DOI: 10.26346/1120-2726-247

# Generative Interactions: No change no gain

Martina Wiltschko

ICREA, Universitat Pompeu Fabra, Barcelona, Spain <martina.wiltschko@icrea.cat>

Chesi argues that generative linguists deserve Piantadosi's dismissal of the Chomskyan approach because of what he considers the generativists original sin: being content with incomplete pseudo-formalizations and data fragment explanations. This criticism is based on a narrow perspective on generativism. Here I outline my own (broader) perspective and the questions I ask, all inspired by core generative ideas. The discussion evolves around three areas: (i) language variation, (ii) interactional language, and (iii) language acquisition. I argue that very large language models are not suitable to replace the theoretical assumptions I hold to explore these areas of research. Some of the assumptions I hold are not consistent with a narrow perspective on generativism *a la* Chesi, and one might conclude that this reflects his view that it is "the end of generativism as we know it". However, I invite a different conclusion: they simply reflect change. And just as change in language is a sign that the language is alive and spoken, so is change in a theoretical framework a sign that it is alive and used.

KEYWORDS: linguistic variation, interactional language, language acquisition, cognitive revolution.

## *1. Introduction*

By way of introduction, let me start with three quotes from Chesi's target article, as they represent the claims that I will take issue with in this response.

(1) Three quotes
   a. In the end, on the one side of the field, computational linguists depend on statistical predictions obtained from vast corpora and have shown that the core syntactic engine, PF, and LF, are effectively distinct only within the theoretical 'T-model'.
   b. To truly understand what a sentence means – crucial for tasks like machine translation or answering questions – it is essential to rely on robust machine learning methods, which are more solid than any formal theory on the market.
   c. On the other side of the field, experimental linguistics has refined its methods, significantly improving the observational capabilities and, ultimately, enhancing the analysis of nearly all sources of linguistic data, whether implicit or explicit, categorical or gradual.
   (Chesi *this issue*: 38-39)

The criticism of generativism expressed in these quotes boils down to three aspects summarized in (2):

(2) a. MODELLING: the T-model is wrong-headed
b. FORMALIZATION: machine learning methods (and hence vLLMs) are superior
c. METHODOLOGY: data sets are more complete when obtained experimentally

In what follows I address these three aspects of Chesi's criticism, and I argue that they are ill-founded, using examples based on my own work to make my point. Hence what follows is a personal perspective on the generative enterprise. I proceed as follows. In section 2, I start by reviewing what I take to be the core tenets of generative linguistics. This will set the stage for the following discussion which is based on the premise that generative linguistics is not confined to the narrow view Chesi portraits. In other words, not every generativist is also a minimalist. In sections 3-5, I present core aspects of my own work that exemplifies this broader stance. As I will show, the assumptions I adopt, and which are decidedly generative, have led me to a series of questions and observations, which arguably would not have been possible if we were to abandon the essence of the T-model and if vLLMs were all there is to linguistic theory. In section 3, I explore linguistic variation in the inventory of grammatical categories. In section 4, I discuss the grammar of interactional language. In section 5, I discuss some relevant aspects of child language acquisition. In section 6, I conclude.

## *2. What is generative linguistics?*

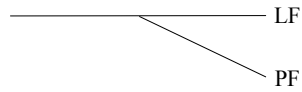
According to Chesi, it is the end of generative linguistics as we know it and that this is our own fault. To evaluate this claim it is essential that we agree on what we are talking about when we talk about generative linguistics. In this section, I provide a brief overview of what I take to be the essence of generative linguistics.

One of the core tenets of generative linguistics is its cognitive stance. That is, Chomsky's initial ideas were developed in response to structuralist linguistics and in opposition to behaviorist ideas about language development. In *Syntactic Structures*, Chomsky (1957) breaks with purely structuralist assumptions by postulating abstract syntactic representations and rules meant to generate all well-formed sentences of a given language but rule out ill-formed ones. And, in his Review of B. F. Skinner's *Verbal Behavior* (1959), Chomsky criticizes the behaviorist stance according to which the human mind is a black box not to be stud-

ied but introduces language as a window into the mind. Thus, since its earliest days generative linguistics is a cognitive enterprise contributing to the cognitive revolution. One of the core tenets of American structuralism was to describe languages in their own terms to not impose classic notions of grammar to languages in which they play no role. In contrast, within a cognitive stance it is essential to approach each language in a way that makes it possible to determine the range and limit of variation. Specifically, on the assumption that there is a cognitive basis for language, linguistic universals are good candidates for being rooted in an (innate) language faculty. And this is precisely what generative linguistics has postulated, in part based on an argument from language acquisition (the classic poverty of stimulus argument). Notably, distinguishing between those aspects of language that are universal (and hence by hypothesis due to our cognitive make-up) and those aspects that vary across languages (and hence must be acquired) requires a formalization that not only allows us to model adult language but also the way it develops in the course of language acquisition.

Another consequence of taking a cognitive stance is that it raises the question as to how to model the relation of language to other cognitive domains. Classic generative linguistics has held the view that the language capacity is autonomous and that it interfaces with other cognitive domains (rather than being reducible to them). Within the minimalist program (Chomsky 1995 and subsequent work), which is the view that Chesi takes as the basis for his criticism, these interfaces are taken to be with the conceptual-intentional system and the articulatory-perceptual system. Syntactic computation is taken to derive abstract representations that are legible at these interfaces (e.g. LF and PF, respectively). This follows the long tradition within generative linguistics of assuming no direct relation between sound (PF) and meaning (LF); rather syntax mediates these levels of representation (i.e. the classic generative T-model, and its minimalist incarnation). This is schematized in (3).

(3) The T-model in its minimalist incarnation




Finally, the cognitive stance of the generative enterprise has methodological consequences. Since what is at stake is the exploration of knowledge of language (rather than its use), there has always been a separation of language competence (the domain of study) from its performance. Since language in use is influenced by external factors, data

collection requires a way to control for such factors. This is the reason why generativists have traditionally relied on well-formedness judgements by native speakers as they are meant to tap into a speaker's knowledge of their language.

Based on these core tenets of generative linguistics and the cognitive stance that defines it, we can identify requirements for modelling, formalization, and methodology, as summarized in (4).

(4) Generative requirements rooted in its cognitive stance
   a. MODELLING: requires a way to capture the relation between language and other cognitive capacities
   b. FORMALIZATION: requires a way to capture the range and limits of language variation and how it is acquired
   c. METHODOLOGY: requires a way to tap into a speaker's knowledge of language and to control for factors that may interfere through language use.

Crucially, Chesi's criticism of generative linguistics is based on one of its instantiations, namely MINIMALISM. In what follows, I argue that taking the requirements summarized in (4) as our benchmark for evaluating generative linguistics, then much of Chesi's criticism falls flat. Specifically, vLLMs do not contribute much towards answering some fundamental questions about language. What is the range of linguistic variation and what are its limits? Should we incorporate little words (like *huh*), which are essential to linguistic interaction, into grammatical analysis, and if so, how? What is the path of language acquisition and how can we model it? How does language relate to other cognitive capacities?

These are the questions which I have been led to exploring by and with generative assumptions.

## *3. Language variation*

I have been fortunate enough to think about the question regarding language variation by exploring languages indigenous to North America, with grammars that appear to be fundamentally different from what I was familiar with, at least at the surface. When I first started working on Halkomelem Salish, the question I got stuck on almost immediately was that of a generative syntactician: How do I draw a tree for the simplest of sentences? And I was not even concerned with establishing the c-command relations yet. I was concerned with trying to understand how to map the categories of Halkomelem onto the functional categories that

made up the trees I was familiar with from having worked on Germanic languages. There were two key issues that made this task so puzzling.

First, Halkomelem has categories that are not found in Germanic languages. For example, verbs obligatorily mark the control the agent has over the event as part of its transitivizing system, and auxiliaries obligatorily mark the location of the event relative to the utterance location. This highlights a more general problem often used by typologists as evidence against universal grammar (Evans & Levinson 2009, Haspelmath 2007): Grammatical categories of the languages of the world appear to vastly differ.

On a purely merge-based approach the question regarding these differences does not arise since labels play no role. Hence, one of the classic generative questions regarding the range and limits of language variation could not even be asked if we ignore category labels. At the same time, generative models which assume trees with labelled categories were not useful either. This holds true for the standardly assumed clausal architecture (CP, TP, vP) as well as its cartographic versions, with their rich inventory of functional categories whose labels mirror traditional grammatical categories. Neither of these approaches includes categories that could accommodate the categories of Halkomelem. This state of the art left me with the puzzle as to what accounts for the Halkomelem inventory of categories and how to understand it within a universalist approach.

Second, Halkomelem has some categories (like past and plural marking) which seem to parallel their Indo-European counterparts, and which led to the postulation of the functional categories TENSE (Pollock 1989) and NUMBER (Ritter 1991). However, on closer inspection this parallel is deceptive as these categories significantly differ in terms of patterns of meaning and distribution (see Ritter & Wiltschko 2014 for past marking and Wiltschko 2008 for plural marking). Following classic linguistic argumentation, according to which categorial identity is deducible from distributional patterns, we are led to conclude that these categories are not the same. Again, differences of this type are insignificant in a merge-based system and hence would not lead us to even ask questions regarding differences in categories. And approaches with labelled structure (classic or cartographic) have nothing to say about distributional differences in seemingly identical categories.

Faced with this conundrum, I developed the Universal Spine Hypothesis (USH, Wiltschko 2014), according to which there is a universal, hierarchically organized structure (the universal spine) where each layer of structure is characterized by an abstract function, such as classifying, point of view, anchoring, and linking. The spine serves as

the universal basis for the construction of language-specific grammatical categories by associating language-specific units of language. Thus, control marking in Halkomelem can be analysed as the functional equivalent of viewpoint aspect. And the location marking in auxiliaries can be analysed as the functional equivalent of tense: it serves to anchor the event to the utterance and has the same distributional patterns as tense in English.

While this analysis of variation in grammatical categories departs from standard generative assumptions, and especially its minimalist instantiations, it nevertheless follows the basic tenets of generative linguistics (Wiltschko 2018). Specifically, it seeks to address the question regarding the range and limits of variation in the realm of grammatical categories. Moreover, it is firmly rooted in the generative assumption underlying the T-model, according to which the relation between form and meaning is mediated by grammar. Specifically, the way the USH implements this insight is through the assumption that the spine comes with functions that enriches the meaning of lexical elements. This allows us to understand why certain aspects of the sound-meaning relation are structurally determined in a universal way. For example, past morphemes are not intrinsically deictic, rather, on the USH, they receive their deictic interpretation by virtue of being associated with the anchoring category of a root clause. In contrast, when associated with embedded clauses past morphemes lose their deictic meaning and turn into markers of dependence. Thus, the core insight of the T-model is essential in explaining the multi-functionality of lexical items: The sound-meaning relation in complex expressions is syntactically mediated and under the USH the spine serves as this mediator.

Consider now Chesi's argument that machine learning methods (and hence vLLMs) are superior to the formalizations generative linguists postulate. It remains to be seen if and how vLLMs can model and formalize linguistic variation such as the differences in grammatical categories. To date, vLLMs are far from being able to provide us with an empirically adequate theory of the differences between English and Halkomelem. One might argue that this is simply because there are not enough data available for minority languages such as Halkomelem and hence vLLMs cannot be adequately trained. But suppose this problem were addressed and we were to have a vLLM for Halkomelem. It is unclear how vLLMs would allow us to understand or model the range and limits of variation, let alone its cognitive underpinnings. The problem I anticipate has to do with the fact that vLLMs are going to be restricted to the surface data in comparing languages to each other. And this will likely result in exactly the state of affairs typologists are facing

and which led to the claim that grammatical categories are not universal. Of course, vLLMs could be trained and the data could be annotated accordingly to reflect the underlying patterns. But this training and annotation would require the intervention of a human linguist.

## *4. Interactional language*

Another domain of research that I came across (quite accidentally) in the Canadian context has to do with the little words that are only found in linguistic interaction. Specifically, I became fascinated by the difference between Canadian *eh* and North American *huh*, first observed and brought to my attention by Strang Burton (personal communication). Both *eh* and *huh* can be used to request confirmation for the truth of the propositional content in the host clause, as in (5a). Significantly, only *eh*, but not *huh* can be used to request confirmation that the addressee is aware of the truth of the propositional content, as in (5b).

(5) a. You have a new dog, {eh/huh}?
b. I have a new dog, {eh/*huh}?

There are many things that make this contrast intriguing from a generative perspective. For example, the fact that there even is a difference in acceptability is striking. Even speakers who are not users of Canadian *eh* share the judgement: *eh* sounds perfectly acceptable in (5b) while *huh* sounds like the speaker is not sure if they have a new dog. Of course, in contexts where the speaker is in fact unsure if they have a new dog, the use of *huh* becomes acceptable. (See Wiltschko & Heim 2020, Wiltschko 2021a for a detailed discussion of the context of use of *eh* and *huh*). The contrast in acceptability suggests that there is a system that underlies that use of these sentence-final particles, a grammar of sorts. Yet, they have been neglected within generative linguistics. There are at least two reasons for this. First, the separation of language competence from its performance implicitly led to the equation of performance with linguistic interaction. Second, the unit of analysis of generative syntax has always been the sentence, an assumption that goes back to classic grammatical treatments of the ancient Greek and Latin tradition. While generative linguistics broke with many of these classic assumptions, the concept of the sentence remained unquestioned. Since *eh* and *huh* (and similar units of language) are restricted to language in interaction and typically appear in sentence-peripheral positions or in isolation, they have never entered the domain of generative investigation.

There are however several properties of these units of language (which I refer to as 'interactional language') that suggest that they, too, are part of the linguistic competence that generative linguists seek to model. First, just because interactional language is used in linguistic interaction does not automatically make it a performance phenomenon. Rather it displays all the hallmarks of grammatical knowledge (Wiltschko 2022), including intuitions about their well-formedness (suggesting that they are part of our linguistic competence), restrictions on their distribution (suggesting that they are regulated by grammar), and systematic patterns of multi-functionality (suggesting that their form-meaning relation is syntactically mediated).

Second, just because interactional language is often realized outside of the sentence proper does not automatically mean that it lies beyond grammar. In fact, since the Principles and Parameters framework (Chomsky 1981), it is one of the core generative assumptions that surface constructions are not primitives of the theory. Arguably, the notion of a sentence, too, is a type of construction that should not be taken as a primitive but is to be decomposed. If so, there is no reason as to why units of language that are used in sentence-peripheral position should be excluded from analysis.

Suppose that interactional language is indeed part of our grammatical knowledge. The question arises as to how to model it while keeping with the core tenets of generative linguistics. This was the question that has inspired me to develop the Interactional Spine Hypothesis (ISH), an extension of the universal spine (Wiltschko & Heim 2016, Wiltschko 2021a). The interactional spine has layers of structure whose function pertains to core properties of linguistic interaction: grounding and turn-taking. While certain assumptions that are at the core of the ISH run counter some minimalist assumptions, it still adheres to the core tenets of generative linguistics.

Consider now Chesi's argument that machine learning methods (and hence vLLMs) are superior to the formalizations generative linguists postulate. It remains to be seen if and how vLLMs can model the intricacies of interactional language (such as the difference between *eh* and *huh*). While LLMs are developing rapidly, I doubt that they will ever be able to fully master interactional language in the way humans do. This is because interactional language is used to regulate the construction of common ground as well as turn-taking. The former requires a theory of mind as the relevant units of interactional language are dedicated to keeping track and marking who knows what in a conversation. For this reason, interactional language is sensitive to who is talking to who and in what context. Currently, vLLMs can crudely mimic the use

of interactional language as can be witnessed by the automatic podcasts generated by NotebookLM. What is clear from listening to such podcasts is that some of the core features that underly our conversational competence are violated. But suppose that this problem can be addressed, and that we end up with vLLMs that can perfectly mimic language in interaction. I doubt that vLLMs would ever allow us to understand and model the universal patterns of interactional language, its variation across languages, dialects, or sociolects, let alone its cognitive underpinnings.

## *5. Child language acquisition*

My interest in interactional language has led me to another strand of research which has been completely unexplored within generative linguistics. That is, once we consider interactional language to be part of our linguistic competence it behooves us to study its acquisition path. Even the most cursory look at child speech reveals a fascinating conundrum. From a very early age, children use interactional language, including the sentence-final particle *huh*. Consider the example in (6), from the Bates corpus. The child uses *huh* in final position at 1;08, long before tense and agreement inflection are in place.

(6) Chuck: Out ball, huh? (1;08 – Bates Corpus)
Mother: Ball out!

From a theoretical perspective this use of interactional language is puzzling. On the one hand, most theories of language acquisition assume some version of maturation such that syntactic structure matures upwards. On the other hand, interactional language is found in the very top of the syntactic structure. Hence, we would expect interactional language to be acquired last, contrary to fact. One way to resolve this puzzle is to assume that syntactic structure matures inwards rather than upwards. This is precisely the idea I have been pursuing in collaboration with Johannes Heim (Heim & Wiltschko 2025). Specifically, we propose that language acquisition starts by linking the highest layer of structure (responsible for regulating turn-taking) with the lowest layer of structure (responsible for classification). Arguably, the first instance of linking is expressed through the pointing gesture around 9 months of age. Once the spine is linked in this way, structure unfolds at both ends with a grounding layer above linking and an anchoring layer below. The inward growing spine hypothesis allows us to understand the acquisition of *huh*: It first serves as a marker to request response, which significantly differs from its

function in adult English. Predictably it can only serve to also mark the epistemic states of the interlocutor once grounding is in place.

Incorporating interactional language into the domain of generative linguistics thus allows for a more comprehensive exploration of the acquisition path. That is, even a cursory look at child language data reveals the seamless combination of units of language that belong to the traditional sentence with those that belong to interactional language. By ignoring interactional language, we miss out on important aspects of language acquisition and arguably depriving us of important clues regarding its cognitive underpinnings and thus regarding our cognitive development. While the inward growing spine hypothesis departs from minimalist assumptions, it nevertheless adheres to the core tenets of generative linguistics. It seeks to explore the cognitive basis for language, and it assumes that the data we observe (in adult as well as in child speech) are regulated by an underlying system (i.e. our linguistic competence).

Consider now Chesi's argument that machine learning methods (and hence vLLMs) are superior to the formalizations generative linguists postulate. Again, it remains to be seen if and how vLLMs can model the language acquisition path in humans. Crucially, children do not acquire language by being exposed to large amount of data nor by being trained. And the human acquisition path looks very different from that of vLLMs. Children do not go from gathering large amount of data to the adult like linguistic state. They go through various stages including babbling, communicating with only one word, etc. And these stages are significant and require modelling. It remains to be seen if vLLMs could ever mimic the child language acquisition process, or model its patterns, including the integration of interactional language and its multi-modal expressions let alone allow us to understand its cognitive underpinnings.

## *6. Conclusions*

I hope to have shown that generative linguistics is so much more than what Chesi makes it out to be. It goes beyond modelling sentences within a single language like English. It is fundamentally a cognitive enterprise. In keeping with the core tenets of the generative enterprise I have been led to a series of fascinating questions and I have developed some answers, which in turn have helped me to uncover new empirical domains, have served as heuristics for discovery and analysis of novel data within a universalist approach (Wiltschko 2018, 2021b), and has led me again and again to ask new questions, some of which have

required me to adapt and change my assumptions about language. To me, this is the ultimate litmus test for a good model: does it allow us to ask questions that open new avenues of research.

In light of this, let me return one last time to Chesi's claim that vLLMs are more solid than any formal theory on the market. I do not see that any of the questions that I have asked over the last 20 years could have been initiated through vLLMs. They cannot be used as heuristics to discover the similarities and underlying differences across the languages of the world including those that are understudied. They cannot be used to explore how interactional language fits in our language competence. And they cannot be used to model the path of language acquisition observed in children.

All of these questions go beyond the questions that vLLMs allow us to ask, let alone answer. And my approach towards language and the generative enterprise differs significantly from the one Chesi portraits as generative linguistics. Maybe these differences could be used in support of Chesi's conjecture that we are facing the end of generative grammar as we know it. But the way I view it is that it signifies not the end but simply change. In my case, it is a change that allowed for the broadening of the empirical domains. Significantly, the change this broadening brings along is not the kind of change Chesi envisions. In the domains I have discussed, generative linguistics is not replaced with vLLMs and arguably it is not replaceable. At the root of this irreplaceability is arguably the core of the generative enterprise, namely its cognitive stance. Clearly, vLLMs do not have the same cognitive architecture as humans do. If they did, then we would be much better at math or at analysing a large amount of data. What we are better at though is the mastery of language as a communicative device which includes so much more than simply the flawless production of sentences.

And even if vLLMs could indeed model language (better than us) in a way that mimics human cognition, this still leaves us with a lack of understanding of human cognition and the role language plays in it. This is because of the lack of transparency of vLLMs. They work, but it is not always clear how they work, either because even their designers do not know or because the industry will not reveal it. And this may be the original sin of vLLMs: its capitalist stance which is diametrically opposed to the scientific enterprise.

*Bibliographical References*

See the unified list at the end of this issue.