DOI: 10.26346/1120-2726-238

LLMs and syntactic theory: Lexical Functional Grammar has no fear

Miriam Butt

University of Konstanz, Germany <miriam.butt@uni-konstanz.de>

This paper responds to Chesi's paper *Is it the end of (generative) linguistics as we know it?* from the perspective of a computational linguist who works within one of several generative syntactic frameworks, namely Lexical Functional Grammar. Chesi's general conclusions and recommendations to Minimalism for the best way forward are found to hold, yet the diagnosis of the precise causes and therefore also the suggestion of remedies differs. Overall, I propose that the chances and opportunities offered by advances in machine learning in general and Large Language Models in particular should be studied carefully and integrated into a model of language which combines a rule-and-symbol driven approach with gradient and probabilistic information.

KEYWORDS: LLMs, Linguistic Theory, computation, generative syntax, LFG.

1 Introduction

In his thoughtful paper *Is it the end of (generative) linguistics as we know it?*, Chesi responds to Piantadosi (2023). Chesi partly agrees with some of Piantadosi's assessments and concludes that generative linguistics needs an 'update' in terms of both methodology and attitude. In this paper, I lay out a perspective from within generative linguistics, but from a minority theory. I find that I agree with Chesi's conclusions on the whole, but differ with respect to the causes and recommendations of remedies.

Chesi focuses on three major issues: (i) the state of generative syntax; (ii) the connection to computational feasibility and standards of evaluation; (iii) the most recent manisfestation of the Poverty of the Stimulus argument with respect to LLMs. I point out that the measures and remedies Chesi proposes already exist within generative linguistics in general and generative syntax in particular. However, researchers working within 'mainstream Minimalism' in practice have tended to isolate themselves from this work, much to the detriment of linguistics. My overall recommendation is thus to take on a pluralistic perspective, including with respect to LLMs, which will consequently allow for a broader engagement and deeper understanding of the underlying issues.

2 Generative linguistics

2.1 Who gets to be a generative linguist?

Chesi begins with a reference to a 2013 roundtable designed to identify the Hilbert questions of syntax. I followed blogs and discussions around that conference and, like Chesi, came away disappointed. Chesi also lists the invited participants, all prominent syntacticians. However, notably absent are prominent syntactic thinkers like Joan Bresnan or Ray Jackendoff.[1] Both have contributed significantly to our overall understanding of syntax and each has (co)formulated a theory of syntax that works within the general principles and understanding of generative linguistics. If one had really wanted to come up with a generally applicable list of target problems and to delve into a deep understanding of syntax, shouldn't a plurality of perspectives have been invited?

Indeed, while Chesi includes multiple references to 'generative grammar', 'generative linguistics' and 'generative linguists', he does not actually refer to any theoretical framework other than the Minimalist Program and only means Minimalists when making reference to 'generative linguists'. If the idea is to worry about the state of Minimalism, then why not say that directly – why frame the issues as if they should apply to all generative linguists?

I have no ready answer to this and find the systematically inward looking practice of Minimalism puzzling. In my own work, I have found it incredibly rewarding to discuss ideas, opinions, and analyses with researchers who do not share the same theoretical framework or quite the same underlying assumptions. While I work mainly within Lexical Functional Grammar (Dalrymple 2023), I have published work using Minimalist assumptions (e.g. Butt & Ramchand 2005) and have worked on prosodic and semantic/pragmatic issues that go beyond my core syntactic interests. In each case, my understanding of syntax has been enriched and deepened.

Lexical Functional Grammar (LFG) is just one of several generative syntactic theories. Others include Combinatory Categorial Grammar (CCG), Head-Driven Phrase Structure Grammar (HPSG), Tree Adjoining Grammar (TAG) and the Lexicalized Tree Adjoining Grammar (LTAG) version. Each of these work with different assumptions as to grammar architecture, but each of these is very definitely an instantiation of a generative approach to syntax.

The problem with Chesi framing his discussion as pertaining to generative linguistics in general is that the issues raised in Chesi's paper do not apply equally to all of these alternatives. In their sum total, they apply only to Minimalism. A generalization from Minimalists being in crisis to all generative linguists being in crisis is thus unwarranted to say the least.

> While generative linguistics struggles to accommodate gradual judgments, online effects, and other kinds of implicit data, these are the daily bread of computational and psycholinguistic models. (Chesi *this issue*: 39

For example, the above assertion is simply untrue for LFG, which was designed to be computationally implementable and psycholinguistically realistic from the outset. Subsequently, Bresnan's pioneering work on the dative alternation, for example, showed that gradual judgements and probability effects needed to be accounted for (Bresnan *et al.* 2007; Bresnan 2016). Similarly, while working on the development of industrial-strength large computational grammars (Butt *et al.* 1999; Sulger *et al.* 2013), it was necessary to weight rules and lexical items in terms of preferences and probabilities. There was no 'struggle' to do this – rather, it involved designing an extension of the theoretical framework in combination with clever engineering to combine knowledge about probabilities and preferences with a basic rule and symbol driven approach (e.g. see Frank *et al.* 2001; Riezler *et al.* 2002). In very recent work, the incorporation of gradient information is being extended to the prosody-syntax interface, e.g. see Bögel *et al.* (2024); Bögel & Zhao (2025).

Stepping out into a meta-perspective, it is therefore interesting to observe: (a) how researchers within Minimalism on the one hand tend to marginalize or even 'erase' alternative ideas within generative syntax, casting themselves as the only game in town; (b) at the same time worry about their own marginalization.

Recommendation 1. Researchers working within Minimalism should be aware that other approaches within generative syntax exist and they should endeavor to achieve some degree of fluency in these frameworks. In effect, they should become multitheoretically competent so as to be able to take on ideas and insights coming from other parts of generative syntax, rather than revisiting issues and discussions potentially fruitlessly within an echo chamber.

2.2 Achievements obscured by jargon

Related to the above recommendation is the use of jargon in presenting analyses and insights. Every discipline is of course entitled to its own specialized vocabulary and needs to be able to engage in in-depth and highly technical discussions that will be more or less incomprehensible to non-specialists. However, at some point, one should be able to take a step back from one's own research and be able to articulate exactly why this research is important and why non-specialists should care about it. Generative linguistics was extremely successful at this in the early days, but seems to have all but lost this knack (*pace* recent efforts of scholars such as David Adger and Ian Roberts).

Example for jargon. In recent years there has been a welcome rise in numbers of generative linguists working on South Asian languages, my field of expertise. Much is being discovered, there is a palpable feel of a field moving forward. However, more often than not, I find that when engaging with early career researchers working within Minimalism, they are unable to articulate their findings without using theory-internal concepts such as Merge, Agree or invoking theory-particular constructs such as the EPP. When I then ask them what my take-home message should be under the assumption that I am not convinced of their particular theoretical approach, I have far more often than not been met with a stunned silence. That is, I find that researchers within Minimalism tend to be unable to separate out which parts of their analysis are entirely due to theory-internal reasons and would therefore find little resonance outside of the theory, and which constitute genuine insights that need to be transported across theories.

Recommendation 2. Linguists within Minimalism should be taught to articulate insights in a manner that can step away from theory-particular reasoning and assumptions.

Achievements obscured. I recently found myself at a conference dominated by typologists and descriptive linguists working on a host of interesting languages and phenomena. However, they were unable to distinguish between control/raising, adverbial predication, serial verbs and complex predicates. I found myself explaining the basics of control/raising to a number of very intrigued researchers who had never been taught to look beyond the surface. I came away depressed: there are so many more functional and descriptive linguists in the world than there are generative linguists and if the considerable achievements of generative linguistics have not been able to make their way into a broader community decades after having been discovered, then how can linguistics as a whole progress? Indeed, with respect to the Hilbert roundtable, one thought I had is that rather than trying define tricky and abstract problems in linguistics to be set out as challenges to the next generations, it would be good if generative linguists could come together and instead make a list of all the major things we have already found out about how languages work. Frans Plank's *Universal's Archive*[2] aimed to do just that: collect the generalizations posited by linguists, along with possible exceptions to the generalization, pointing towards avenues for further necessary research.

Recommendation 3. Find ways of articulating the findings of generative linguistics so that they resonate outside of the field.

Articulating both achievements and challenges should then also include taking on new insights in terms of the role of gradient phenomena and probabilistic information, as well as results from research into the interfaces of syntax with other components of grammar (e.g. prosody, semantics, pragmatics), rather than either arguing strenuously against the existence of such effects (section 4) or clinging to a model of grammar that does not do justice to the sum totality of existing observations (section 3.5).

3 Standardization and consolidation

3.1 Computational feasibility

> Furthermore, the general underevaluation of experimental and computational advancements by leading scholars in the generative field has contributed to a perception of generative linguistics as marginal within both computational and experimental language research communities. (Chesi: 6)

This observation by Chesi is entirely true. Going even further, one can observe that of the various generative syntactic frameworks that have been formulated, Minimalism and its precursors are the only ones that have never had any significant computational presence. The reason for this seems to be that despite its early computational promise, Minimalism and precursors were never formalized enough to make them computationally feasible. As noted by Chesi, the work done by Edward Stabler is exceptional in this respect. However, as also noted by Chesi, most Minimalists would not feel entirely comfortable with the computational version that Stabler has crafted.

The situation is entirely different for the other approaches within generative syntax. They were not only created with computational feasibility in mind, from the get-go they were accompanied by serious formal mathematical modeling. A small community of specialized scholars is looking at this within Minimalism, but there is a clear disconnect between the linguistically oriented practitioners of Minimalism and the mathematically oriented ones. Again, this is different for the other approaches, where Aravind Joshi (TAG), Ron Kaplan (LFG), Carl Pollard (HPSG) and Mark Steedman (CCG), for example, were always interested in both mathematical foundations and linguistic insights (e.g. see Kaplan 2019 for a history about the foundations of LFG).

Given that the jargon surrounding Minimalism (section 2.2) is filled with mathematical symbols and uses terminology like 'computation' and 'enumeration', it is more than ironic that this is the only approach within generative syntax that strugges to establish computational feasibility (except for in Stabler's version). Again, this is entirely different for CCG, TAG and to some extent HPSG, where the computational realization of the framework is inextricably bound up with the formal linguistic analyses. Within LFG, the computational and theoretical communities exist side by side, with some practitioners such as myself straddling both. However, the learning curve for a theoretical LFG linguist interested in using the computational version is not large: the computational version more or less mirrors the theoretical version.

3.2 Computational applications

Moreover, while CCG, HPSG, LFG and TAG have very clear interests in explanatory adequacy they also all have interests in applying this knowledge towards computational applications for Natural Language Processing (NLP). Beyond grammar development, traditional applications have included parsing, generation, machine translation (MT), Question-Answering, Computer Assisted Language Learning (CALL), text summarization, anaphora resolution, natural language understanding (NLU) and inferencing (NLI). It has also included the creation of computational resources, for example, in the form of lexicons, treebanks or other types of annotated corpora. The Minimalist footprint in these areas is minimal and so it is no surprise that Minimalism is at best seen as marginally relevant within computational language research.

3.3 Syntax as a solved problem

Compounding the perception of marginal computational relevance is that Minimalism sees syntax as an unsolved problem with a number of exciting and interesting challenges ahead (hence the Hilbert roundtable), but from the computational perspective syntax is essentially a solved problem. This is true both for machine learning, on which the current LLMs are built, and the rule-and-symbol based approaches.

I gave a talk in 2008 at the annual meeting of the German linguistics society,[3] in which I noted that syntax was essentially solved from the computational LFG perspective and that the current interesting frontiers for research revolved around the integration of semantics and reasoning. Indeed, while the annual LFG conferences[4] see their share of 'core' syntactic work, most of the interesting discussions within LFG currently revolve around interfaces with prosody, morphology, semantics and pragmatics.

This stands in stark contrast to Minimalism, which seems to revisit the core mechanisms posited for syntax time and again, with a latest iteration found in Chomsky (2021) and Chomsky *et al.* (2023). From a computational perspective, the perception that Minimalism is not even sure yet how best to do syntax makes it an unviable proposition. This is compounded by the existence of very robust parsers with reasonable linguistic output based on machine learning like the Stanford Parser,[5] let alone the capabilities of current LLMs. The existence and good performance of LLMs and the smaller machine-learning based parsers renders any Minimalist foray into computation as irrelevant at this stage in the game.

Indeed, even the rule-and-symbol based technology of the other generative syntax approaches has been able to generate large-scale, fairly robust grammars for at least English. For example, the LFG technology was able to scale up sufficiently for the start-up *Powerset* to emerge out of the Natural Language Research group at PARC and to eventually be bought for an estimated $100 million by Microsoft in 2008.

The English grammar at the core of that technology is still freely accessible and can be interacted with via the INESS infrastructure site.[6] The INESS site also collects a number of other grammars, all developed with the same grammar development environment and with a common set core guiding principles as part of the ParGram (Parellel Grammars) effort (Butt *et al.* 1999; Sulger *et al.* 2013). Much of the ParGram effort involved creating benchmarks, agreeing on standardized analyses across languages and the creation of common resources: something Chesi recommends Minimalism adopt.

3.4 The importance of dependency information

The collection of generative syntactic approaches instantiated by CCG, HPSG, LFG, Minimalism and TAG each offer up very different ideas for linguistic representations and how to arrive at them. Generally this includes a mix of hierarchical relations, linear order, dependencies and semantic information. It turns out that the most useful type of information for most NLP applications is either in terms of low-level n-gram collocations (types and instances of tokens cooccurring in a window of n tokens, this is essentially what LLMs are built on) or in terms of dependency relations.

LFG represents dependencies directly in terms of its core syntactic f(unctional)-structure representation. Within NLP more generally, the Universal Dependencies framework[7] has become extremely successful, providing manually annotated corpora of dependency relations for a host of languages. These treebanks are used as the basis for machine learning of dependency parsers, which can then automatically identify information about dependencies to be fed into NLP tasks such as Question-Answering or NLI. The dependencies provide basic information about who did what to whom, where, how and (possibly) why. This clearly very useful information is only very indirectly recoverable from Minimalist representations, rendering Minimalism uninteresting from this NLP perspective as well.

3.5 Measuring progress

3.5.1 Empirical benchmarks

Let us now turn to a very important point made by Chesi. He points out that in comparison to the standards in computation and experimentational approaches, Minimalism lacks benchmarks and concomitant efforts at standardization. Chesi already points towards some useful ways to ameliorate this situation, but from my perspective the heart of the problem lies in the following observation.

> I think the original sin of most generative linguists is that they have gotten used to incomplete pseudo-formalizations and data fragment explanations. (Chesi: 40)

For any field, there should be a way to measure progress. In linguistics, we can measure progress in terms of how much we understand about the structure and use of language and we can measure this concretely by amassing data sets that can serve as bench marks. Collecting such data into so-called ‘testsuites’ by which coverage and progress can be measured takes a bit of effort. But even in the absence of such a concerted effort, it should be considered standard practice that if a theory has been developed with respect to a certain body of data, for example, a theory of control and complementation (Bresnan 1982), then any extensions of that same theory or a new theory altogether should be able to account for at least that body of data and, in an ideal world, for an even larger body of data.

Minimalism is alone in the collection of generative syntactic theories in not following this good scientific practice. The use of "most generative linguists" in the above quote is therefore deceptive. The observation is indeed probably true of most generative linguists, but what the discussion does not mention is that all of the set of generative linguists for whom the statement is true is contained within the set of Minimalists. The statement thus obscures the fact that there are very different practices and attitudes available within generative linguistics as a whole.

3.5.2 Falsifiability

A great strength of the Minimalist approach (and its precursors) is the formulation of strong hypotheses driven primarily by theoretical considerations. This stands in contrast to LFG, for example, whose mode of scientific inquiry is more inductive: one examines a body of data, makes hypotheses based on that data, refines or discards the hypotheses given more data, etc. One can see Minimalism as being driven primarily by top-down considerations, while LFG works more bottom-up. Yet this great strength of Minimalism is marred by the practice in the field, which as Chesi observes, is full of partial formalizations and explanations of fragments of data. Given the many only partially specified and often incompatible versions of Minimalism, each with different sets of assumptions of how things might work if one got around to working them out properly, the predictive power of the approach is weakened considerably.

> If a theory is not fully explicit – i.e. formalized – there is no way to make precise predictions. For computational linguists accustomed to running their models on a computer, it is a well-known fact that no external oracle can ever fix a bug or a gap in the theory. (Chesi: 16)

The quote above points out that if a linguist is not forced to have their ideas checked systematically either against bench marked data or by running their analyses in a computational system, then a certain predilection for either overlooking or ignoring a gap in the theory is bound to occur.

In the worst case, what happens is that a theory becomes unfalsifiable even in the face of empirical evidence and formal issues, as is the case with the problems pointed out by Abels & Neeleman (2012) with respect to Kayne's Linear Correspondence Axiom (LCA) (Kayne 1994). Chesi discusses this issue and concludes:

> While Abels and Neeleman's critique raises significant questions about the formal legitimacy of LCA, offering what appears to be a more elegant solution to the generalizations of GU20, an ultimate comparison of these proposals – considering simplicity, descriptive adequacy, and

> explanatory power – must be achieved. (Chesi: 37)

Though Abels and Neeleman present strong arguments, there is clearly a reluctance in the field to let go of the nice idea represented by the LCA, apparently simply because it was a nice idea. From the perspective of a comparative outsider, it is also astonishing to observe that Kayne's original formulation was in 1994, Abels and Neeleman's critique in 2012 and yet here we are in 2025, concluding that the state of the art is that an ultimate comparison of these proposals still remains to be achieved.

Recommendation 4. (a) Follow Chesi's recommendations in terms of bench marks to measure progress; (b) include a computational modeling system by which the repercussions of analyses can be checked, e.g. along the lines of what was done for Optimality Theory;[8] (c) hold work to greater accountability in terms of both empirical coverage and formal implementation.

Following these recommendations will certainly help to alleviate one large problem identified at the Hilbert roundtable, namely:

> More crucially, it was practically impossible to present all the problems in a concise and coherent manner within a consistent framework: in nearly every instance, although each problem statement came with a proposed solution, the underlying assumptions were often at odds with the premises of others. (Chesi: 4)

4 LLMs: Poverty of the Stimulus reloaded

In this section I turn to the original reason for Chesi's paper: Piantandosi's claim that the modern LLMs refute Chomsky's approach to language. As pointed out by Chesi, this is the old Poverty of the Stimulus argument reloaded. The argument essentially always pits a rules-and-symbols approach against machine learning, with the current LLMs just the most current instantiation of machine learning.

Much has been said at every iteration of the Poverty of the Stimulus argument, including for this latest version (e.g. see Kodner *et al.* 2023 and keep an eye out for Bender & Hanna 2025). This includes pointing out the vast amounts of data and the vast amounts of energy resources that need to go into training the LLMs, as well as what is known as the 'grounding problem'. LLMs are essentially large probabilistic predicting machines that are extremely good at predicting the next most likely byte, character, word, sentence, text, etc. They do this by dint of figuring out distributional patterns in large amounts of data, essentially solving the problem of predicting and modeling syntactic distribution (cf. section 3.3). However, by themselves, they do no more and no less than that (though this basic capability can be put to extremely effective use in a multitude of NLP tasks). While the problem of recognizing distributional patterns in a language is thus essentially solved, LLMs fall short in terms of independent reasoning and semantic understanding. This is why NLU and NLI are currently hot areas within NLP.

Bender & Koller (2020) point out the distinction between predicting form and understanding via a thought experiment. This involves two humans, each stranded on a separate island, but who can communicate with one another via an underwater cable. A sea creature has found a way to receive and send signals via that cable and has 'listened in' long enough to be able to successfully impersonate a human communicator. One day one of the humans is attacked by a hitherto unknown creature and asks for help. The question is, in the absence of being 'grounded' in the reality of living on an island (rather than in the sea), will the sea creature be able to come up with an innovative and successful proposal for a defense? Bender and Koller's answer is in the negative.

We might similarly ask whether LLMs are capable of language change and whether they would have been able to come up with the concept of emoticons completely out of the blue, as humans did.

In any case, LLMs do not pose a threat to any linguistic enterprise that seeks to understand how language works in the human brain. LLMs do not model a human brain and they are not grounded in the human world of experience. They do not breathe, they do not feel hunger, they do not age and they are not driven by urges to procreate. Looking at papers like that of Dennett (1978), which could have just as well been written today, it seems we keep having to reiterate the same conclusion: no matter which kind of technological advances we make, if we are interested in understanding how humans work, then our primary object of research needs to be humans. Of course, we can and should adduce information from the implementation and results of the technological advance to aid us in our search for understanding human language.

It is perhaps this last point that underlies the flurry of papers surrounding each iteration of the Poverty of the Stimulus argument: what (if anything) should linguists take on board from technological advances, particularly with respect to machine learning? As with many debates, I suggest the answer lies not in the extreme positions, but somewhere in between. It is clear that humans operate with symbols (e.g. Deacon 1997). However, it is just as clear that humans are very good stochastic predictive machines, given that there is ample evidence from psycho- and sociolinguistics that humans attend to frequency information and the comparative likelihood of items occurring together.

The obvious conclusion is that any model of human language needs to be able to account for both the rule-governed symbolic parts and the predictive, experientially learned parts. LFG in effect has already gone down the road of building such ‘hybrid’ models by positing a rule based core that can be overlaid with information as to preferences and probabilities. Going even a step further, we can investigate what the computational advances might offer up in terms of further informing our models and thereby our understanding of language.

For example, in recent work Kalouli (2021) worked on the problem of NLI and found that while BERT (the state-of-the-art LLM at the time) does poorly on logical reasoning, it does well on adducing so-called ‘world knowledge’ and applying this towards NLI problems like deciding whether the statement *A child is buying a dog* can be true in light of previous knowledge like *A girl is purchasing a pet*. While manually created lexical resources like WordNet[9] go a long way towards helping resolve inferences like this (e.g. see Bobrow *et al.* 2007), a very strong suit of LLMs is their ability to extract and store information about the Distributional Semantics of words (also referred to as ‘vector semantics’ or ‘word embeddings’). The input and thereby also the interim representations contained in LLMs are high-dimensional vectors. Vectors are mathematically well understood objects that can be represented in a two-dimensional space and that can be calculated with easily. The vectors containing information about the distributional properties of a word can thus be represented in a vector space and can be used for calculations as to the comparative distance between meanings (using cosine similarity, for example). The overall effect is that words occurring in similar contexts, like *dog, cat* and *pet*, will be clustered closer together in a vector space, as opposed to words like *tiger* or *chair*.[10]

Kalouli ended up building a hybrid model, which combined the power of symbol-driven syntactic and semantic parsers for logical reasoning with the power of subsymbolic (neural net) driven information about distributional semantics. Her model outperformed the state-of-the-art at the time (Kalouli *et al.* 2020) and thus provides a strong argument for pursuing hybrid models of language.

With respect to LFG, for example, Mark-Matthias Zymla is currently looking at integrating information from word embeddings directly into the current computational model of LFG, which has (among other things) extended the grammar development platform XLE to include a semantic construction module (Dalrymple *et al.* 2020; Zymla 2024) that utilizes ‘glue semantics' (Dalrymple *et al.* 1999). While LFG (naturally) makes extensive use of lexicons, the precise meanings of lexical items have been notoriously difficult to describe. Word embeddings offer a neat new way forward with respect to this long-standing problem (cf. also Gehrke & McNally 2019): one can approximate the meaning of a word via information about its distribution in manifestations of language use (i.e. corpora). Whether this potential new avenue will be successful remains to be seen. However, excluding such an avenue of research on principle seems to be the entirely wrong way to go. In recent work, for example, we have been able to not only theoretically model, but also computationally implement a system within LFG that allows us to take a real speech signal as input and produce a semantic representation as output (Butt *et al.* 2024). As far as I am aware, LFG is to date the only approach within generative syntax that allows for a seamless building of syntactic and semantic analyses directly on the basis of information coming from a speech signal. Achieving this required putting into a place a well-defined syntax-semantics interface and a well-defined prosody-syntax interface that can deal with gradient information.

Recommendation 5. Rather than worrying about imagined threats of current LLMs to the generative syntax enterprise, they should be investigated to determine if or how they could inform formal linguistic theorizing/modeling.

5 Conclusions

I would like to conclude with a final observation. The primary concern that drove the development of Minimalism is that its precursors were getting unnecessarily complex under the assumption that a theory of language should be maximally efficient and therefore compact. Much of Chesi's discussion is given over to this concern. However, as a computational grammar engineer I made the experience that grammar simply is complex and that any design decision in one area impacts other (sometimes unexpected) areas. Keeping your phrase structure rules simple (e.g. doing only Merge) means that you have to complicate your feature and constraints system. Slimming down your features and constraints means that you have to complicate your treatment of phrase structure, etc.

We now know much more about the syntax of languages than we did in the 1950s (e.g. Chomsky 1957). In the early days of formal syntactic inquiry it was undoubtedly important and perhaps even crucial to focus exclusively on the mechanisms of syntactic analysis. However, over half a century later, I find that it is time for generative syntax to embrace the interfaces to the other, messier components of grammar in their own right, rather than trying to account for prosodic (e.g. contrast, stress, prosodic boundaries) and semantic/pragmatic phenomena (e.g. speaker/hearer orientation, etc.) within the syntax and with only the tools and representations of syntactic analysis. That is not elegant Minimalism – that is mistaking the problem to be solved. From this perspective the newest instantiation of the Poverty of the Stimulus argument with respect to LLMs is just a windmill distracting from deeper issues.

Abbreviations

CCG = Combinatory Categorial Grammar; HPSG = Head-Driven Phrase Structure Grammar; LCA = Linear Correspondence Axiom; LFG = Lexical Functional Grammar; LLMs = Large Language Models; NLI = natural language inferencing; NLP = Natural Language Processing; NLU = natural language understanding; TAG = Tree Adjoining Grammar.

Acknowledgements

I would like to thank David Adger, Mary Dalrymple, Tracy Holloway King, Hinrich Schütze, Peter Svenonius, Annie Zaenen and – in particular – Gillian Ramchand, for very many stimulating conversations about syntax, syntactic theory and grammar architecture over the years.

Notes

[1] I could list more, but the point here is to provide examples, not an exhaustive list.
[2] <typo.uni-konstanz.de/rara/category/universals-archive>.
[3] <ling.sprachwiss.uni-konstanz.de/pages/home/butt/main/papers/20years.pdf>.
[4] <lfg-proceedings.org>.
[5] <nlp.stanford.edu/software>.
[6] <clarino.uib.no/iness/xle-web>.
[7] <universaldependencies.org>.
[8] <brucehayes.org/otsoft>.
[9] <wordnet.princeton.edu>.
[10] See <vectors.nlpl.eu/explore/embeddings/en/> for one nice way of exploring distributional semantics in LLMs.

Bibliographical References

See the unified list at the end of this issue.