Adopting Large Language Models as a Theory of Language *Does* Refute Chomsky (But Not Like You Think)

Veno Volenec, University of Zagreb & Concordia University, Canada
Charles Reiss, Concordia University, Canada

This paper is a response to Ambridge & Blything (2024), Piantadosi (2024), and similar recent papers which claim that Large Language Models (LLMs) explain how language works. We provide a series of arguments showing that LLMs are not theories of language at all, and therefore cannot be "better at theoretical linguistics" than theoretical linguistics (Ambridge & Blything 2024: 33). We clarify that the object of study of theoretical linguistics is the human mental (brain-based) language faculty, known as 'linguistic competence' (Chomsky 1965) or 'I-language' (Chomsky 1986), and not 'languages' construed as imitative behaviors, conditioned habits, sociopolitical conventions, texts, corpora, etc. What little can be learned from LLMs about the nature of the language faculty directly corroborates generative linguistics, e.g., that the competence–performance dichotomy and some form of Universal Grammar are indispensable. We show that LLM-driven approaches to the study of 'languages' fall prey to the Platonic and externalist delusions that arise from ignoring the I-language perspective, where the subject matter is "a real object rather than an artificial construct" (Chomsky 1986: 28).

Keywords: *large language models; ChatGPT; theoretical linguistics; generative linguistics; syntax; phonology; language acquisition; universal grammar.*

1 Introduction

A recent paper with a self-acknowledged attention-seeking clickbait title — Ambridge & Blything (2024; henceforth A&B) — bombastically proclaims that "theoretical linguistics is dead" (p. 45). The alleged cause of its death are Large Language Models (LLMs) such as GPT-4o. LLMs are claimed to offer a better scientific theory of language "by a country mile", while "traditional linguistic theories don't come close" (p. 45). To prove this point, A&B examine how LLMs perform on acceptability judgement tasks that focus on "the only domain that [they] happen to know quite a bit about: verbs' argument structure privileges" (p. 44). Because LLMs perform as well as humans on this task, they conclude that LLMs explain language. Specifically, A&B take LLMs to be "the best currently available theories of speakers' representation and learning" (p. 34) of linguistic phenomena. The only linguistic approaches that come close to the scientific success of LLMs are the "exemplar-, input- and construction-based" approaches (p. 33). In contrast, "traditional linguistic theories" (i.e., any theory except such usage-based approaches) are inferior because they "are not specified at anything close to the level of detail that would be required for them to make precise *quantitative predictions regarding the relative grammatical acceptability* of individual sentences" (p. 35; emphasis added).

Likewise, Piantadosi (2024: 353) claims that "modern language models implement genuine theories of language" and informs us that "large language models have attained remarkable success
at discovering grammar without using any of the methods that some in linguistics insisted were necessary for a science of language to progress". Triumphantly, he declares that "[t]he unmatched success of an approach based on probability, internalization of constructions in corpora, gradient methods, and neural networks is, in the end, humiliating for a subfield of linguistics that has spent decades deriding these tools" (p. 384). The humiliated subfield of linguistics in question is, of course, 'Chomskyan' generative linguistics, and Piantadosi's (2024: 353) verdict is that "[m]odern language models refute Chomsky's approach to language".

In this paper, we provide a series of arguments showing that LLMs are not theories of language at all, and therefore cannot be "better at theoretical linguistics" than theoretical linguistics (Ambridge & Blything 2024: 33). We clarify that in theoretical linguistics, the term 'language' corresponds to the *human mental (brain-based) language faculty*, which is also known as 'linguistic competence' (Chomsky 1965) or 'I-language' (Chomsky 1986), and not to any set of behaviors, habits, sociopolitical conventions, texts, corpora, or any other extra-mental phenomena. Contrary to Piantadosi's (2024: 353) unsubstantiated assertion that "large language models have attained remarkable success at discovering grammar", we demonstrate that what little can be learned from LLMs about the nature of *the human language faculty* actually reinforces the claim of generative linguistics that the competence–performance dichotomy and some form of Universal Grammar are indispensable. Finally, we show that LLM-driven approaches to the study of 'languages' in the extra-mental sense fall prey to the Platonic and externalist delusions that arise from ignoring the I-language perspective, where the subject matter is "a real object rather than an artificial construct" (Chomsky 1986: 28).

2 The object of study in theoretical linguistics: the human language faculty

The first and foremost problem with Ambridge & Blything is that they show no indication of understanding what "theoretical linguistics" of the sort they're criticizing is supposed to explain. Indeed, the image that comes to mind after reading that paper is that of two hunters standing over a meerkat they have just shot, proudly declaring that they have slain a notorious mighty lion. Hunters should at least know their prey. Theoretical linguistics *does not* set itself the task of explaining how people (let alone machines) perform on acceptability judgement tasks. So, the fact that, say, generative linguistics doesn't "make precise quantitative predictions regarding the relative grammatical acceptability of individual sentences" (Ambridge & Blything 2024: 35) is a good thing. We don't want it to.[1] We want it to capture grammaticality, not acceptability — a crucial distinction.[2] What theoretical linguistics is

[1] We *do not* claim that quantitative research methods are not legitimate within generative linguistics, but that the theory itself shouldn't make quantitative predictions about the likelihood of some particular linguistic *behaviors*. We also *do not* claim that generative linguistics is the only representative of "theoretical linguistics". We think that the arguments we make in this paper are *generally* compatible with other linguistic theories such as role and reference grammar, dependency grammar, etc., even if these theories would diverge from some of our particular claims.

[2] "The general conclusion is that in dealing with the grammar of natural language, one must distinguish acceptability from grammaticality. The latter is a theoretical notion relating to the actual rules characterizing the mental representation of linguistic entities. The former is a far less theoretically loaded concept. Significantly,

supposed to explain, however, is how the human language faculty works. Humans use their language faculty, of course, on acceptability judgement tasks, but at the same time we use other cognitive, sensory and motor systems on such tasks, and these other factors that contribute to how a person judges a sentence's acceptability are obviously not under the purview of theoretical linguistics. In other words, theoretical linguistics is about language knowledge (its content and acquisition), and not about the use of that knowledge on acceptability-measuring tasks. If this looks like we are invoking the notorious (cf. Evans 2006: 108) distinction between *linguistic competence* and *linguistic performance*, that is because we are. It is precisely because of their failure to understand this distinction that Ambridge & Blything's characterization of LLMs as a theory of language doesn't hold. The reasons for the necessity and validity of the competence–performance dichotomy have been stated *ad nauseam* by now (e.g., Chomsky 1965; 1980; 1986; Boeckx 2010; Isac & Reiss 2013; Smith & Allott 2016; Volenec & Reiss 2020; Firestone 2020; Dupre 2021) and are obvious given even the most basic observations of human behavior, so we won't repeat them here. We simply point out that an *immeasurably great array of intractable factors* influences linguistic performance, including performance on acceptability judgement tasks. Performance on such tasks is as arbitrary and variable as, say, a poet with a playful attitude toward language use performing much differently than a literal-minded person. Evidence from linguistic fieldwork clearly shows that even the same speaker sometimes gives different judgements for the same sentence on different occasions (see Vaux & Cooper 2003: 116ff for details and examples). For this reason alone, one cannot draw *direct* conclusions about mental grammar based on acceptability judgements tasks.

Ambridge & Blything are victims here of an error common among non-specialists attempting to weigh in on linguistic issues. They fail to appreciate the distinction between the *object of inquiry* (the human language faculty) and *sources of evidence* (acceptability judgements, eye-tracking studies, brain imaging results, and countless others). Chemists don't study litmus paper, but rather use litmus paper to determine if a substance is an acid or a base; and physicists don't study cyclotrons, but they use them to study fundamental particles. Similarly, linguists do not study acceptability judgements, but rather use them to study the human language faculty (see Volenec & Reiss 2020: 7ff for further discussion on the confusion between the object of study and sources of evidence). Of course, Ambridge & Blything are free to define linguistics in a manner that is completely at odds with the definition used by generative linguists, but then it is an error to presume that conclusions reached using their definition should be relevant to the object of study defined by theoretical linguists. Ambridge & Blything's use of the term 'linguistics' is an example of what logicians call the fallacy of equivocation (Hansen 2024) – the use of a key word in an ambiguous manner. Ambridge & Blything rely on their idiosyncratic definition of the field to draw conclusions about the standard definition used by theoretical linguists.

A *productive* approach in linguistics is to use acceptability judgements as a source of evidence in the painstaking and indirect process of inferring the properties of linguistic competence, carefully trying to control the variables and to disentangle the various contributing factors.[3] An *unproductive* approach is to use acceptability judgements to argue that linguistic

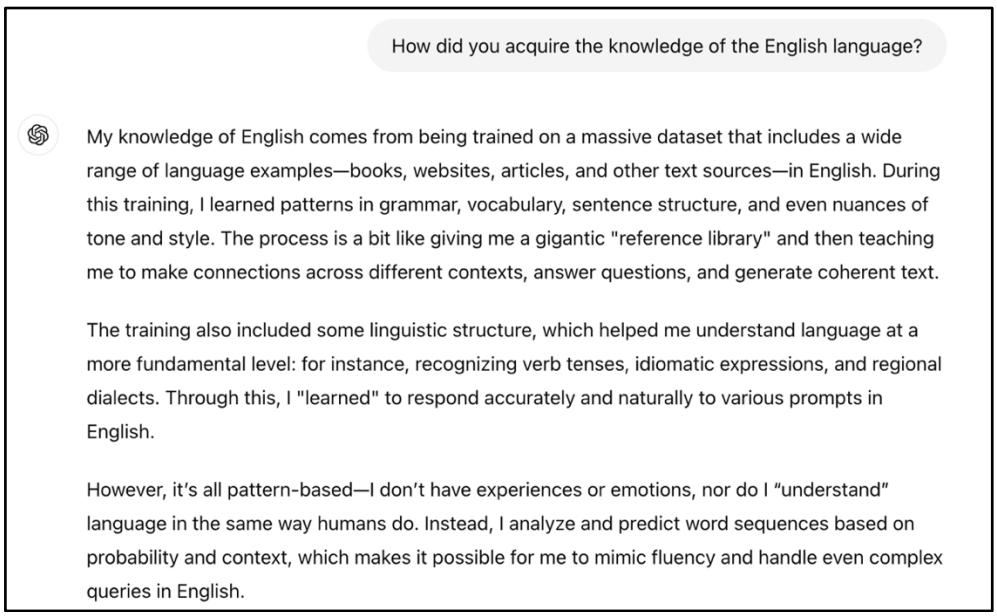
sentences that are not grammatical may still be acceptable, though their acceptability will be due to *extragrammatical* factors such as processing." (Hornstein 1984: 35; emphasis in the original)

[3] "In practice, we tend to operate on the assumption, or pretense, that these informant judgments give us 'direct evidence' as to the structure of the I-language, but, of course, this is only a tentative and inexact working hypothesis, and any skilled practitioner has at his or her disposal an armory of techniques to help compensate for the errors introduced. In general, informant judgements do not reflect the structure of the language directly;

competence doesn't exist, that mental grammar (as opposed to behavior) is probabilistic, and that theoretical linguistics doesn't work. A&B's inflammatory statement that "large language models are better than theoretical linguists at theoretical linguistics" (p. 33) reveals a profound ignorance that explains the logical fallacy of equivocation to which they succumb: they don't know what the object of inquiry of theoretical linguistics is, and therefore they don't know what they're criticizing. (Whence the hunters' delusion: they can't tell a meerkat from a lion.)

3 'Language acquisition': human brains vs. LLMs

The second reason why Ambridge & Blything are wrong in claiming that LLMs are better at explaining language than theoretical linguistics is that LLMs and human brains (where the actual object of study of linguistics is located) have nothing relevant in common (loose metaphors about 'neural networks' notwithstanding), especially in the context of how knowledge of language is acquired (Friederici 2017). Ironically, a very good argument for this comes from an LLM itself. Here is how ChatGPT-4o described its own 'language acquisition' (on October 28, 2024):



"Massive data-set", "gigantic reference library", "training", "teaching [it] to make connections", "[it does not] understand language in the same way humans do", "[it] mimics fluency". ChatGPT even put scare quotes around 'learned'! It is clear that this is *nothing* like what goes on in the brain of a human infant. The human brain doesn't have access to a "gigantic reference library" in its infancy, and even if it did, it wouldn't be able to make use of it (Piattelli-Palmarini & Berwick 2013; Bever et al. 2023). Notice also that LLMs acquire the ability to mimic human language fluency solely through reading, while no child has ever acquired language through reading (solely or even largely). Thus, LLMs' path toward the ability to

judgements of acceptability, for example, may fail to provide direct evidence as to grammatical status because of the intrusion of numerous other factors." (Chomsky 1986: 36)

produce human-like expressions is completely different from a child's path. A&B even describe how they optimized an LLM to perform better on acceptability judgement tasks: "Following a training session designed to familiarize the model with the task of rating sentences on a 5-point acceptability scale [...] the model is given (counterbalanced) prompts" (p. 38). What does that have to do with human language? Do A&B seriously think that babies undergo training sessions to acquire language? In line with A&B's reference to Popper's criterion of falsifiability (p. 34), even one baby that acquired language without explicit training is sufficient to refute A&B's 'theory' of language acquisition. In fact, every baby that ever existed refutes it. Humans engineered LLMs to be as human as possible in their overt communicative behavior; it is precisely in achieving that goal that LLMs have revealed how different they are because *their path to achieving it was profoundly inhuman*.

4 Universal Grammar and domain-specificity

Ambridge & Blything did, however, get one important point right. They've admitted that even LLMs require *a priori*, *predetermined*, *in-built* principles and mechanisms that make the language learning process possible upon exposure to data. In other words, they agree that even LLMs, which were trained on corpora containing "hundreds of billions of words" according to ChatGPT's estimate, cannot function without some form of Universal Grammar. In A&B's own words:

> "We" – or at least the software engineers who built LLMs – made hundreds of decisions about the precise architecture and learning mechanisms that should be used. These engineers could have made different choices; and – depending on those choices – the models would have simulated human acceptability judgments either better or worse. These choices – fossilized in thousands of lines of computer code – are a theory of human language acquisition.
>
> (Ambridge & Blything 2024: 42)

It is gratifying to see they've finally reached the conclusion that UG is inevitable after the misguided attempt to explain "why universal grammar doesn't help" in child language acquisition (Ambridge et al. 2014).[4]

Only by virtue of having a predetermined universal set of basic units and operations can the categorization and representation of the infinitely variable external linguistic stimuli proceed, and only by assuming (and then further investigating and refining) such units and operations can the science of language succeed.[5] This point was made even more broadly by Hammarberg (1981), who argued that it applies to all aspects of cognition and scientific method:

> [N]o IPS [information processing system, such as mental grammar] could have access to matters not representable in the 'language' in terms of which the IPS functions. Matters not representable are not accessible, and matters accessible are so only in

[4] Of course, the need to build 'priors' into machine learning systems is a truism, whether engineers explicitly acknowledge it or not (Versace et al. 2018; Rawski & Heinz 2019; Wolpert 2021). See Volenec & Reiss (2020) and Reiss & Volenec (2022) for a more detailed discussion.

[5] For an elaboration on how the unavoidable assumption of universality pertains to research methods in linguistics, see Reiss (2025, forthcoming).

> virtue of being presented in the 'language' of the IPS. Thus from the point of view of any IPS, *its* data are going to appear ultimate to *it* [...], simply because it cannot 'see things' in any other way. The fact of these matters seems to be that an IPS [...] is a prisoner of its own representational processes: We can never escape a point of view. [W]e are paradigm-bound and not only in doing science, but in all our cognitive-perceptual activities.
>
> (Hammarberg 1981: 262–263).

True blank slates, like rocks, cannot learn. Crucially, *one cannot learn that which is required for learning*, an obvious logical truism recognized by Plato, Kant, Fodor, and many others. Therefore, for any learning to take place, there must be some *a priori*, predetermined, in-built learning mechanisms. This is true of anything that learns, be it biological or artificial, and it is true in any domain of learning, linguistic or otherwise. In the domain of language, for historical reasons (Chomsky 1966), these innate mechanisms happen to have a name: Universal Grammar. In virtually all other areas of (neuro)biology, they are taken for granted and thus remain unnamed (e.g., Gazzaniga et al. 2019): no one would claim that the visual faculty, the auditory faculty, or the object recognition faculty are *not* partly determined by innate factors.

Piantadosi (2024: 390) also agrees that "[t]here are no doubt *some* [in-built] principles required for language." The question, according to him, is whether they are specific to the language faculty or domain-general. But ascribing only domain-general language-learning mechanisms (DLMs) to humans leads to insurmountable empirical and theoretical problems. Empirically, given their generality, DLMs predict that the output of language acquisition (i.e., mature I-languages) will vary in ways that are never actually attested. A simple example illustrates this point. All I-languages contain phonological segments (mental representations of consonants and vowels) that are built from a small set of universal distinctive features (e.g., [±CONTINUANT], [±SYLLABIC], [±HIGH]). Could these distinctive features be learned via DLMs? Presumably, the relevant DLM here would be general audition: the brain's ability to differentiate between different frequencies, intensities and durations of sound, and to extract categories from that acoustic information (see Mielke 2008 for an attempt to develop such an approach to feature learning). For concreteness, consider the feature [±HIGH], which in acoustics correlates with the frequency of the first formant (F1). In various experiments, the human auditory system, the putatively relevant DLM here, was reported to be able to discriminate between at least 1300 levels on a single frequency scale (Fastl & Zwicker 2013). But this capacity has nothing to do with phonological competence: we do not have 1300 different levels for vowel height. There are at most five contrastive levels (Gussenhoven & Jacobs 2017: 81), and languages usually manifest only two or three. Note that the *observed* phonetic variability of vowels along the F1 dimension is so vast (Hillenbrand et al. 1995: 3104) that we actually *could* have hundreds of vowel height levels. But we don't, and this is not a matter of statistical probability: we *always* have only a few phonologically relevant levels. More generally, phonology has a handful of discrete categories, despite the fine-grained perceptual sensitivity that the human auditory system furnishes.

This empirical failure of DLMs extends to all levels of linguistic organization: they fail to explain *why*, for example, all languages have hierarchical syntactic structures, a small set of phrase types (NPs, VPs, PPs, etc.), discrete lexical items (morphemes), verbs with argument privileges, a limited set of phonological features, consonants and vowels, ordered phonological processes, a universal syllable structure, and so on. There is objectively an infinity of alternative ways to parse the primary linguistic data (PLD; the linguistic data available to a language-acquirer) using DLMs, but the process of language acquisition always settles on the

same universal categories. These highly specific categories, like negative polarity items, anaphoric pronouns, and *wh*-elements, recur in language after language, and can serve no conceivable function in any other, non-linguistic domain of cognition, such as visual interpretation, decision-making, value judgement, facial perception, intentionality, attention, etc. Also, the listed linguistic universals[6] could not have arisen because of communicative needs since they do not optimize communication in any way,[7] so functionalist or usage-based 'explanations' are of no use.

The theoretical problem with DLMs is even worse than the empirical one. Piantadosi (2024: 368) claims that LLMs make syntax "reducible to general statistics between words". But showing that supercomputer-driven AI can find statistical patterns in gigantic corpora of text and can then reproduce those patterns to chat with humans has no scientific value for linguistics for at least two reasons. The first one has already been mentioned in the previous section: that is not even remotely similar to how human brains work, and it has nothing to do with how children acquire language. The second reason is that it fails to explain *where these patterns come from and why they are the way they are*. The patterns are in the PLD (for humans) or in corpora (for machines) because human minds contain mechanisms that produced those patterns; human minds have acquired those mechanisms, according to DLM models, by statistically analyzing *their* PLD; *that* PLD comes from older minds, whose mechanisms come from even older PLD, and so on indefinitely. Obviously, that's not an explanation, but a mere fallacy of infinite regress. Domain-general statistics doesn't (cannot) explain why the human language faculty is the way it is; consequently, LLMs don't either. Only by pursuing a biolinguistically plausible theory of Universal Grammar, formulated in terms of innate units and operations specific to the language faculty,[8] can we hope to explain why the language faculty is the way it is, how it develops in individuals, and how it might have evolved in our species.

5 Implications of theoretical linguistics for syntax

Let us move on now to the syntactic phenomenon discussed in Ambridge & Blything's paper, namely the "domain that [they] know something about: learning and representing verbs' argument structure privilege" (p. 33). Unfortunately, they do not seem to know much about that domain either, because their identification of the domain is sloppy and self-contradictory. The problem again derives from a lack of understanding of the object of study of theoretical

[6] By *universals* we mean the innate options provided by UG which are likely to manifest in most I-languages. A common misunderstanding is expecting that *all* the options provided by UG need to be overtly present in *all* I-languages. They do not, just like all of the phenotypes made possible by the human genome are not overtly present in every individual. Importantly, we are not referring to the typological universals (descriptive generalizations) of the kind Greenberg (1963) discussed. For an insightful discussion on these different notions of universals, see Baker (2011).

[7] And in many ways they impede communication. Opaque rule interactions in phonology, phonological neutralization, structural ambiguity in both syntax and morphology, syntactic displacement, barriers to displacement (islands), homophony, etc., make communication using language *harder*, not easier.

[8] Such a research program might be complemented by so-called 'third factor' principles that shape language but are not part of UG; rather, they are laws of nature, such as computational efficiency (Chomsky 2005; 2025, in press). So, all in all, this framework sees the growth of I-languages inside of human brains as arising from an interaction of three factors: innate domain-specific units and operations provided by the human genome (UG), external experience that serves as the primary linguistic data (PLD), and general laws of nature such as simplicity and efficiency (the third factor).

linguistics — the human language faculty and individual I-languages. Linguists often talk informally about 'languages' like English, and in that context, it is perfectly reasonable to talk about **the** verb *roll* in the sentences *The ball rolled* and *Someone rolled the ball* appearing with two different sets of arguments. But linguists also sometimes consider that the two verb forms might contain unpronounced differences, differences that are sometimes reflected in the pronunciation of such intransitive-transitive pairs as *lie / lay*. Deciding whether the two printed words spelled *rolled* are actually type-identical is a legitimate question discussed by linguists (e.g., Borer 1994, 2004; Levin & Rappaport Hovav 2004),[9] but we are concerned here with a much simpler issue.

First, let's agree that a verb, or any morpheme, must be understood as having at least a tripartite structure. There must be a lexical meaning, expressing things like the distinction between, say, *roll, spin* and *cry.* There must also be a phonological representation, relating to things like the difference in pronunciation between *flee, fly* and *flow*. Finally, there must be a syntactic representation that indicates that a word is a verb, and perhaps, depending on one's theory, what its argument structure is, what arguments it takes. If our two tokens of *rolled* do not encode a difference in argument structure, there is a chance that they are type-identical. However, if they do encode such a difference, then by definition, they cannot be tokens of the same word. As we said, this is a legitimate issue that morphologists and syntacticians can discuss.

With these preliminaries in mind, consider the following sequences of 'English' words:

(1) a. We are not allowed to run here.
b. We are not allowed running here.

(2) a. We are done with our homework.
b. We are done our homework.

Most speakers of English say that the (a) sequences are acceptable, and that is what ChatGPT-4o tells us, too. And most people, like ChatGPT-4o, will tell us that the strings in (b) are not acceptable, they are not sentences of English.

However, English speakers in Montreal, and many other speakers of 'Canadian English' judge the (b) examples as perfectly acceptable (Fruehwald & Myler 2015). To simplify the discussion, let's assume that the Canadian speakers use only the (b) forms and reject the (a) sentences. (They do not reject them, probably because they are regularly exposed to more than one dialect and so treat the (a) forms as *acceptable* — they are typically not aware that the (b) forms are not used by other English speakers.) For (1a) we might say that 'the past participle verb *allowed* occurs with an argument structure that has an infinitival complement,' whereas for (1b) we might say that 'the past participle verb *allowed* occurs with an argument structure that has a gerund complement'.[10] And in (2a) we might say that 'the verb *done* takes

[9] LLMs don't even recognize that such phenomena are puzzles that need to be explained, let alone provide an explanation for them. This is yet another reason why LLMs cannot be "better than theoretical linguists at theoretical linguistics" — they neither raise nor answer important research questions.

[10] The argument structure of a verb uses primitives such as 'external argument' and 'internal argument'. Arguments are typically nominal, but they could also be clausal. Whether an internal argument is nominal or clausal is captured in the subcategorization feature of that verb, not in the argument structure. Even if we assume a laxer version, in which the argument structure of a verb does specify whether that argument is clausal or not, the

a prepositional phrase complement', whereas in (2b) 'the verb *done* takes a noun phrase complement'. However, in both cases, we would be talking nonsense. Unlike the case of *rolled* above, there is not even a remote possibility that we are talking about **the** past participle of **the** verb *allow* or **the** verb *do.*

In each case, a Canadian speaker has a morpheme that has a similar or identical semantic component as a morpheme of an American speaker, and the two have similar or identical phonological representations for the corresponding morphemes. However, the difference in argument structure in each case must be distinctly encoded. Canadian *allow* has different properties from American *allow*, so by Leibniz's Law concerning the Identity of Indiscernibles (Forrest 2025), the two must be different.

Similarly, within a single speaker's mental lexicon, *couch* and *sofa* share a common syntax and semantics, but they cannot correspond to the same morpheme because they differ phonologically. By the same reasoning, the word spelled *dog* for the two present authors cannot be tokens of the same morpheme across our lexicons, because for one of us, *dog* rhymes with *morgue* but not with *log* or *frog*, as it does for the other: they differ phonologically. If *any* aspects of our morphemes differ, they cannot be the same. We need to distinguish everyday talk that we all engage in — where we might say that English *mutton* and French *mouton*, or English *hound* and German *Hund* are 'the same word' — from scientific discourse in which 'language' is a technical term that designates a particular component of the human mind/brain. Despite everyday informal talk, only a 'Chomskyan' I-language approach, which sees 'language' as mind-**i**nternal, **i**ndividual, **i**ntensional unconscious knowledge, can handle such facts. LLMs are fed *written* sources that obscure pronunciation differences as well as many other differences like the Canadian vs. 'Standard English' examples above. Human learners acquire lexicons and grammars that reflect the data to which they are exposed. An LLM might assign a string like *We are not allowed running here* a very low probability of occurrence, but that does not reflect what is actually going on.

6 The status of I-language in theoretical linguistics

A&B fail to understand the reasons for the necessity of adopting the I-language perspective in order to do linguistic science and the reasons for rejecting the Platonic conception of language, also known as 'P-language'. The difference between the two is explained by Chomsky (1986: 33–34):

> Sometimes it has been suggested that knowledge of language should be understood on the analogy of knowledge of arithmetic, arithmetic being taken to be an abstract "Platonic" entity that exists apart from any mental structures [...] What is claimed is that apart from particular I-languages, there is something else additional, what we might call "P-languages" (P-English, P-Japanese, etc.), existing in a Platonic heaven alongside of arithmetic and (perhaps) set theory, and that a person who we say knows English may not, in fact, have complete knowledge of P-English [...].
>
> In the case of arithmetic, there is at least a certain initial plausibility to a Platonistic view insofar as the truths of arithmetic are what they are, independent of any facts of

argument structure will certainly *not* specify morpho-syntactic differences between various types of nominals or clauses, i.e., whether an argument is a definite or indefinite nominal constituent, or whether an argument is a finite or non-finite clause, or what type of non-finite clause it is. This is captured in the selectional or subcategorization feature of that verb. See, e.g., Hornstein (2009) and Carnie (2021) for more details.

> individual psychology, and we seem to discover these truths somewhat in the way that we discover facts about the physical world. In the case of language, however, the corresponding position is wholly without merit. There is no initial plausibility to the idea that apart from the truths of grammar concerning the I-language [...] there is an additional domain of fact about P-language, independent of any psychological states of individuals. [...] Of course, one can construct abstract entities at will, and we can decide to call some of them "English" or "Japanese" and to define "linguistics" as the study of these abstract objects, and thus not part of the natural sciences, which are concerned with such entities as [I-languages and UG]. But there seems little point to such moves.

And yet, precisely "such moves" are being made by those who believe that LLMs explain 'language'. Ambridge & Blything tacitly adopt the 'P-language' conception which includes a mystical pseudo-Platonic idea that an 'English' verb like *allow* exists outside of human minds. Under A&B's assumptions, we would conclude that Canadians and Americans have different partial knowledge of the argument structure privileges of the P-English verb *allow.*

Moreover, Ambridge & Blything's perspective also reflects the 'E-language' conception, another scientifically untenable notion (Isac & Reiss 2013: §3–4) in which a language might be thought of as an externalized collection of actions or utterances, perhaps recorded on paper or in computer files. Since strings of text on paper or in computer files are related to the morphemes in human minds in an exceedingly indirect manner, they cannot be taken as direct evidence about how people learn and represent argument structure privileges. So, all in all, Ambridge & Blything can't be "better than theoretical linguists" (p. 33) at analyzing language because they don't know what language is or even what a verb is for a linguist.

In addition to tacitly adopting a mix of P-language and E-language conceptions, Ambridge & Blything's paper also fails to avoid all of the related problems of adopting the everyday sociopolitical notion of language, one related to identity, history, tradition, geography and politics (Chomsky 1986: 15–16; Lightfoot 1999: §3; Hale 2007: §1–2; Boeckx 2010: §1).[11] Ambridge & Blything, and virtually all work that relies on corpora, fall prey to these same well-known problems. How do Ambridge & Blything decide what counts as English? Should we include in an 'English' corpus Trump's speeches, Toni Morrison's novels, Shakespeare's sonnets, Chaucer's *Canterbury Tales, Beowulf*, and so on? The findings of theoretical linguistics are not affected by the outcome of the wars that broke up Yugoslavia — we never believed that an entity called 'Serbo-Croatian' existed in the world, but rather, we believed in a bunch of more or less similar I-languages. LLM researchers have to make a linguistically arbitrary decision on whether to train a model on data that is called, say, 'Croatian', or additionally on data that is called 'Serbian'. (In practice, such decisions are usually made on the basis of what is deemed politically correct and financially more beneficial.) Whatever decision is made will generate a set of probabilities that has no bearing on how the mental grammars of the individual speakers in those regions work. As theoretical linguistics has made

[11] "[T]he commonsense notion of language has a crucial sociopolitical dimension. We speak of Chinese as "a language," although the various "Chinese dialects" are as diverse as the several Romance languages. We speak of Dutch and German as two separate languages, although some dialects of German are very close to dialects that we call "Dutch" and are not mutually intelligible with others that we call "German." A standard remark in introductory linguistics courses is that a language is a dialect with an army and a navy (attributed to Max Weinreich). That any coherent account can be given of "language" in this sense is doubtful; surely, none has been offered or even seriously attempted. Rather, all scientific approaches have simply abandoned these elements of what is called "language" in common usage." (Chomsky 1986: 15)

clear, the shift of perspective to I-language as the object of inquiry in linguistics is a shift towards realism, "the study of a real object rather than an artificial construct" (Chomsky 1986: 28).

7 A note on phonology

Dabbling in phonology to reinforce their point also didn't work well for A&B. They misleadingly cite a paper by one of us: "In phonology, for example, the once-mainstream idea of a universal set of phonological features (analogous to the categories assumed in the domain of verb argument structure) is 'very much a minority position today, even among phonologists trained in the generative tradition' (Reiss 2023: 9)" (p. 44). In fact, the passage from Reiss (2023) continues thus: "For example, Reiss & Volenec (2022) is the sole contribution to a recent volume on phonological primes that adopts and defends the nativist position for features expressed by Chomsky and Halle." Aside from clarifying the point that a minority position is not necessarily the wrong position (if it were, everyone with a new idea would necessarily be wrong!), we argued that the *only* coherent way in which two linguistic forms can be considered as type-identical (phonologically, semantically, and/or syntactically) is in terms of a universal, innate set of representational primitives (e.g., phonological features). No one who does phonology invents features from scratch for every language or for every analysis. That would be as senseless as inventing a new periodic table or a new set of subatomic particles for every new chemical analysis.[12] Indeed, it has been known for a long time that "all phonology breaks down if we do not assume analysis [...] in terms of universal phonetic features [which is an old name for *phonological* features — V&R]" (Chomsky & Halle 1965: 119). Even saying something as trivial as "English, Spanish, and Japanese all contain the phoneme /p/", only makes sense if features are universal. A&B's reliance on spelling and arbitrarily delimited corpora is a non-starter for scientific inquiry.

Needless to say, LLMs also didn't learn the primitive elements of their 'phonological' representations. In this case, since the LLMs under discussion only deal with text, these 'features' correspond to the characters of ASCII and Unicode that LLMs use to encode and process text. And these representational primitives were of course built into the LLMs by the engineers who made them. By ChatGPT's own admission (on May 7, 2025): "I didn't learn Unicode standards like 'é' being 'U+00E9'; they were built into my architecture". As we argued in section 4, the existence of in-built domain-specific primitives is inevitable in any system that has the capacity to learn.

8 Conclusion

In the generative linguistics literature, it is a commonplace that all entities, such as particular verbs, syntactic categories, phonological segments, syllables, etc., are mental constructs of individual I-languages. Even in the more 'concrete' domains of phonetics and phonology, we find statements such as: "it should be perfectly obvious by now that segments do not exist

[12] The analogy references Jackendoff's (1994: 60) characterization that "the discovery of distinctive features, and the continual refinement of their formulation over some decades, [is] a scientific achievement on the order of the discovery and verification of the periodic table in chemistry".

outside the human mind" (Hammarberg 1976: 355). However, this view does not mean that segments and other linguistic entities are not real:

> Science aims for a theory of the real, and to base one's descriptions and generalizations on a fictional taxonomy could only lead to one's theories being fictional as well.
>
> Current phonological theory holds that segments are the entities in terms of which the phonological component of a grammar operates, and the grammar, in turn, is a cognitive mechanism.
>
> (Hammarberg 1976: 355).

In other words, phonological segments are not "convenient fictions" (Laver 1994: 568) or 'merely mental entities' in the sense that unicorns are; they are real mental entities in the sense that they are the inputs and outputs of mental computations, somehow encoded in the brain (Idsardi & Monahan 2016). This view extends to all levels of linguistic analysis: "Language, as far as I can tell, is all [mental] construction" (Jackendoff 1992: 164). And notably, as Chomsky (2015: 126) puts it:

> No one is so deluded as to believe that there is a mind-independent object corresponding to the internal syllable [ba], some construction from motion of molecules perhaps, which is selected when I say [ba] and when you hear it.

It is thus clear to theoretical linguists that the phoneme /l/ of the English verb *allow* does not exist as a mind-independent object; the final syllable of the English verb *allow* does not exist as a mind-independent object; the argument structure privileges of the English verb *allow* do not exist as a mind-independent object; the English verb *allow* does not exist as a mind-independent object; and finally, English does not exist as a mind-independent object. Ambridge and Blything's (2024) critique of theoretical linguistics, which is fundamentally dependent on the existence of mind-independent verbs with argument structure 'in English', gives the lie to Chomsky's claim cited above. To circle back to the title of Piantadosi's (2024) recent viral paper, modern language models do indeed refute at least one of Chomsky's views: such delusions do exist.

Acknowledgements

We are very grateful to Elizabeth Tobyn, Dana Isac, Gabe Dupre, Jon Rawski, Neil Myler, and two anonymous reviewers for their valuable feedback on earlier versions of the paper.

References

Ambridge, Ben & Liam Blything. 2024. Large language models are better than theoretical linguists at theoretical linguistics. *Theoretical Linguistics* 50(1–2). 33–48.

Ambridge, Ben & Julian M. Pine & Elena Lieven. 2014. Child language acquisition: Why universal grammar doesn't help. *Language* 90(3). 53–90.

Baker, Mark. 2011. The interplay between universal grammar, universals, and language specificity. *Linguistic Typology* 15. 473–482.

Bever, Thomas & Noam Chomsky & Sandiway Fong & Massimo Piattelli-Palmarini. 2023. Even deeper problems with neural network models of language. *Behavioral and Brain Sciences* 46. e387.

Boeckx, Cedric. 2009. *Language in Cognition. Uncovering Mental Structures and the Rules Behind Them.* Malden: Wiley-Blackwell.

Borer, Hagit. 1994. The projection of arguments. *University of Massachusetts occasional papers in linguistics* 17(20). 19–48.

Borer, Hagit. 2004. The grammar machine. *The unaccusativity puzzle: Explorations of the syntax-lexicon interface* 5. 288–331.

Carnie, Andrew. 2021. *Syntax: A Generative Introduction*. London: Wiley Blackwell.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chomsky, Noam. 1966. *Cartesian Linguistics.* New York: Harper & Row.

Chomsky, Noam. 1980. *Rules and Representations*. Oxford: Basil Blackwell.

Chomsky, Noam. 1986. *Knowledge of Language. Its Nature, Origins, and Use.* New York: Praeger.

Chomsky, Noam. 2005. Three factors in language design. *Linguistic Inquiry* 36, 1–22.

Chomsky, Noam . 2015. *What kind of creatures are we?*. New York: Columbia University Press.

Chomsky, Noam. 2025, in press. *The Miracle Creed and SMT*. In Giuliano Bocci, Daniele Botteri, Claudia Manetti & Vincenzo Moscati, V. (eds.), *Issues in Comparative Morpho-syntax and Language Acquisition*. Cambridge: Cambridge University Press.

Chomsky, Noam & Morris Halle. 1965. Some Controversial Question in Phonological Theory. *Journal of Linguistics* 1. 97–138.

Dupre, Gabe. 2021. (What) Can deep learning contribute to theoretical linguistics?. *Minds and Machines* 31(4). 617–635.

Evan, Vyvyan & Melanie Green. 2006. *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

Firestone, Chaz. 2020. Performance vs. competence in human–machine comparisons. *Proceedings of the National Academy of Sciences U.S.A.* 117(43). 26562–26571.

Forrest, Peter. 2025. The Identity of Indiscernibles. In Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2025 Edition). URL: [https://plato.stanford.edu/archives/sum2025/entries/identity-indiscernible/].

Fruehwald, Josef & Neil Myler. 2015. I'm done my homework – Case assignment in a stative passive. *Linguistic Variation* 15(2). 141–168.

Gazzaniga, Michael S. & Richard B. Ivry & George R. Mangun. 2019. *Cognitive Neuroscience: The Biology of the Mind.* New York: W. W. Norton & Company.

Greenberg, Joseph. 1963. Some universals of grammar, with particular reference to the order of meaningful elements. In Joseph Greenberg(ed.), *Universals of Language*, 58–90. Cambridge: MIT Press.

Hale, Mark. 2007. *Historical Linguistics: Theory and Method*. Oxford: Blackwell.

Hammarberg, Robert. 1976. The metaphysics of coarticulation. *Journal of Phonetics* 4. 353–363.

Hammarberg, Robert. 1981. The cooked and the raw. *Journal of Information Science* 3(6). 261–267.

Hansen, Hans. 2024. Fallacies. In Ed Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2024 Edition). URL = [https://plato.stanford.edu/archives/fall2024/entries/fallacies/].

Hillenbrand, James M. & Laura A. Getty & Michael J. Clark & Kimberlee Wheeler. 1995. Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America* 97(5). 3099–3111.

Hornstein, Norbert. 1984. *Logic as Grammar.* Cambridge: MIT Press.

Hornstein, Norbert. 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge: Cambridge University Press.

Idsardi, William J. & Philip Monahan. 2016. Phonology. In Gregory Hickok & Steven L. Small (eds.), *Neurobiology of Language*, 141–151. London: Elsevier.

Isac, Daniela & Charles Reiss. 2013. *I-Language. An Introduction to Linguistics as Cognitive Science*. Second edition. Oxford: Oxford University Press.

Jackendoff, Ray. 1992. *Semantic Structures.* Cambridge: MIT Press.

Jackendoff, Ray. 1994. *Patterns in the Mind: Language and Human Nature.* New York: BasicBooks.

Laver, John. 1994. *Principles of Phonetics.* Cambridge: Cambridge University Press.

Levin, Beth & Malka Rappaport Hovav. 2005. *Argument Realization.* Cambridge: Cambridge University Press.

Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution*. London: Wiley Blackwell.

Mielke, Jeff. 2008. *The Emergence of Distinctive Features*. Oxford: Oxford University Press.

Piantadosi, Steven T. 2024. Modern language models refute Chomsky's approach to language. In Edward Gibson & Moshe Poliak (eds.), *From fieldwork to linguistic theory: A tribute to Dan Everett*, 353–414. Berlin: Language Science Press.

Piattelli-Palmarini, Massimo & R.C. Berwick (eds.) 2013. *Rich Languages from Poor Inputs.* Oxford: Oxford University Press.

Rawski, Jonathan & Jeffrey Heinz. 2019. No free lunch in linguistics or machine learning: Response to Pater. *Language* 95(1). 125–135.

Reiss, Charles. 2023. Research methods in Armchair linguistics. Lingbuzz Preprint: [https://lingbuzz.net/lingbuzz/007568].

Reiss, Charles. 2025, forthcoming. Research methods in Armchair linguistics. In Gabe Dupre et al. (eds.), *The Oxford Handbook of Philosophy of Linguistics*. Oxford: Oxford University Press.

Reiss, Charles & Volenec, Veno 2022. Conquer primal fear: Phonological features are innate and substance-free. *Canadian Journal of Linguistics* 67(4). 581–610.

Smith, Neil & Nicholas Allott. 2016. *Chomsky: Ideas and Ideals*. Cambridge: Cambridge University Press.

Vaux, Bert & Justin Cooper. 2003. *Linguistic Field Methods*. Eugene: Wipf and Stock Publishers.

Volenec, Veno & Charles Reiss. 2020. Formal Generative Phonology. *Radical: A Journal of Phonology* 2. 1–65.

Versace, Elisabetta & Antone Martinho-Truswell & Alex Kacelnik & Giorgio Vallortigara. 2018. Priors in animal and artificial intelligence: where does learning begin? *Trends in cognitive sciences* 22/11. 963–965.

Wolpert, David H. 2021. What is important about the no free lunch theorems?. In *Black box optimization, machine learning, and no-free lunch theorems*, 373–388. Springer International Publishing.

Veno Volenec
Department of Classics, Modern Languages and Linguistics
Concordia University
1250 Guy Street (FB1000.05)
Montreal, QC H3G 1M8
Canada
E-mail: vvolenec@m.ffzg.hr

Charles Reiss
Department of Classics, Modern Languages and Linguistics
Concordia University
1250 Guy Street (FB1000.05)
Montreal, QC H3G 1M8
Canada
E-mail: charles.reiss@concordia.ca

In SKASE Journal of Theoretical Linguistics [online]. 2025, vol. 22, no. 1 [cit. 2025-06-30]. Available on web page http://www.skase.sk/Volumes/JTL58/01.pdf. ISSN 1336-782X