DOI: 10.26346/1120-2726-244

On the complementarity of Generative Grammar and Large Language Models

Luigi Rizzi

Collège de France, Paris, France
University of Siena, Italy <luigi.rizzi@unisi.it>

This note intends to stress two complementary aspects of the issues raised by Cristiano Chesi's provocative paper:
(i) Generative grammar and Large Language Models are two separate scientific endeavors, with different goals and methodologies: the first aims at the scientific description and explanation of a natural object, the human language faculty; the second is a technological program aiming at expressing linguistic knowledge in machines, in view of an efficient man-machine interaction. They should be kept carefully distinct. As far as I can tell, the second cannot determine the end of the first, much as the technological discovery of airplanes did not determine the end of the scientific study of flight in nature.
(ii) The two endeavors both deal with the same object, natural language, and have common roots in the theory of computation (the common use of the adjective 'generative' in generative grammar and in generative artificial intelligence presumably is not a mere lexical accident). Rather than being considered in competition, they should be thought of as complementary in many ways. Various forms of collaborations should be envisaged in the future.

KEYWORDS: Generative Grammar, Large Language Models, Minimalism, explanation.

1. Two distinct scientific endeavors

Let me first give my take on the provocative question which entitles Cristiano Chesi's paper. Generative grammar and Large Language Models (LLMs) are two distinct scientific endeavors, which can interact in many ways, but differ profoundly in methods and aims: they should be kept separate and independent, and, as far as I can tell, one cannot determine the end of the other.

Consider an analogy in the domain of natural and artificial flight. The study of natural flight in birds, certain mammals, and insects is a distinct scientific endeavor from the engineering project of building flying machines. The two enterprises are connected, as they were in Leonardo da Vinci's early studies on birds and flying machines, and obviously they can interact, as the results in one domain can inform the other. But they are distinct in goals and methods, and, as far as I know, nobody seriously thought of reducing one to the other. Ornithologists were not deflected from studying natural flight in birds when airplanes were invented.

I believe the same conclusion should hold for generative grammar and LLMs. Generative grammar represents the attempt to describe and understand a natural object, the human cognitive capacity for language, and to explain its properties through the identification of principles regulating the structure and functioning of the linguistic system. The goal of about 70 years of generative grammar has been, and is, to try to reproduce in the study of language the paradigm of explanation that imposed itself in the natural sciences, primarily through Galileo's and Newton's ideas and discoveries. The goal is to deductively connect the empirical observations on language structure and function (gathered through naturalistic observations, or through experiments) to general formal models of the language faculty. The latter notion can be construed in the largest possible way as including whatever makes it possible for humans (but not for chimpanzees, dogs, parrots or other animals) to acquire and use a natural language. So, we are seeking a system of mental principles and properties which are operative in the human mind, some of which may be specific to language, or to human cognition, whereas others may be particular cases of principles operative in a wider array of complex systems. The general goal may be characterized as the study of the 'Faculty of language in a broad sense', in the terminology of Hauser, Chomsky & Fitch 2002. Part of this system will plausibly turn out to be species-specific and task-specific, the 'Faculty of language in a narrow sense'. But for the purposes of this note, we may just focus on the broader system.

Generative artificial intelligence and LLMs are a technological endeavor to have machines learn and internally represent languages in the aim of permitting a smooth and fluid communication between man and machine through natural language, and for a variety of practical tasks, including translation. The progress made possible by such models is astounding: three years ago I would not have thought that I would see in my lifetime a truly successful artificial language processing system, capable of generating texts and interact in natural language with humans, in a way that could fool a human observer on its natural or mechanical origin, thus qualifying as serious candidates to pass the Turing test. Now such systems exist, and they acquire an ever more pervasive influence on so many aspects of our lives.

Sometimes, the success of LLMs is interpreted as meaning that we can do away with formal theories of language, and perhaps with linguistics as an independent domain. After all, if very general computational devices able to learn all sorts of patterns can figure out human language, wouldn't that already provide all we need for capturing language, without having to assume formal models dedicated to language, or even an independent discipline devoted to the scientific study of language? Arguments of 'eliminative reductionism' from the 1980's resonate in this reasoning.

2. Explanation and intelligibility

Does the existence of such technological marvels imply that we do not need formal linguistic models anymore? I do not think so. Chesi mentions intelligibility in passing a couple of times in the target paper, but the point deserves center stage, in my opinion. The general goal of generative grammar is to deduce the empirical generalizations observed in linguistic data from systems of general and abstract principles. In this particular way of understanding scientific explanation, intelligibility plays an essential role: the principles must be intelligible, and each step of the deductive connection between principles and empirical facts (of acquisition, and of adult language knowledge and use) must be accessible and transparent to the researcher. In other word, the primary aim of the researcher is to look inside the black box of the human mind/brain, make it as transparent as possible, and elucidate its components and functioning in intelligible ways.

If we continue to understand 'explanation' in the Galilean sense, then questions of explanation are not automatically and directly addressed by systems mimicking the empirical pattern.

In fact, explanatory questions may be legitimately asked of artificial systems as well: Why do artificial systems significantly succeed in mimicking the human capacities? How do they capture the empirical generalizations that the empirical linguistic work uncovers? What internal structural properties do successful devices have? Why not other imaginable characteristics? How does the structure and functioning of artificial systems compare to the structure and functioning of the natural system, implemented in the human brain? Mimicking a system is not sufficient for explanation: we want to understand why and how the artificial system works. And, first of all, we want to assess in detail how well they work. Here the contribution of scientific linguistics is essential.

3. Some results of generative grammar

Why did linguists bother with the building of explanatory models for language? The search for explanation through intelligible principles has permitted, through 70 years of generative grammar, to gain much insight on the structure and functioning of natural language. The deepening of explanation also showed a remarkable heuristic capacity, leading to the discovery of innumerable empirical generalizations holding for hundreds of languages submitted to rigorous generative analysis. Let me just mention three major areas of empirical discovery:

a. The HIERARCHICAL ORGANIZATION of linguistic expressions. A hierarchical relation like c-command (Reinhart 1978) has a pervasive role in

- syntax: syntactic locality is checked on hierarchical representations, not on linear sequences (Rizzi 2013, 2021 for discussion);
- morphosyntax: c-command and hierarchical locality govern the whole functioning of the case-agreement system (see, e.g., the discussion in Baker 2013);
- interface with meaning: all referential dependencies, binding, coreference/non-coreference, etc. are ruled by c-command, as shown by an immense literature stemming from Tanya Reinhart's seminal work (see, e.g., the discussion of Demirdache *et al.* 2024, also in connection with the performance of LLMs on referential dependencies); similarly, interpretive interactions between different quantificational elements are ruled by c-command;
- and even the interface with sound: phono-syntactic phenomena are sensitive to c-command (Manzini 1983, Rizzi & Savoia 1993).

Whenever a linear and a hierarchical analysis compete, the hierarchical analysis unerringly turns out to be empirically correct. For instance, verb agreement with a nominal expression never is with a linearly adjacent noun, but with a noun in a certain grammatical relation, the head noun of the subject noun phrase (i.e. in the sentence *The picture of the trees is here* the verb agrees with *picture*, not with the linearly adjacent *trees*). The antecedent of an anaphor is a c-commanding expression, not a linearly close expression (e.g. in *The man who saved the boy hurt himself* the antecedent of *himself* is the man, not the boy). And so on and so forth. Moreover, the study of the hierarchical structures led to the identification of rich sequences of functional elements characterizing the internal structure of clauses and phrases, an aspect studied in detail in cartographic research (Cinque & Rizzi 2010, Rizzi & Cinque 2016).

b. The ubiquitous manifestation of the DISPLACEMENT PROPERTY: the fact that certain expressions are pronounced in a position different from the position in which they are interpreted. For example, the fact that a *wh*-phrase is pronounced in initial position in languages like

English, but must be construed with the argument structure of a verb which can be indefinitely far away (*Which book do you think that we should read _?*). Whether or not the particular grammatical model one adopts has an independent and dedicated operation of 'movement' (e.g. the Government Binding model does, whereas Minimalism, Lexical-Functional Grammar and Head Driven Phrase Structure Grammar do not), such discrepancies between sound and meaning are pervasive in natural language. Displaced configurations give rise to systematic 'reconstruction effect' (i.e., in *Which picture of himself do you think John likes _?*, the phrase *which picture of himself* is interpreted in the trace position, the object position of *likes*, where *himself* is c-commanded by *John*, hence it can be properly interpreted with *John* as antecedent): the mind 'sees' the displaced element in a position that is not pronounced for the computation of numerous interpretive properties.

c. Syntactic dependencies obey LOCALITY CONSTRAINTS, which give rise to a precise and detailed typology. On displacement dependencies we have strong islands, giving rise to severe unacceptability on all types of movement: for instance, extraction is generally barred from a relative clause (**Who did you talk to the man that saved _?* meaning 'Which is the person such that you talked to the man that saved this person?'). Whereas other environments give rise to weak islands, typically showing argument-adjunct asymmetries (for instance, extraction of an argument from an indirect question can be marginally acceptable, as in *?Which problem do you wonder how to solve _?*; whereas extraction of the adjunct is more severely ill formed (e.g. the previous example with the two *wh*-expressions *which problem* and *how* that exchange places: **How do you wonder which problem to solve?*). There are partially similar locality constraints on interpretive procedures such as the binding of anaphors (*John said [that Bill praised himself]*: of the two c-commanding nominals only the structurally closer one, *Bill*, can be the antecedent of *himself*); and the control of null subjects in embedded infinitives (*John convinced Bill [_ to leave]*: Bill, not John, is the one who will leave).

Why do we find such properties in natural languages, rather than many imaginable alternatives (e.g. systems with purely linear dependencies, systems without displacement, where everything is interpreted where it is pronounced, systems with different, or no locality principles)?

The Minimalist Program put forth the hypothesis that this is a matter of simplicity. Suppose that natural languages use an extremely simple combinatorial device, an operation called Merge that says 'take two expressions A and B and form the complex expression [A B]'. The operation is recursive, and this captures the unbounded nature of language. It automatically determines the hierarchical structure expressed by syntactic trees. If we assume that the operation can apply with maximal freedom, it can both put together two separate elements (*see* and *Mary*, to form the verb phrase *see Mary*), or take an element that already is part of a larger structure and 'remerge' it with the whole structure, thus capturing the displacement property: external and internal Merge. Locality constraints operate on Merge to make derivations of sentences as economic as possible in terms of computational resources. If all this is on the right track, many basic constitutive properties of natural language are deduced from extremely simple hypotheses on the combinatorial system.

4. Formalization and levels of empirical adequacy

Chesi underscores the fact that generative analyses tend to be not fully formalized. This has the consequence that such analyses tend not to reach observational adequacy, i.e. the capacity to assign a structural representation to each sentence in a given domain (the Language Problem, in Chesi's terminology). LLMs on the other hand, are fully implemented and as such offer an exhaustive coverage of the sentences in a given domain. On the level of observational adequacy, LLMs can therefore be said to be superior to generative analyses. This is correct, but it is important to understand why it is so. In fact, neglecting observational adequacy is a precise choice that formal linguistics has made, in order to be able to focus on explanatory principles: that kind of endeavor inevitably leads to neglecting certain aspects of the empirical domain, to focus on aspects that permit the elaboration of an explanatory model. After all, other disciplines work like that: physics does not aim at capturing all the phenomena that take place in a cubic meter of space: rather, explanatory physical models will select certain patterns and elucidate them through the interplay of abstract principles. There is a trade-off between exhaustiveness (in the sense of observational adequacy) and explanatory depth: in a formalization covering the totality of the empirical domain, explanatory principles would be drowned by the details, and would fail to emerge. Once again, the intelligibility of the explanatory system has been and is the driving force in generative grammar. The engineering project, given its practical goals, cannot avoid empirical exhaustiveness (observational adequacy); consistently, it does not aim at the intelligibility of the underlying principles. Or, at least, intelligibility is not its immediate aim. But nothing precludes the possibility that explanatory questions may be asked of artificial systems.

5. Forms of collaboration between generative grammar and Large Language Models

LLMs need precise benchmarks to test progress in the mastery of natural language. As far as I can tell, major benchmarking tools are largely based on generative research (e.g. Hu *et al.* 2020 and the other cases discussed by Chesi in the target paper), for good reasons. Generative grammar can offer 70 years of analytic experience with the fine details of linguistic structures across hundreds of different languages. This is of critical importance to set up efficient testing grounds for computational models. It would be irrational for the engineering projects not to use the formidable expertise with linguistic structures that generative grammarians can offer. That is why natural language processing projects need formal linguists, and (*pace* Frederick Jelinek's famous dictum) big companies regularly hire students trained in generative grammar.

Reciprocally, I am convinced that what LLM research has to offer to generative grammar and the cognitive science of language is of crucial importance, provided that one avoids drawing hasted analogies between natural and artificial systems. An artificial neural network bears only a very vague resemblance to the neural structures of the human brain for numerosity, organization, internal structure and functioning of natural and artificial neurons. Moreover, the size of the training set in LLMs typically is several orders of magnitude bigger than the primary linguistic data that children have access to (the utterances the child hears in the course of language acquisition). And LLMs may well be at ease with 'impossible languages', systems with rules and properties that no natural language includes, and that children never conjecture in acquisition (Moro *et al.* 2023). These differences are all too obvious, and they should not be forgotten, otherwise any inference based on the analogy would be unwarranted.

Having said that, we should not underplay the points that the two endeavors have in common, and make collaborative projects possible and desirable. They both deal with the same object, natural language, and have common roots in the theory of computation. The common use of the adjective 'generative' in generative grammar and in generative artificial intelligence may not be a mere lexical accident: even though the term 'generative artificial intelligence' may be primarily intended to stress the capacity of the artificial systems to produce texts, etc., I like to think that this terminological choice pays a tribute to the pioneering role that generative grammar had in the study of the computational foundations (the Chomsky hierarchy, etc.) and in the scientific study of language.

I believe LLMs offer an opportunity for addressing from a new angle questions linked to the nature and acquisition of knowledge. Why are LLMs so successful? How do they learn languages (and many other things)? What quantity and quality of empirical evidence is necessary for them to learn languages successfully? How does that compare to the evidence that natural learners require? Are there areas of language that are problematic for artificial systems and easy for natural learners? or vice versa?

In general, I think the search for explanatory principles can legitimately be pursued for artificial systems, much as it has been pursued for natural systems. Linguists have tried to open the 'black box' of the language faculty, elucidate its internal structure, identify in an intelligible manner the principles constraining its functioning. Both abstractly, with functional abstract models, and more and more concretely, with neurolinguistic models. One can imagine pursuing the same logic with artificial systems, trying to open the black box of artificial neural systems and study how the knowledge of language gets organized in such systems, and on the basis of what intelligible principles. Current work using 'ablation' techniques (deactivation of certain artificial neural structures), inspired by lesion studies in neurolinguistics, looks promising and suggestive (Lakrets *et al.* 2019). And the study of 'learning biases' that may be necessary for artificial systems to acquire certain structural properties (e.g. Mitchell *et al.* 2019 on Principle C and referential dependencies) invites a comparison with principles of the language faculty postulated by linguists.

Among other things, such a comparison will help disentangle principles specific to the human language faculty from more general principles organizing complex systems (Chomsky's 2005 'third factor' principles), biological and not. So, a comparison between natural and artificial intelligent systems for language (and other domains of knowledge), using the best tools made available by linguistics and computer science, may be of decisive importance for a better understanding of what 'learning' means, a crucial question for the future of cognitive science.

Bibliographical References

See the unified list at the end of this issue.