Language Models, Theory, and Tools: A Brief Note

Matthew Fulton

March 23, 2023

1 Introduction

In a short reply to Piantadosi's argument that "Modern language models refute Chomsky's approach to language" this month, Rawski and Baumont (R&B) critique the stipulated logic between what those models are and their correlation with the empirical. It is noted by way of illustration that the geocentric model of the solar system had high predictive power with respect to planetary trajectories, and that the heliocentric theory "predicted worse, yet explained far more." I had myself wondered when reading Piantadosi in what sense language models such as ChatGPT are theories; in certain ways, they seem to be more of a black box than the language faculty itself. To him, a tool which generates the appropriate predictions can be in itself the theory– and I do not doubt that theory is involved in its inferring and setting parameters, nor do I deny the importance of such work. Yet as the critique goes, does it hold suitable explanation? Piantadosi seems to equate a model dubbed relatively unconstrained as not appropriately explanatory, and thus a constrained model with explanation; this logic does not follow either.

2 Prematurity

Roni Katzir points out that Piantadosi's fervor is premature (p.7). In general it is presumptuous and common to dismiss the work of another person or community. Generative grammar has heard every radical dismissal under the sun, so much so that it is a convenient starting point for an undergraduate paper topic– *How are generative ideas completely wrong?* It goes without saying that if one engages in this kind of critique, their ideas should be maturely developed and informed. The author of the diatribe makes the comment that "most parts of generative linguistics" are not sufficiently precise or formal for computational implementation, when it is the case that the generative aims are not necessarily applied; because of this, there is no cause to boast here, and he does not cite anything. It is also the case that certain pressures (to speak, to publish, to actually write something in a timely manner) work against maturity as an ideal.

This offhand reply, as I type it, is puerile as well. But in fact every theory of natural language, with no exceptions, is premature; this claim is uncontroversial. We are all advancing notions in a hybrid process of excavation and imagination. Whether that theory or model of language is embedded in a properly formal, functional, or machine-learning context–or any other distinct or overlapping approach–this is an axiom, trivial but important. Chomsky, Gallego, and Ott (2019) bring this up for generative grammar in this way: "Even within the expressly narrow focus of GG on linguistic competence, virtually every aspect of (I-)language remains a problem" (p.25). This sometimes makes me melancholy. I don't have to tell you that our work jumps the gun; however, we will not wait for the gun but proceed and be justified later (or not); science is deferral. In the same way that theory is premature, *every* dismissal is also premature– and they should not stop dismissing. However, our grace does not extend to claims which will not hold or are not suitable. There is optimism to be taken in the illumination provided by strains of myriad studies (within and without generative circles) over the years, which trace the truth. The theory/model being premature is as much a reason for hope as a reason for sobriety. It is relevant to consider not only what kind of theory can aptly describe what is the case, but also how its explanations may be crafted for understanding by scholars and the public, as well as what kind of explanations we *want* to have. The telos of any theory is arguably not an account of most accuracy and probability (in the sense of the geocentric analogy), but belief in explanations not only powerful in themselves, but also suitable to occupy human minds.

3   Technology

R&B refer to a "theory" contained in the workings of a language model as a "tool-driven kludge." A language model should not itself be the theory nor hold within itself the workings which would hold the content of a theory, because it is a non-explanatory tool. We should not lose sight of the fact that theory is a tool as well, technology, just as language itself is. Both are extensions of ourselves. Katzir writes that language models are successful only in their role as a engineering tools. You cannot make a scarecrow and then begin to investigate human anatomy from there. However, in the case of language models, they are informed by partial ideas about how human language may work (ideas still shaped by their applied purpose), and the scarecrow here is not a still tool like a hammer but is a system with dynamics which are more than the sum of the input in its construction. So there is both the familiarity of the input in its construction, and possible unfamiliarity in its operations and output– unfamiliarity to be observed. Then there is whether the model can work, whether it can be successful not by the parameters you have set for it, but the parameters that you might expect for natural language. But this approximation is purely heuristic as a means to an end, to direct one to a particular kind of testing. This is where scarecrows have their place.

For example, artificial languages have at times illuminated natural languages in ways more or less direct. Artificial learning systems also seek to investigate acquisition. They are of course different from the object of study, and are tools of approximation or contrast. Studying artificial neural networks does not necessarily entail anything about neural networks. But technology will shape the form of knowledge it holds: various kinds of trees and embeddings of matrices, statistical models, and taxonomies, constraint tables and their algorithms, algorithms of linearization and labeling– these being the most apparent tools. You get the picture. Although no one would ever calls these theory-internal tools "theories," the logical leap from what they convey to reality is not so different from the leap from language model to reality, *except that* they were designed with epistemic aims for testing and not as explanans– i.e. the intent of their purpose is important. Some language models (maybe not explicitly called by that name) are designed to be– well, models of language, and escape their engineering purpose; this fact has been neglected, and I see no reason against their application in a search for a model of the truth. Given proper grounding, it might be questionable what the difference between these and the average theory. I also believe that they are useful to tell us the limits (or extensions on limits) of what possibly can be the case for organic computation, as Piantadosi notes on purported barriers to the reducibility of syntactic dynamics to statistics (p.15). These are all applications.

4   Grounding

There is for Piantadosi a problem of explanation, of language models being more than simply tools. But there is (more centrally) a problem of epistemically *grounding* those tools, thought to resemble natural language, in natural language itself– discussed by R&B. I myself have no background in grounding theory (contra "grounded theory"); I must leave the discussion of a method alone. Suppose that you want to a machine which can play cards, and with careful thought as to what we know about card players, reading through all the literature, you flip all the switches necessary to generate the kinds of moves that a professional player would make. You then take a sample of every professional card player in the world with proper controls over an appropriate intervals of time, and with some quantitative reasoning you conclude that the probabilistic correspondence of the machine to the sample is extremely high. Will the machine's principles be an explanation for the mind of a card player? Not that the study is not important, why should they be? Here arises a situation vaguely reminiscent of the issues of commensurability between the objects and processes in linguistic theory and neural systems in Poeppel and Embick (2005). A theory supervenes over facts and would be a scarecrow if its grounding were not clear. Nevertheless our contemporary theories are partially scarecrows, just as a cyborg is partially mechanical, partially organic. There is indeed fiction in our theories, even apart from hypothesis, with the verdict as to their truth deferred. This is necessary (although "fiction" is a purposely inflammatory choice

of word), and we should be comfortable with it even as we seek to fulfill mere forms and stipulation by fact.

Yet grounding is still a vexing shadow. The first time you were in a syntax class, how confused (or angry) were you at the sprawling territory before you? And like a child, the endless question *Why?* To the curious person not previously initiated, finding the ways that contemporary generative syntactic theory supervenes over and is grounded in facts requires an unreserved delve into the extensive and eclectic history which brought it about. At times we may have wanted to painstakingly assemble an understanding by integrating each relevant atomic substantiated claim into a kind of logical super-structure of proof. Aside from what can be proven is: the aptness of representation, certain proposed features and operations, boundaries, the reality of interactions, etc. In the extreme case we would say that we do not know anything about something until we know everything about that something, a form of weak holism– or until we know everything in general (the strong case).

Both are impossible; either amounts to the admission of defeat in the face of premature grounding. So there is necessary fiction. It is true that starting with fiction first may sometimes be preferred for practical, heuristic reasons; it is not actually the backwardness of a logical move from model to reality which is the problem. Piantadosi's aims are not so strange or foreign. In an Abralin Ao Vivo panel talk ("The Minimalist Program: Achievements and Challenges"), David Pesetsky mentions that the Minimalist Program is best viewed as "exhortative speculative fiction." In other words, technology in a broad sense. Of course, a history preceded the program; it wasn't brought forth out of nothing. But it is the heuristic part of the inquiry which was built in as a limb, or even the head, of a scarecrow. But it is difficult to see current language models as exhorting, let alone *undermining* theory.

5   Directions

The judgment for the regrettably ill-equipped and backwards task of using language models to refute theory is solemn: supply a proper theory separate from the model itself to ground (even partially) its fiction in fact in a causally explanatory way; here I add "causally" because of a contrastive discussion of "explaining" regularities in Haspelmath (2021:p.3). I welcome modifications to the criteria above. It will be a set of hypotheses of the most strong severity, demanding of the facts even more than the facts demand of a theory. It is better if the model is designed with that purpose, and further, designed to reveal unexpected patterns from a minimum of input conditions. If refutation is not the goal, the model should be explicitly designed with epistemic aims, and not as a theory. If you do succeed in satisfying such criteria, these skills are useful in grounding (e.g.) the symbolic in the artificially neural, and the artificial in the actually neural; grounding secondary generative notions built

on primary notions, and primary notions on fact; and the transitively effected relations in between. However we should continue to accommodate the proper use of technology in inquiry, given that our own theories are an imperfect form of technology. My last thoughts are that I am disappointed with Piantadosi's ending his article with a call to evaluate Chomsky's legacy (speaking of prematurity), that we should greatly appreciate his grace in highlighting generative contributions, but also that the field is not equivalent to what Chomsky believes. Much less do I speak for generative linguists. But I value the discussion.

References

Chomsky, A.N., Gallego, Á.J., and Ott, D. (2019). Generative grammar and the faculty of language: Insights, Questions, and Challenges. Published in *Catalan Journal of Linguistics* Special Issue 2019. https://doi.org/10.5565/rev/catjl.288

Haspelmath, M. (2021). How to tear down the walls that separate linguists: Continuing the quest for clarity about general linguistics. *Theoretical Linguistics* 47(1/2):137-154. https://doi.org/10.1515/tl-2021-2011.

Katzir, R. (March 2023). Why large language models are poor theories of human linguistic cognition: A reply to Piantadosi (2023).

Piantadosi, S. (March 2023). Modern language models refute Chomsky's approach to language. Retrieved from: https://lingbuzz.net/lingbuzz/007180.

Poeppel, D., & Embick, D. (2005). Defining the relation between linguistics and neuroscience. In A. Cutler (Ed.), Twenty-first century psycholinguistics: Four cornerstones (103–118). Lawrence Erlbaum Associates Publishers.

Rawski, J. and Baumont, L. (March 2023). Modern language models refute nothing. Retrieved from: https://lingbuzz.net/lingbuzz/007203.