise can continue to be nurtured and from where it can contribute to multidisciplinary questions.

Returning to Chesi (*this issue*), on the question of whether theoreticians, or generative grammar (or specifically the Chomsky inspired linguistic tradition) need to change in response to recent advances in this technology, my answer is yes, but with a perspective slightly different from that articulated in Chesi (*this issue*). I agree with Chesi that there are problems both in formalization, and in the nature of the data that can be used as falsifying evidence. While generative grammar can claim to be a theory (and we need theories!), it is fair to note that it has not being doing a particularly good job of showing that it is in fact a 'good theory'. This is because, in my opinion, it has not really made good on its own goals of 'explanatory adequacy', and ironically for a theory that started off by placing the scientific object of study within the realm of individual cognition, it has not really seriously engaged with the results or observations from cognitive science. It is true that generative grammar has always situated itself squarely at Marr's computational level (Marr 1982), and has used this to justify the lack of theorizing to the next step algorithmic and implementational levels. But even Marr, in his work on vision, saw filling in those other levels as part of the scientific enterprise he was engaged in, and he theorized about those too. The problem is that if we as scientists do not form linking theories between the computational level and how these tasks are achieved in real brains, then we are in possession of theories that make no predictions whatsoever about data gathered by psycholinguists or neurolinguists. This in turn means that we cannot claim that such theories have higher levels of explanatory adequacy than others with the same descriptive coverage. It seems to me that many of the early claims to 'explanatory adequacy' in the rhetoric of the Government and Binding era rested on the potential to account for the acquisition of language. In fact, it turned out to be much more difficult than anticipated to formalize a concrete implementation of a learning algorithm under a principles and parameters conception that achieves the right results deterministically given the input data and the kinds of cues available to the child (Gibson & Wexler 1994, but see also Fodor 1998). Moreover, acquisition is not the only explanatory adequacy benchmark. Generative grammar (and most especially the minimalist program), needs to begin to seriously theorize about the relationship between its computational theories and how they are embedded within more domain general theories of mind/brain. In other words, it needs to engage with a variety of different performance tasks directly and produce theories of them. The aim should be to get to the stage where our best theories do make predictions about the cognitive processing behaviours we can measure when we are deploying our 'knowledge of language' whether in production or comprehension *or* acquisition. Only in this way can we bridge the commensurability gap between linguistic theory and the cognitive sciences, and only in this way can we assess the explanatory adequacy of these theories. So far, generative grammar has failed to do this. Not because it has tried and failed, but because it seems to have exerted a lot of rhetorical effort in arguing that it should not be required to try.

It is here that I wholeheartedly agree with Chesi that generative grammar needs to up its game if it is to take its place at the table where scientists take the next step towards understanding how language functions within human cognition. Chesi argues that currently the state of formalization within minimalist theorizing lags behind what is necessary if minimalist grammar is to join the conversation at the highest level. This is because formalization is not consistent and is often piecemeal. I agree with this, but to it I would add that a more precise and complete formalization of the computational theories we have is not enough, because such theories do not make predictions beyond that of grammaticality at the level of the sentence. Grammatical theory needs to take the step toward a more algorithmic understanding, and thus open itself up to potential falsification from a wider range of data types. In doing so, it needs to avail itself of all the new mathematical tools and methodologies on offer. I would argue that one of the things we are seeing in this new era of big data and LLMs is that the computational tools and methodologies at our disposal are greater and more sophisticated than ever before, and the data we are able to measure from human brains is getting more and more precise and detailed. It is no longer a pipe dream to think of making good on the promises of real explanatory adequacy. For the same reason, it is no longer defensible to stick cosily to our computational level and virtuously deny any ambitions to make predictions about 'processing'. In making the transition, we will have to engage in new mathematical techniques and methodologies, including computational modelling. And here we come back again full circle to artificial intelligence, neural networks and computational modelling, this time not as theories, but as tools for cognitive science. This is the position laid out by van Rooij *et al.* (2024) who point out that this was originally part of the idea of computational modelling and later connectionist networks (see also Bates *et al.* 1996), not as rivals to a particular theory, or rivals to our human intelligence (AGI anyone?), but as a tool for testing the different theoretical ideas concerning computation and cognition. Language scientists need to reclaim the space for real theory, and not be afraid to use computational modelling as part of their toolbox. Collins

(2024) argues that an expansion of the methodological toolbox has already begun within formal grammar, for reasons that are independent of the sudden success of LLMs. For example, he notes a number of recent approaches that fuse formal linguistic models with information theoretic methodologies (Levy 2008, Goldsmith & Riggle 2012, Rasin *et al.* 2021). He also points out work like Allott *et al.* (2021) which discusses quantitative approaches to sentence processing and how they depend on the Chomskyan program. Collins himself advocates for generativists adopting methods from statistical physics, which can be used to tackle issues of complexity and emergence from a robustly theoretical and rationalistic point of view. When it comes to using the most advanced tools available, modelling via neural nets themselves are valuable tools in the pursuance of the Chomskian questions relating to human minds. I am in complete agreement with van Rooij *et al.* (2024) that we need to reclaim AI from the artificial general intelligence builders and put them to work in the service of answering the scientific questions we care about. Much of what is going on in the AI community is highly relevant and shows great potential for modelling and testing theoretical hypotheses. I sincerely hope that theoreticians and generative grammarians embrace these new possibilities and engage with the advances in the field of computational modelling instead of interpreting LLMs as a rival theory to be attacked. It certainly does not help if LLMs are being hyped as actual theories that are 'refutations' of a whole linguistic tradition and its research questions.

The bottom line is that we have a unique opportunity in this new era to create a genuinely multidisciplinary science of language which involves a whole host of methodologies, with distinct though related research questions. There will of course be the inevitable challenges that arise for generativists in moving into this more multidisciplinary space, and not all generativists need choose to do so. It is necessary to emphasize that there is still also important work to be done in pure language description and comparison that theoretical generativist linguists of all stripes continue to do, using shared or at least intertranslatable analytic vocabularies. I would not be unhappy however to be witnessing the demise of lazy repetition of ideological tropes and thought experiments associated with certain tribal memberships, which stand in the way of genuine re-examination of the issues from first principles as we learn more and more from different fields. If we are lucky, and if we have time to do it before capitalism consumes the planet, then we are on the brink of making genuine breakthroughs in cognitive neuroscience.

Understanding how language fits into cognition is one of the important aspects of cognitive neuroscience more generally, and I think that theoretical linguists can and should be a central part of that scientific conversation going forward. Some of those will be generativists.

Note

[1] Chesi (*this issue*) explains: "One platform designed for performing such linguistic benchmarks is SyntaxGym (Hu *et al.* 2020): an on-line, open-source repository that includes a significant set of linguistic contrasts – 39 test suites that include a total of about 4k sentences. For each relevant contrast included, human generalizations have been gathered in various studies. Direct comparisons of these data with the predictions provided by the models under evaluation is then possible."

Bibliographical References

See the unified list at the end of this issue.