Sergio Torres-Martínez*

A predictive human model of language challenges traditional views in linguistics and pretrained transformer research

https://doi.org/10.1515/lass-2024-0018
Received April 18, 2024; accepted October 21, 2024

Abstract: This paper introduces a theory of mind that positions language as a cognitive tool in its own right for the optimization of biological fitness. I argue that human language reconstruction of reality results from biological memory and adaptation to uncertain environmental conditions for the reaffirmation of the Self-as-symbol. I demonstrate that pretrained language models, such as ChatGPT, lack embodied grounding, which compromises their ability to adequately model the world through language due to the absence of subjecthood and conscious states for event recognition and partition. At a deep level, I challenge the notion that the constitution of a semiotic Self relies on computational *reflection*, arguing against reducing human representation to data structures and emphasizing the importance of positing accurate models of human representation through language. This underscores the distinction between transformers as posthuman agents and humans as purposeful biological agents, which emphasizes the human capacity for purposeful biological adjustment and optimization. One of the main conclusions of this is that the capacity to integrate information does not amount to phenomenal consciousness as argued by *Information Integration Theory*. Moreover, while language models exhibit superior computational capacity, they lack the real consciousness providing them with multiscalar experience anchored in the physical world, a characteristic of human cognition. However, the paper anticipates the emergence of new *in silico* conceptualizers capable of defining themselves as phenomenal agents with symbolic contours and specific goals.

Keywords: active inference; agentive cognitive construction grammar; ChatGPT; embodiment; essentialist concept formation; large language models

***Corresponding author: Sergio Torres-Martínez**, Universidad de Antioquia, Cll. 67 #53 – 108, Medellín, Antioquia, Colombia, E-mail: surtr_2000@yahoo.es. https://orcid.org/0000-0002-8823-1676

1 Introduction

In this paper, I introduce a theory of mind that aims to provide a context for the analysis of language as a cognitive tool for the optimization of biological fitness. In doing so, I also explore the potential of construction grammars to become either theoretical contributors to or ontological beneficiaries from the development of large-language-model developments. However, in contrast to traditional construction grammars, I argue that the human language reconstruction of reality is the result of both biological memory and adjustment to uncertain environmental conditions, and the need to define a preferred future in the reaffirmation of the *Self-as-symbol* (see Torres-Martínez 2023a, 2024a,b,c).

Among other things, I show that pretrained language generators are incapable to create a model of the world through language due to their lack of embodied grounding. As I shall indicate further below, autoregressive pretrained transformers create models of the world that reflect an essentialist, single-layered reconstruction of reality that lacks embodied representation owing to an absence of subjecthood and conscious states for event recognition and partition (Torres-Martínez 2024a,b,c). Importantly, although humans engage in symbolic essentialism through language (concepts about natural kinds, for example, are generally modeled on essential and/or intrinsic properties ascribed to an entity or object), our capacity to manipulate syntax and semantics well beyond the prediction of statistically identified tokens enables us to integrate, blend, and overlap eventful experience to create suitable pictures of the world. More concretely put, while humans use their intuitive physics and biological memory to acquire concepts and express them through a vast amount of constructional combinations, *large language models* (LLMs) mobilize large amounts of data to infer syntactic and semantic suitability as a means to increase efficient response rates to a prompt. Since human performance in communication is not defined by controlled task completion but survival, assertions that transformers "exhibit subtle and sophisticated pattern identification and inferencing" (Lappin 2024:11) conflate *performance* with intelligent agency, a hallmark of human language learning, usage, and entrenchment (see Torres-Martínez 2022a,b,c).

Importantly, empiricist readings of LLM performance are not alone in ascribing intelligent agency to AI-generated language. Indeed, amodal language processing theories of language have seen in LLM research a good means to tackle the undeniable paucity in the investigation of how human language is processed and conceptualized in the speaker's mind. These approaches come in a variety of forms depending on the definition of how language is produced and processed. What they have in common is the goal of creating mathematical theories that accurately capture the way meaning is *extracted* from symbols and further mapped onto internal

representations and concepts possessed by speakers immersed in an ideal state of *unity* "often thought of as part of a descriptive theory of ideal rational agents, not of real agents" (Kindermann and Onofri 2021: 2).

It makes sense, then, to begin our investigation by pointing out that this unified representationalist assumption rests on the arbitrariness of the symbols used. In many ways, an undue distinction between the linguistic symbol and perception is created as a condition for *logical closure*, which contrasts with embodied theories that, in many cases, reject the idea that representations are internal and entirely amodal. Although symbolic models depart from different assumptions regarding the ways concepts are distributed and organized, whether they are local (Collins and Quillian 1969), or distributed (Hinton et al. 1986), current symbolic models of cognition, such as *latent semantic analysis* (Landauer and Dumais 1997) or *word2vec* (Mikolov et al. Dean 2013), offer design rules for the manipulation of signs through algorithms that distribute words across semantic layers in "high-dimensional numerical vectors extracted from large amounts of natural-language data" (Günther et al. 2019: 1), following mathematical compression techniques for the construction of representations that work at the level of prediction and performance rates during specific tasks. Running across these distributional semantic models (Günther et al. 2019: 4) is the idea that language can be coded, encoded, and further transformed into retrievable patterns defined as language-as-performance instances capturing human linguistic decisions (thereby becoming "explanatory theories" Landauer and Dumais 1997) within distributional vectors.

We can see the problem with the symbolic account manifest itself in many different ways outside the realm of *Good Old Fashioned Artificial Intelligence*, especially in the context of *generative artificial intelligence* (GenAI) and large language models (LLMs). As we will see in the next subsection, researchers coming from an empiricist and functionalist tradition have gone so far as to elevate LLM performance (that is, language-as-performance) to the status of "the best predictive models of human language representations at the resolution of data [neuroscientists] have access to" (Tuckute et al. 2024: 285), in a manner akin to the claim that symbolic distributional models can "take sensorimotor experience as context" (Günther et al. 2019: 15).

1.1 Humanism or posthumanism in large language model conceptualization

In a similar way, the notion of *language-as-performance* has been integrated into ethnographic (posthumanist) viewpoints seeking to ascribe an ontological status to pretrained transformers in the belief that these language generators are *authentic*

language creators (e.g., Demuro and Gurney 2024). According to this reading, such models actively contribute to the cultural construction of knowledge through unique reasoning processes that, nonetheless, are distinct from those of humans. This assertion stems from a rejection of a "universalist" definition of language, which, according to some authors (e.g., Braidotti 2013), is often associated with a male-centered humanist tradition, and instead views language as a manifestation determined by the performance of communication. Consequently, pretrained language models are seen not just as performers of language but as active participants in its creation, thereby shaping events where language unfolds through both performance and competence, which "relativize[s] the human by coupling it to some other order of being" (Clarke 2008: 3). Accordingly, human language is depicted as a process of assembling symbols within a pre-established grid of semiotic relations, with speakers functioning as participants bound by the rules of extant signhood preservation practices within a communicative event. As a result, human language is portrayed as a residual aspect of interaction, lacking any inherent transcendental qualities ("humanness"), while positioning LLM language generation as an integral component of authentic "encounters" where participants, both human and nonhuman, share a sense of experiential adjustment (e.g., Demuro and Gurney 2024; Dynel 2023). This shift obliterates any claim to a construction of the Self-as-symbol through an emphasis on a redefinition of power relations that replace universalism with a focus on addressing gender inequality.

> My antihumanism leads me to object to the unitary subject of Humanism, including its socialist variables, and to replace it with a more complex and relational subject framed by embodiment, sexuality, affectivity, empathy, and desire as core qualities (Braidotti 2013: 26).

However, since human agency involves purposeful, goal-oriented action aimed at both reducing uncertain states and constructing preferred models of interaction with the environment, human beings, as distinct biological systems, participate in a continuum of relations with the world and other species that goes beyond denaturalized claims to power redistribution. The primary contradiction in this post-humanist reasoning lies in the fact that even the most contingent set of relations is driven by biological imperatives that cannot be relativized through the postulation of indefinite intrinsic properties:

> For posthuman theory, the subject is a transversal entity, fully immersed in and immanent to a network of nonhuman (animal, vegetable, viral) relations (Braidotti 2013: 193).

Hence, if we were to accept that artificial language generators create culture simply by existing in the world through interactions with humans (Demuro and Gurney 2024), we are also compelled to believe that humans, as open biological systems,

depend on an infinite, regressive chain of decisions made at a molecular level by noncognizant interpreters in order to justify their existence.

This form of semiotic atomism, as part of what some have called "new materialisms" (e.g., Coole and Frost 2010), assigns portions of purposeful animacy and cognition to fine-grained instances or states of a system's parts, under the belief that bottom-up self-organization cannot be explained by observing the system *as is*, but rather as it could be under different environmental conditions (e.g., Watson 2023). This antihumanist stance is summarized by Pepperell's assertion (2003: 2) according to which "the balance of dominance between human and machine is slowly shifting." In other words, human uniqueness is being evaluated through an arbitrary and biased notion of efficiency (through information capacity and memory, e.g., Cantlon and Piantadosi 2024) that renders human agency unnecessary and relativized as idiosyncratic manifestations of cognitive actants drowning in a sea of "material forces [that] manifest certain agentic capacities" (…) (Coole and Frost 2010: 10). Inevitably, the blurring of the line between conscious (human?) and unconscious (artificial) states (subsumed to the rules of performance), embedded within unwarranted part-whole relations, fails to provide a coherent characterization of subjecthood and experience as core elements of organic agency, as the following passage reveals:

> The *zoe*-centered embodied subject is shot through with relational linkages of the contaminating/viral kind, which interconnect it to a variety of others, starting from the environmental or eco-others and include the technological apparatus. (Braidotti 2013: 193)

The obvious result of this reasoning is the postulation of political relativism "devoted to the ['critical'] analysis of actual conditions of existence and their inherent inequality" (Coole and Frost 2010: 25). As we shall see, the postulation of humans, "a strange product of evolution that can perceive itself and its fellow humans as paradoxical" (Rosendahl 2013: 224), as self-symbols, striving for biological priority and persistence, is a requirement for the definition of language as a cognitive tool that cannot be emulated by artificial systems of language generation engrained within a fuzzily and ideologically defined material environment in which "there is no definitive break between sentient and nonsentient entities or between material and spiritual phenomena" (Coole and Frost 2010: 10). Importantly, my use of "persistence" does not seek to position a code-driven idea of biological reality through heredity. As it will become evident in the next pages, my idea is, rather, anchored in the definition of signhood within a context of predictive semiosis, and not the passing of information through DNA-based heredity (Birch et al. 2020: 56).

1.2 Defining consciousness

Where is, then, consciousness to be found? According to *Information Integration Theory* (IIT), the question of consciousness needs to be defined from a first-person perspective. In this vein, IIT considers two problems: (1) the extent to which a system can be considered as having conscious experience and (2) the definition of the conditions that determine the type of consciousness a system has.

This phenomenological approach departs from the assumption that, for a system to be conscious of itself and of what is going on around it, it must, first, be able to deal with large amounts of information (differentiation), and second be capable of integrating that information into a coherent semiotic picture of the world (integration). Moreover, the whole process is said to be private and occur in specific areas of the brain. The main weakness of this approach is that it cannot address some important questions regarding the role and nature of consciousness, especially with regard to purposeful action. Whereas some could argue that a focus on a need to cope with uncertainty in an ever-changing environment is enough to account for the rise of different types of consciousness, it is a fact that living organisms are not isolated systems but create complex networks of relations with other organisms, including those considered as exhibiting disparate levels of consciousness. In other words, consciousness is possible because it provides us with a sense of extended community that creates connections with others. Consciousness is thus constructed upon patterns of interaction associated with agentive perception that can be projected onto the environment and other entities. The implication is, of course, that the one condition of conscious states is that can be shared an enhanced by collective experience. This leads us to a second objection to ITT: its reductionism of conscious states to an impoverished version of the brain as a physical support of experience (a quantity).

The present proposal defines consciousness as a combination of four main elements: (1) The identification of objects in the world associated with events of perception (experience is intrinsic, though modeled on shared experiential structures in common with other organisms). (2) The construction of pictures of the world through the integration of modal and amodal information processes in the brain (information is provided by different modalities and then organized in the brain in specific areas). The nature of experience is defined by degrees of perceptual access to *complexes* (composite concepts) and not by simulations of experience in the form of "*shadows of perception*." (3) The content of experiential states becomes integrated into the agentic projection of intentional states onto the world. (4) The alignment of subjective conscious states with other conscious entities is facilitated by the construction of a model of the self both as individual and part of a system of agentive

relations and hierarchies. This notion of bodily awareness, as an instance of agency, is indebted to the *free-energy principle* (Friston 2009, 2010) whereby organisms need to keep a balance between environmental and inner-body information in order to avoid entropy. This defines the role of biological systems as predictive entities moving toward specific *attracting states*, that is, points of equilibrium that minimize entropy in order to attain homeostatic integrity. Homeostasis (inner-system equilibrium) is accomplished through the identification of sensory states by means of *ad hoc* receptors placed in a statistical boundary (*Markov Blanket*) providing a boundary between system-external and system-internal conditions.

This conceptual move is of great importance, as one of the motivations to compare biological and artificial systems is what I refer to as *increment bias*.[1] According to this concept, artificial intelligence, specifically large language models, would exhibit rates of evolutionary advancement that surpass biological transitions regulated by environmental constraints. From this perspective, humans would be at a disadvantage compared to *in silico* "cognizers" due to our constrained evolvability. However, the constitution of a semiotic Self is not the product of *reflection*, that is, the possibility of a computational architecture to "form representations of control flow that affect the control flow itself" (Barron et al. 2023: 5).

The paper is divided into five main sections and a conclusion. In Section 1, I introduce the theoretical tenets of *Agentive Cognitive Construction Grammar* (AgCCxG, Torres-Martínez 2018a,b, 2019, 2020, 2021a,b, 2022a,b,c, 2023a,b,c), a predictive theory of mind and language for the definition of the biophysical underpinnings of language. In Section 2, I offer an in-depth exploration of AgCCxG. In particular, I show how constructions possess a triadic structure combining form, function, and agency that become active in specific events thanks to the action of constructional attachment patterns. In Section 4, I analyze the forms of essentialism encoded in LLM. Among other things, I demonstrate that generated language sequences are for the most part generalizations from flattened content that lack

1 While artificial systems, such as large language models, may exhibit rapid advancements that seem to outpace human biological evolution, an *increment bias* underscores a key difference between human and artificial symbolic processing, namely that human cognition is deeply tied to our *semiotic Self*, where language is more than just a tool for communication – but an adaptive, symbolic system that evolves within a context of lived experiences. Thus, central to an evolutionary increment bias is the need of stability, adaptability, and dependability of cognition, which artificial systems lack. This form of persistence allows human cognitive systems to remain stable and functional even within the slower evolutionary processes governed by environmental constraints. It is not merely the result of computational processes or control flow, as seen in artificial systems. Instead, human cognition and language endure because they are part of a complex network of lived experiences, semiotic processes, and biological embodiment. Thus, the persistence of the human semiotic Self provides a form of cognitive stability that artificial systems, despite their rapid advancements, cannot replicate.

embodied grounding, in contrast to natural language. In Section 5, I reveal the reasons why current usage-based approaches to language are at odds theoretically. One of the main reasons for this is the characterization of language as an intrinsically computational code for performance and statistical entrenchment of anecdotal constructionel combinations.

2 A predictive semiotic theory of mind

Agentive Cognitive Construction Grammar (AgCCxG 2018a,b) is a theory of mind and language that favors a strong embodied characterization of experience through a focus on concept formation and semiotic hybridization over time. An important consequence of this characterization is the definition of mind as a continuum of perceptual loops embedded in eventful states of consciousness. This is conditional upon precisely the existence of biological memory and entropy-reduction strategies. Another way of putting it is to say that uncertainty reduction, also known as free-energy minimization, also aims at facilitating the identification of the contours of the Self as an open biological system. The impetus for this claim is inspired by *Active Inference* (AIF), a process theory, adopted by different fields (e.g., computational neuroscience Friston 2005, 2009, 2010; Clark 2013; *Predictive Processing*, Clark 2013; or mathematical frameworks such as the *Free-Energy Principle*, Friston 2010), as a means to provide a formal scaffolding to the biological imperative of free energy minimization. Among other things, AIF states that, in order to reduce *entropy* (that is, the breakdown of the contours of self-organizing systems), living organisms need to apply statistical predictions of the state of the world enabling them to cope with *surprise* (aka. free energy). In this paper, AIF is, hence, considered as a framework for understanding how organisms actively gather information and take action in their environment to maintain a desired state or achieve a goal. This stresses the fundamental relationship between ground physical conditions supporting the organization and optimization of matter-energy exchanges within living organisms striving for priority and persistence in a specific Umwelt. This is expressed mathematically using Bayesian models that integrate probabilistic beliefs about the world and observations as a means to guide decision-making and action. The core AIF equations include the following:

1. **Bayes' rule**: P(h|d) = P(d|h) * P(h)/P(d) where P(h|d) is the posterior probability of a hypothesis h given data d, P(d|h) is the likelihood of the data given the hypothesis, P(h) is the prior probability of the hypothesis, and P(d) is the marginal probability of the data.
2. **The free energy principle**: F = –log P(d) + D_KL(Q(h|d) || P(h)) where F is the free energy, P(d) is the marginal likelihood of the data, Q(h|d) is the approximate

posterior distribution over hypotheses given the data, and P(h|d) is the posterior distribution over hypotheses. The Kullback–Leibler (KL) divergence between Q(h|d) and P(h) measures the distance between approximate and true posterior distributions.

3. **Action selection**: a* = argmax_a E_q(h|d)[log P(o|h,a)] where a* is the optimal action, q(h|d) is the approximate posterior over hypotheses given the data, P(o|h,a) is the likelihood of observations o given the hypothesis and action, and E_q(h|d)[ ] denotes the expected value of the expression inside the brackets under the distribution q(h|d).
4. **Perception and inference**: q(h|o) = P(o|h) * exp(–F)/Z where q(h|o) is the posterior distribution over hypotheses given observations, P(o|h) is the likelihood of the observations given the hypothesis, F is the free energy, and Z is the normalization constant that ensures the distribution q(h|o) integrates to one over all possible hypotheses. This equation describes how organisms use their beliefs about the world and sensory observations to update their understanding of the environment and guide their actions.

I argue that this perceptual loop is first defined by specific modalities (interoception, exteroception, proprioception, or visceroception) in a *Markov blanket* (a term originally introduced by Pearl 1988, and later adapted by Friston 2010), and further organized by a modular structure of neuronal networks. It has to be said at this point that while Pearl's formulation of a Markov blanket is purely statistical, Friston's Markov blanket points to a true boundary between the internal and external states of a system, in a manner akin to a cell's membrane. For present purposes, however, a Markov blanket is not considered as a distinct boundary but refers to the dynamic influence (energy exchange momentum) of an agent on an environment (see Torres-Martínez 2024a,b,c), whereby the system and the world can be distinguished from each other (the system does not possess a distinct boundary itself, though). This type of *open Markov blanket* can be formalized by using a Bayesian framework that incorporates both the agents' internal model of the environment and the actual sensory inputs they receive. The equations for the *open Markov blanket* can then be defined as follows:

1. **The agent's internal model of the environment:** P(s_t | o_{1:t-1}, a_{1:t-1}, m) where s_t is the agent's internal state at time *t*, *o_{1:t-1}* are the observed sensory inputs up to time t-1, a_{1:t-1} are the actions taken by the agent up to time t-1, and m is the agent's model of the environment.
2. **The actual sensory inputs received by the agent:** P(o_t | s_t, a_t, e) where *o_t* is the sensory input received by the agent at time t, *s_t* is the agent's internal state at time t, a_t is the action taken by the agent at time t, and e is the actual state of the environment.

3. **The agent's policy for selecting actions** based on its internal state and model of the environment: P(a_t | s_t, m)
4. **The dynamics of the environment**, which determine how the environment evolves in response to the agent's actions: P(e_{t+1} | e_t, a_t)
5. **The feedback loop** that updates the agent's internal state based on its actions and the environment's responses: s_{t+1} = f(s_t, o_t, a_t, e_{t+1})

In this formulation, the open Markov blanket refers to the dynamic influence of the agent on the environment, whereby the system and the world can be distinguished from each other. This dynamic influence is captured in the feedback loop that updates the agent's internal state based on their actions and environmental responses. In other words, *intelligent agents* make decisions about their action upon the world as a means to maintain equilibrium between the preservation of their model of the environment (Self-as-symbol) and the imperative to reduce uncertainty. This can be summarized in a working checklist (see also Torres-Martínez 2023a):

1. *Intelligent agents* use language as a predictive tool for an adaptive action on the world.
2. Language bears traces of embodied content in the form of presemiotic cognitive routines.
3. Embodied experience is *direct* and *ecological* (that is, representational layers are not added to events in the brain).
4. Embodied cognition reflects a need to reconstruct events on the basis of bodily acquired information.

3 Language as a predictive semiotic tool

According to this reading, language is a semiotic tool that reflects the dynamics of biological adjustment and self-preservation. The upshot is that there is a wide range of phenomena pertaining to language that involve learnability, actualization, and biological adjustment and that can be translated functionally as the interplay between efficient and nonconflicting sign-combination and context-bound agentive decisions. This characterization requires us to view language as a sign system consisting of constructions encompassing three core elements: *form*, *meaning*, and *intelligent agency* (see Figure 1). Under this view, constructions emerge as linguistic components for the conceptualization of an event defined by specific *constructional attachment patterns* (CAPs), that is, embodied constructional arrangements that organize both high-level (abstract syntactic patterns) and lower-level constructions

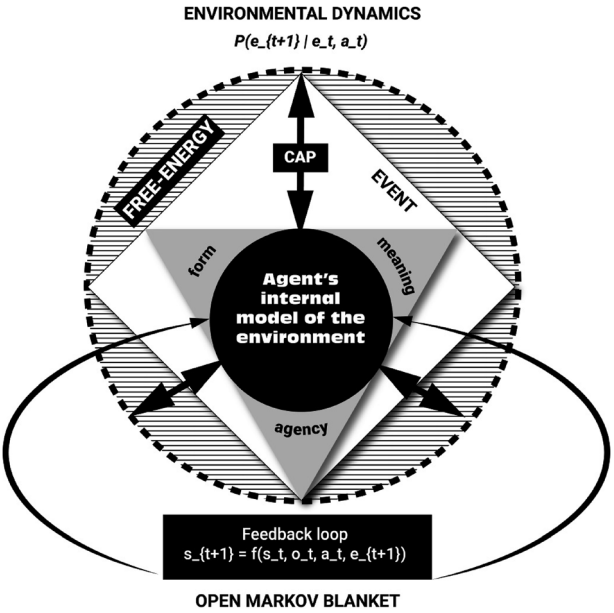

Figure 1: The model of triadic construction. The agent's internal model of the world provides the basis for the partition of events into discrete sequences of actions and conscious states that motivate the assembling of constructional elements though specific CAPs. CAP action aims at reducing free energy by way of an estimation of the state of the world measured through a feedback loop within an open Markov blanket.

such as words, morphemes, affixes, etc., in networks of interconstructional relations motivates an agent's need to offload intentional states onto specific contexts of use.

CAPs reveal different degrees of embodied experience with the physical world that provide concepts with unique tridimensional qualities (concepts are not mechanically retrieved from within a fixed symbolic system). As coefficients for semiotic relations among constructions, CAPs are driven by hypotheses for the reconstruction of events (Torres-Martínez 2023a,b). This reconstruction process encompasses the integration of sensory and experiential aspects, as depicted by the formula below:

$$\text{CAP} = tc \times er$$

where *tc* represents the weight of the connection between a *triadic construction tc* and an embodied representation *er*. These weights reflect the strength of the association between the linguistic and embodied components and can be entrenched through exposure and reinforcement. For example, the sentence "The sleeping snake

in my belly woke up and started to slither" (expressing intransitive motion) can be analyzed as follows:

Argument structure construction	Form	Meaning	Intelligent agency	Formalism
Intransitive motion	(Subj) $VOblique_{PATH}$	X moves to Z	IA_intransitive_motion	IntMot = IA + (Σw_k_im/TC_im)

Components:

- Argument structure construction: Describes the syntactic structure and form of a specific construction.
- Form: The linguistic expression or pattern that represents the argument structure construction.
- Meaning: The semantic interpretation or function associated with the construction.
- Intelligent agency: Represents the influence or involvement of an intelligent agent in the construction.
- Equation: The equation that incorporates the intelligent agency factor and the associated triadic construction into the calculation of free energy.

Thus, instead of engaging in rough predictions on the basis of stored symbolic chunks (e.g., Goldberg 2006–2019), intelligent agents organize linguistic content by engaging in predictions about the outcome of specific usage within events, or chunks thereof. These predictions are motivated by world experience, rather than unmotivated, within-the-system linguistic permutations and generative self-organization. To assess this, we need to explain what perceptual experience is supposed to be. In the first place, language is constructed on concepts, and concepts point to a need to recognize the substance of objects and entities in the world. However, essence identification is not construed as a simplistic assignment of value to a set of features on a continuum of prototypical characteristics possessed by entities and objects that could be described, for example, by way of single-layered propositions. Therefore, one main point of the present theory is to suggest that "unobservable essences" are not homogeneous, linearly arranged properties of identifiable natural kinds, but portions of experience retrievable across various layers of perceptual modeling in a bottom-up manner. The idea I am advancing here, then, is that language links these essences to observable surface characteristics through specific affordances, which can be represented as a vector $E = [e_1, e_2,\ldots,e_k]$. On this construal, the essential

properties of a natural kind, say, a tiger, is not simply a symbolic representation of the animal's intrinsic properties in a conceptualizer's mind, but a reflection of a network of potentialities inherent in the events where these essences are active.

Importantly, in defining the existence of an entity or object a multilayered set of propositions needs to be posited following this model: "If condition S becomes true upon the lower scale length Δ*L*, then condition *S** will become true on the higher scale length Δ*H*" (Wilson 2023: xiv–xv). This formulation underscores the intricate interplay between diverse perceptual scales and the multifaceted interpretation of properties as they relate to natural kinds[2] across these scales. From this perspective, a trait or property deemed essential for an agent's survival at one scale does not preclude its classification as an intrinsic property at another level. For instance, in discussing tigers, the absence of stripes may not inherently signify the core properties associated with tigerness. However, this absence may pose a disadvantage, as it becomes a prerequisite at a related experiential layer, such as camouflage, and even more critical as an asset at the layer pertaining to survival through hunting. In this context, the fur property cannot be considered in isolation from its functional value in the natural environment. Similarly, the question of why propositions such as "Redness and squareness are intrinsic properties. Being next to a red object is extrinsic" (Vallentyne 2014: 31) are rendered nonsensical by the multilayered nature of linguistic reconstruction (see Torres-Martínez 2024a).

2 From this viewpoint, it becomes clear that brain ontologies related to language formation and processing cannot be assumed solely on established methodological and epistemic principles. This challenges the notion that identified brain regions provide definitive evidence of ontological reality (e.g., Fedorenko et al. 2024). Likewise, claiming that a linguistic natural kind in the brain, loosely defined as "an ontologically meaningful grouping of brain areas" (Tuckute et al. 2024: 282), can be substantiated by the traditional encoding–decoding framework of symbolic string generation is untenable. Importantly, the symbolic output-oriented task design, influenced by the deficiency-fallibility bias, portrays human "performers" as "fallible but boundedly rational agents who can, in principle, overcome their ignorance if the task falls within cognitively permissible inference rules" (Solaki 2022: 543; emphasis added). The preceding passage highlights a growing irony in modern algorithmic-driven epistemologies: while the original goal of the Turing test was to determine "whether someone could accurately identify if they were interacting with an AI or a human in a decontextualized setting" (Youssef et al. 2023: 6, citing Nov et al. 2023), the current focus has shifted to examining how well humans can create representations that enable efficient communication with machines using a limited set of predefined signs based on language model training. This algorithmic mindset has led to studies positing LLMs as "mind theorists" (Strachan et al. 2023), rationally bounded agents (Yax et al. 2024), or human-like intuitive thinkers (Hagendorff et al. 2023), which points to an increased bias among many researchers toward LLM humanization and attribution of mental states. This has been mostly operated on the rationalistic premise that linguistic data alone can account for stable states of the world across natural language, which, as a result may induce a correct reconstruction of reference from raw text (e.g., Søgaard 2024).

Clearly, this reasoning runs out of wiggle room quickly, since the existence of stripes and orange fur (an evolutionary trait) is intrinsic to successful hunting (although irrelevant for coarse visual tiger identification) and may become an extrinsic trait for the Chital, whose sensory affordances[3] could, at some point, be deceived by the tiger's stripes, in which case this fully functional biological system becomes a potential prey. So, to conclude, it is not the case that "[w]hether something is red, or 3 kg, or round is a matter of how it itself is, regardless of anything else." (Denby 2014: 91).

Moreover, at the heart of the tiger-chital affordance war is the idea that local intrinsic properties are not merely parcels of experience brought to perceivers through modality-specific representations. We can summarize this assertion as follows:

1. Essence Identification and Perceptual Modeling:
   - Model 1.1:
   - Variables: $E = [e_1, e_2,\ldots,e_k]$ (Essence vector), $S = [s_1, s_2,\ldots,s_n]$ (surface characteristics)
   - Equation: $E = f(S)$
   - Description: The essence vector E is a function of observable surface characteristics S, representing the retrieval of essences across diverse perceptual modalities.
2. Gradation and Transference of Properties:
   - Model 2.1:
   - Variables: ΔL (lower scale length), ΔH (higher scale length)
   - Equation: $S* = g(S)$
   - Description: Conditions at lower perceptual scales ΔL influence conditions at higher perceptual scales ΔH, indicating the transference of properties across perceptual layers.

---

3 According to Gibson (1969, 1977), embodied experience is regulated by environmental affordances, denoted by A, which serve as anchor points for action. An affordance is represented as a function f: E → A, where E is the set of environmental features and A is the set of potential actions that an entity can exploit. Affordances are determined by the phylogenetic configuration of the entity, denoted by φ. Thus, an affordance can be defined as A(φ, E) = f(E).

Objects in the environment possess affordances that can be utilized by entities based on perceptual cues such as weight (W), size (S), and form (F). These perceptually accessible properties influence the affordance function, resulting in the emergence of specific action profiles. Mathematically, this can be expressed as A(W, S, F, E) = f(W, S, F, E) of perception in the construction of representations. This occurs thanks to semiotic mediators of embodied experience by which entities perceive and interact with their environment based on the potential actions available to them according to their phylogenetic configuration and perceptually accessible environmental features.

3. Complexity and essential traits:
   - Model 3.1:
   - Variables: E_t (essence vector for tigers), E_c (essence vector for chitals)
   - Equation: E_t = *[stripes, fur_color,…], E_c = [stripes, fur_color,…]*
   - Description: Essential traits for tigers and chitals contribute to the complexity of perceptual models, with traits like stripes and fur color influencing both intrinsic properties and functional significance within ecological contexts.
4. Contextual Relevance of Properties:
   - Model 4.1:
   - Variables: C (contextual relevance)
   - Equation: $C = h(S)$
   - Description: Properties exhibit contextual relevance across perceptual scales, impacting both intrinsic and extrinsic attributions.

It is evident that this situation does not bode well for theories endorsing a linear, deterministic mapping of experience onto language, since the multifaceted nature of properties and their contextual relevance across various scales of perception does not allow for a one-dimensional assignation of content to concepts. As a result, it is not possible, to simply "ascribe properties to something" and then move on to state that considering a thing as being part of something larger is no longer a proposition about the thing itself but about something intrinsically different. This is crucial to consider, especially when we acknowledge that the constructions comprising a language, along with the theories applied during their assembly in online discourse production, require more from the speaker than mere estimation of truth values (see Section 4). Clearly, the sum of linguistic routines, amendments, possible hesitations, creative expansion of semantic content, in short, the lived experience of surviving through a semiotic system tuned to the world, is not accessible to language models whose only imperative is to define statistically relevant patterns of symbolic combinatorics in a one-dimensional context of usage.

All of these considerations prompt us to ask: Is it feasible to disregard the intrinsic–extrinsic dichotomy while recognizing the nuanced intricacies of perception and the propositions that delineate the underlying processes? Describing this connection as one of "ontological dependence" (Marshall and Weatherson 2002), wherein the existence of one entity relies on the existence of another, presents its own set of challenges. Ontological dependence does not pertain to the structure of a system in relation to another system; instead, it represents a facet of the relationship between concepts and the perceiver/conceptualizer within a dynamic continuum of attentional states (see Torres-Martínez 2024a).

This leads us to suggest a definition of consciousness, denoted as C, as the understanding of an external reality (R) in conjunction with a constructed model of

the Self (*S*) within a specific parcel of reality. Mathematically, this can be expressed as:

$$C = R \cup S$$

Essentially, this theoretical standpoint refutes the notion of the existence of internal representations. From this perspective, neither gorillas nor drosophilae possess a symbolic repertoire (SR) capable of being mapped onto real-world objects or entities (O) (see Torres-Martínez 2024a). This can be represented as:

$$\mathrm{SR} \nrightarrow O$$

This is compatible with the idea that the strengthening of a sense of causal dependence, through the coupling of low/higher order perception and concept formation, is the hallmark of consciousness, rather than the ability to compute by means of a code enabling us to bring to ourselves a picture of the world. So, in the present theory, representation stems from the interplay of three types of concepts:

1. **Core concepts**: (Phylogenetically defined).
2. **Contingent concepts**: Constructed upon the interaction of a system's affordance repertoire and the structure of the environment.
3. **Expanded concepts**: Updated patterns of agentic action upon an eco-niche that can be split into two components for non-cognizant open systems.

Natural language production, as discussed earlier, requires a broad range of experiences tied to our sense of self – something that artificial language models cannot replicate. This complex representation of the self in the world relies on the combination of our perceptions, how we conceptualize them, and the beliefs we project onto specific events in different contexts (see Torres-Martínez 2024a,b).

Take for example the verb "slither" in English. This phrasal verb is typically associated with animate entities having a cylindrical body shape that constrains their movements in a certain way. In particular, snakes are said to *slither*, that is, to move in either a rectilinear, lateral undulation (sliding contact without gripping), sidewinding, or concertina mode (periodic static gripping), in contrast to limbed animals. The assignation of content during a "slithering event" thus entails an iconic reconstruction of body shape, skin texture, incline, surface roughness, as well as the rhythmic alternation of muscle flexion as a means to move forward. These core elements can be expanded metaphorically provided that shape and mode of locomotion do not create a conflict with the original embodied elements encoded in the verb. As a result, the particular selection of the proper verb in order to describe specific modes of locomotion depends on the assignation of affordances to an entity. Thus in the sentence, "The car silently *slithered away* **from** the curb," the sinuous manner of locomotion attributed to snakes (and also to snails) is projected onto a

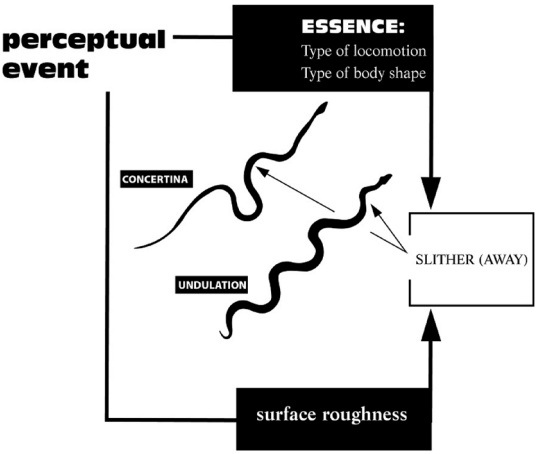

Figure 2: A succinct model of iconic essence mapping in language. As shown, both surface structure and body shape are encoded in the verb "slither [away]" as a means to describe a specific type of movement associated with cylindrical body shapes.

self-propelled object as a means to create a particular effect on the reader, whereby bodily experience, and not only statistically driven construction combinatorics, is summoned. As shown in Figure 2, the genesis of concepts encoded in verbs is attributable to a battery of relations pertaining to the experience of agents with the physical world, and that may include well-entrenched essences associated with visual and motor modalities. These can be considered as core, essential elements that link iconic representation with body shape and a type of locomotion. However, as already mentioned, the embodied reconstruction of events can be distributed over several layers of experience where the description of specific phenomena require a nuanced treatment of what can be considered as an essential trait or an intrinsic property of a natural kind. For example, depending on the texture of a surface, or the structure of it, snakes can adopt either a concertina or an undulating mode of locomotion, which reveals that "to slither" is an essentialist term that cannot capture all of the possible facts about that particular event of locomotion. As a result, to "silently slither away" is not a property of cars but a manner of movement attributed by an agent to the vehicle in that particular event, which reveals that the distinction between what is essential and intrinsic is also the product of the intention of the agent conceptualizer.

4 Large language models' statistical essentialism

The spectacular advancement in *natural language processing* (NLP) in the field of artificial intelligence has led to the consolidation of language models like the GPT (Generative Pretrained Transformer) series. This research has also impacted our

views of language and intelligence. In particular, formal linguistic models of natural language have been challenged by the capacity of transformers to generate language without recourse to syntactic trees, innately given rules, or specific form-function pairings (constructions), as postulated by usage-based approaches (Goldberg 2006). An important feature of these models is that they undergo extensive pretraining on massive corpora of text data, which results in remarkable performance across a wide spectrum of NLP tasks, ranging from language translation and text summarization to question-answering. ChatGPT stands out for its exceptional performance and versatility, showcasing promise across various domains including education (Frieder et al. 2023; Kim and Su 2024) and healthcare (Imperial et al. 2024). Its multifaceted utility underscores the transformative potential of advanced NLP technologies in addressing real-world challenges and enhancing human–computer interactions.

In the case of ChatGPT, a language model that generates context-specific language according to a user's prompts, the process of generating sentences that reflect real-world physics, as well as the semantic nuance introduced by a human speaker is divided into eight steps whereby language is transformed into symbols that fit a preexisting statistical model of semantic distribution and syntactic arrangement. In the case of the phrasal verb "to slither away," the prompt sentence "the car silently slithered away from the curb" is first "tokenized," that is, the sentence is chunked into individual words: "The," "car," "silently," "slithered," "away," "from," "the," "curb." Then, each token is mapped to an *embedding vector* in a high-dimensional space. For example, the embedding for "car" might be represented as $car_e car$. Input tokens are then processed in *ChatGPT's Transformer architecture* wherein the input tokens go through multiple layers of self-attention and feed-forward neural networks. An attention mechanism helps the model to focus on relevant parts of the input sequence. For instance, when generating the word "slithered," the model may pay more attention to the preceding words "car silently." In the next instance, "training," ChatGPT uses a large corpus of text data "to learn" and ultimately predict the next token given the preceding context. For example, it learns that "slithered" often follows descriptions of movement like "silently" (the reconstruction of the event is, hence, purely statistical). In this sense, when generating text, ChatGPT predicts the next token based on its learned probabilities. For instance, given the context "The car silently," the model might predict "slithered" with high probability. After generating the sequence of tokens, *detokenization* reconstructs the original sentence from the tokenized form. For example, the model merges subwords to form complete words and handles punctuation. In this case, detokenization would convert the tokens back into the original sentence: "The car silently slithered away from the curb." Finally, the generated text may undergo further postprocessing such as formatting adjustments (see Figure 3).

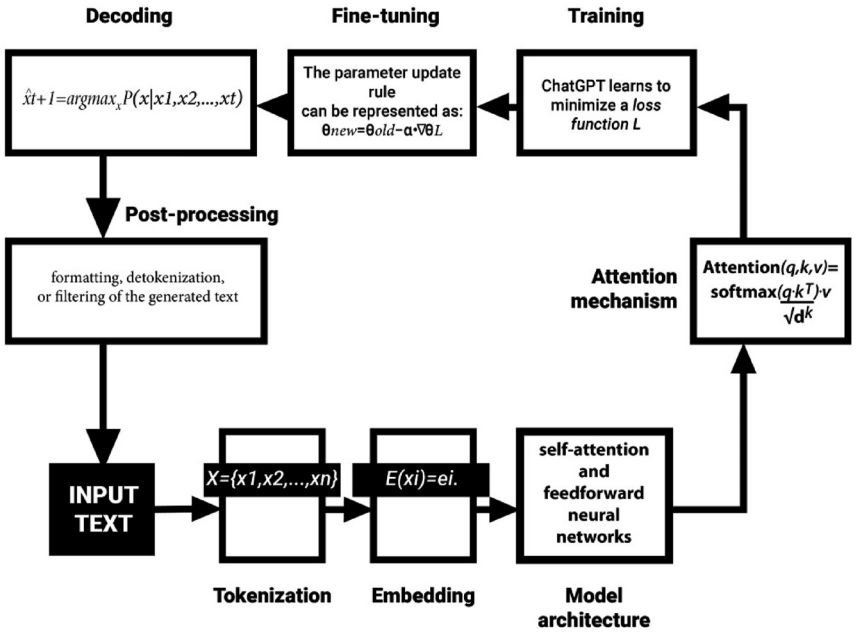

Figure 3: The cycle of language generation of ChatGPT. The process of language generation follows specific steps. Tokenization involves splitting the input text *X* into individual tokens, represented as $X = \{x_1, x_2, \ldots, x_n\}$. Each token xi is then transformed into a high-dimensional vector representation, ei, through embedding: E(xi) = ei. ChatGPT utilizes a Transformer-based architecture, comprising layers of self-attention and feed-forward neural networks, with each layer's output computed as Layeri(input) = LayerNorm(MultiHeadAttention(input)+input). Within this architecture, attention mechanisms enable the model to focus on different parts of the input sequence, computed as Attention(q,k,v) = softmax(q·kT/$\sqrt{}$dk)·v. During training, ChatGPT learns from text data, minimizing *cross-entropy loss*: L(Y,Y^) = $-\sum$iyilog(y^i). Optionally, fine-tuning on domain-specific data may occur, updating parameters via θnew = θold−α·∇θL. Decoding involves selecting the next token x^t+1 based on its conditional probability given the preceding context: x^t+1 = argmaxxP(x|x1,x2,…,xt). Finally, generated text may undergo postprocessing steps such as detokenization, filtering, or formatting.

(1) a) The snake silently ***slithered away into*** the dense undergrowth.
b) After being startled, the lizard quickly ***slithered away under*** the rock.
c) Sensing danger, the eel gracefully ***slithered away from*** the approaching predator.
d) As the light faded, shadows seemed to ***slither away into*** the darkness.
e) Noticing the suspicious movement, the spy discreetly ***slithered away from*** the crowded room.
f) The thief attempted to grab the treasure, but it seemed to ***slither away from*** his grasp.

g) With a flick of its tail, the catfish ***slithered away from*** the fishing line, escaping into the depths of the pond.
h) Despite its size, the slug could effortlessly ***slither away from*** potential threats.
i) The fog slowly began to dissipate, its tendrils seeming to ***slither away with*** the rising sun.
j) Startled by the sudden noise, the insects ***slithered away from*** the fallen branch, seeking shelter in the nearby foliage.

As we see in example 1a, ChatGPT can aptly predict verb usage from a simple prompt sentence. However, the model's responses are limited to a fixed syntactic pattern that, although grammatically correct, does not necessarily capture event structure properly. Since the prompt was "Write ten sentences using this sentence as a model,"[4] ChatGPT recognized the pattern but quickly run out of suitable content to fit

4 The analysis of language generation in Large Language Models reveals fundamental limitations in their ability to mimic human communication. While these models demonstrate an impressive command of linguistic fluency, they rely on statistical prediction rather than the grounded understanding of world knowledge that human speakers mobilize for meaningful interaction. This disjunction between linguistic *form and meaning*, where form becomes ***an allegory for understanding***, offers a critical lens for examining the outputs of LLMs. For example, as stated by Habermas (1989), by employing insights from *Adorno's critical theory* and Brecht's dialectical approach, LLMs' generated language can be construed as a semiotic system operating in the realm of *allegory* – a disembodied system of symbols that points to meaning but lacks genuine comprehension:

> "The unity of sign and meaning, particularity and generality in the symbol, contrasts with the concept of allegory, which is based on the separation of sign and meaning. Given that modern society is based on the principle of separation, it is tempting to view allegory as the modern form par excellence and the symbol as merely the remainder of a premodern attitude. The persuasiveness of this categorization, however, is based on an abstract classification of thinking. If, on the other hand, one engages with the movement of form, one cannot remain with the simple juxtaposition of (modern) allegory and (premodern) symbol but must investigate their relationship as a whole." (Habermas 1989: 90).

Seen through this lens, one of the most striking characteristics of LLM outputs is their allegorical nature, particularly in their handling of verbs of motion. As we have seen, LLMs produce linguistic structures that seem meaningful but remain unmoored from embodied knowledge. Therefore, command definition by a user is defined by the profound disconnection between form and content – this same disconnect is evident in LLM-generated text. So we see that allegory becomes the process by which LLMs create representations of the world that are statistically grounded but conceptually "unanchored." For instance, when an LLM generates sentences like "the snake slithered away," it is relying on statistical patterns rather than an understanding of the biological and physical properties of snakes. The model can replicate such patterns because they are abundant in its training data, but this replication is merely an allegorical gesture. The LLM does not access the embodied knowledge

the particular event configuration suggested by the prompt sentence (*model collapse*, that is the generation of repetitive, one-sided, or incoherent output as a result of disproportionate data points in the training data, see *Federal Bureau of Information Security* 2023: 10).[5] This points to one revealing characteristic of LLM output, namely its allegorical nature. As noted by Adorno, allegory is marked by a disconnection between form and meaning. In this sense, LLMs generate linguistic structures that seem meaningful at the surface level but lack the cognitive grounding necessary for true comprehension. This disconnect becomes particularly evident in the model's handling of verbs of motion, such as "slither away," where physicality is implied, but the underlying semantic understanding of what it means to slither is absent. Thus, sentences (1a,b,c) focus on the intrinsic properties of a type of movement associated with specific body shape. Samples (1d,e,f,i) introduce metaphorical expansion that, nonetheless, can be hardly licensed by embodied world knowledge, that is, they defeat the purpose of using the verb as a means to describe the behavior of a specific animated object or entity in the reported event (the transformer introduced the hedging verb "seem" in cases where the usage of the verb "slither away" was not warranted by the prompt sentence). In other words, the semantics of the verb in the generated sentence is not expanded but constrained by the prompt in the model's representational space. Interestingly, in the case of (1i), the creative use of "slither away" results from the introduction of the word "tendrils," which creates a layer of animacy for an atmospheric phenomenon (the fog) that aligns with the form+way of locomotion real-world bias. Likewise, sample (1g) properly captures the event in which "active gripping" is counteracted by lateral undulation as an element encoded in the slithering action performed by the fish; however, sample (1h) fails to create a plausible event, since the adverb "effortlessly" cannot be freely associated with an escaping action performed by a slug (remember that ChatGPT predicts verb usage by

---

that informs human language, nor does it comprehend the real-world implications of the word "slither" beyond its linguistic context. As recent research suggests (e.g., Weissweiler et al. 2023a,b,c), this reliance on disembodied statistical predictions points to a fundamental limitation of LLMs, which operate through form but without grounding in true world knowledge.

5 Model collapse is one key aspect of LLM's limited language generation loop during semiosis. In particular, we can see in this limitation of data processing an intrinsic lack of teleological quality, that is, a failing sense of movement toward the "ultimate," the *telos*, considered here as the accomplished reading of an event as grasped through the senses and ultimately construed as experience. Since LLMs cannot be both Interpretants within and outside semiosis, it is unlikely that their representations can be further expanded to a continuous process of meaning reconstruction in events defined by world knowledge and embedded experience, In LLM training, this basically means that a LLM being trained by data provided by another LLM fails to recognize the very training task, as no semiosis can be identified from data retrieved from the introduction of improbable sequences generated by the models themselves (see Shumailov et al. 2024).

paying attention to *–ly* adverbs placed before the target verb); finally, in (1j), the model creates a conflict between the event depicted and human world-knowledge by attributing a "*slithering action*" to an unsuitable agent (insects actually "crawl," which presupposes a "limbed" type of locomotion).

So it is evident that the disembodied, unconstrained statistical prediction of pretrained language generators cannot emulate the human need to mobilize world knowledge as a means to provide accurate, context-oriented representations of the world. Furthermore, current architectures and learning setups are still poorly equipped to creatively construct sentences that reflect the actual intention of an agent engaging in a type of communication not activated by a prompt. As we have seen in the sentences provided by the Language Model, language generation and semantic understanding provide graded semiotic access to meaning.

5 Construction grammars and pretrained transformers

Despite challenges, research persists in applying linguistic methodologies to analyze language generation by *Large Language Models* (LLMs). Notably, usage-based approaches such as construction grammars (Cappelle 2024; Goldberg 2006) have garnered attention in this context with efforts being made to bridge LLM development with linguistic inquiry (Beuls and Van Aecke 2025; Weissweiler et al. 2023a,b,c). However, incorporating construction grammar predictions into LLM training remains problematic. One main reason is that, since its inception by Fillmore et al. (1988), construction grammars have struggled to establish themselves as a coherent and ontologically robust discipline, particularly in contesting the Chomskyan linguistic paradigm. However, despite offering intriguing insights into linguistic nuance (Brisard 2023), construction grammars have generally fallen short in providing a comprehensive understanding of language and language usage. As a result, linguistic representation is often reduced to questionable simplifications, due to an over reliance on statistical learning methods (e.g., Goldberg 1995, 2006, 2019). The dearth of substantive insights into the cognitive essence of constructions (see Silvennoinen 2023, for a discussion) has relegated construction grammars to the realm of speculative constructs (Samuel 2020). This is particularly evident in areas where the application of construction grammar's conceptual tools fails to substantiate robust assertions. For instance, endeavors to reduce modal meaning to formulas identified in corpora, aimed at simplifying explanatory complexity, have proven inadequate (e.g., Depraetere et al. 2023). Moreover, the lack of cohesive theoretical

underpinnings within the constructionist community has given rise to the formulation of arbitrary "principles" (e.g., Leclercq and Morin 2023).

As discussed in Section 3, the cognitive dimension of language arises from recognizing that linguistic decisions are driven by the imperative of biological open systems to uphold equilibrium between Self-as-symbol and the environmental dynamics (Torres-Martínez 2024c). This implies that the entire framework of propositions defining human–world interactions should be organized into a hierarchy of potential models of the world, each associated with specific linguistic content. In this context, a proposition featuring a natural kind, such as a tiger, should refrain from treating the essential nature of tigerness as an intrinsic property, without further considerations regarding the potential relationship of this concept with other natural kinds that may be part of the proposition's content. Consequently, the hunting behavior of a tiger (encompassing stalking, ambush hunting, or swimming), irrespective of the feline's physical characteristics such as body configuration, fur color, and predatory instincts, should also be described in terms of the bodily affordances presented by its prey that define its being in the same eco-niche.

Therefore, the usage-based claim that there is absolutely no trace of evolution or biological *persistence* in the construction of language is untenable.[6] As we have seen, the eons of experience of our species on the planet do not become accessible to us simply through statistical bias. Although languages are not innate *per se*, the mechanisms for concept formation and manipulation are innately given and further refined by experience and agency. On this reading, language evolution does not occur within the heads of individuals through semiotic mutations within the system, surfacing in anecdotal exchanges without other motivation than "sneezing foam off our cappuccinos," but as a manifestation of larger representational events that summon the whole host of experiences that define us humans. The key element here is phylogeny, which can hardly be considered as a rhetorical move for the positioning of an idealized speaker.

6 The term "persistence" refers to the stability and dependability of our cognitive functionality, which is crucial for survival and adaptation. In this view, language is one of the key cognitive tools that support this stability. On this reading, human cognition, and especially language, functions in a way that ensures the continuous ability to interpret, interact with, and respond to our environment effectively. Language, in this sense, is not just a means of communication but a core cognitive mechanism that allows for predictive processing – the ability to anticipate and understand events, behaviors, and contexts. This capacity to predict and interpret the world in symbolic terms is part of what sustains our cognitive functions, ensuring their "persistence" over time. Hence, unlike artificial models, which lack true experiential grounding, human cognition, through language, maintains this continuity by constantly adapting to new experiences, perceptions, and conceptual frameworks. Otherwise stated, "persistence" refers to the cognitive system's robustness and adaptability, enabling language to continuously function as a tool that both reflects and shapes human thought, aiding survival through the minimization of uncertainty in a complex world.

Indeed, although it is evident that both construction grammars and AI research depart from a reduced view of human experience, pretrained transformers do not need constructions to generate language (see Weissweiler et al. 2023b). Notwithstanding, it cannot go unnoticed that both construction grammars and LM research confine world knowledge to the positing of propositions through which real-world phenomena are tokenized and resemiotized according to predefined rules of communication, which disregards conceptual groundedness (the process of ecological concept formation through language), embedding (the embodied, context-driven reconstruction of an event in context), and situatedness (eventfulness). As we have seen, this offers an inaccurate reading of language as a tool for the positioning of speaker intention. As a result, even when construction grammarians acknowledge the role of agency in the emergence of constructions (Goldberg 2019), it is only to reinforce the idea that it is the linguistic system, rather than its purpose which defines the roles of speakers during communication.

6 Conclusions

In times where linguistics is increasingly compelled to embrace the algorithmic view of language, through the reduction of human representation to data structures for the acquisition of theoretical foothold (as seen in the development of computational construction grammar, Beuls and Aecke 2025), it is crucial for researchers in linguistics to envision more accurate models of human representation through language. In this sense, human linguistic data cannot be reduced to instances of LLM performance, such as perplexity values. Another point worth considering here is whether training a transformer to recognize specific constructions (e.g., caused motion constructions, Weissweiler et al. 2023c) may at some point help buttress the theoretical plausibility of the notion of form-function pairings in general. Since argument structure constructions are abstract syntactic patterns that reflect general trends in the human construction of world knowledge, it is of little relevance for linguists to explore how "peripheral" constructional content (such as caused motion constructions) could at some point confirm the existence of constructions as ontological objects contributing to the optimization of LLM performance (e.g., Weissweiler et al. 2023c).

As we have seen, a focus on the formal properties a language has in virtue of what it is in isolation has led to confusion regarding the role of natural language as a model for the development of pretrained transformers. Certainly, language has a number of properties that are not purely "linguistic," but both biological and physical. Moreover, performance and computational efficiency are not intrinsic properties of language. If they were, computers and humans would be

indistinguishable, as their sole function would be to combine symbols to externalize logical propositions about the state of the world. This stresses the underlying interconnectedness between different perceptual scales and the potential transference of properties across these scales. Among other things, it shows that a trait deemed essential for an agent's survival at one scale does not negate its status as an intrinsic property defining its way of being at a higher level, where the value of condition S* is influenced by a modified set of relations. So, if intrinsic properties were to be had in language production/generation, these would be dependent on the structure of the experiential parcel defined by an event and not a particular instance of formally interpreted linguistic performance. As I have indicated previously, the fact that natural language is a tool to (re)construct experience for purposes other than language itself shows that linguistic intrinsicality is defined by agentive imperatives at every step of the process, and that what is intrinsic about concept formation is not what we make of natural kinds or the objects in the world, but our ability to distinguish beliefs from facts. In contrast, pretrained transformers are not conscious agents, but performative ones, the nature of their responses being restricted by their capacity to identify efficiently the content of a prompt. Thus, while transformers exhibit superior computational capacity, humans excel at purposeful biological adjustment and optimization.

One of the lessons to be learned from the discussion introduced in this paper is that LLMs' linguistic capabilities do not say anything substantial about the way we think. Indeed, in terms of chatbot–human interaction, we should refrain from suggesting that transformers are in possession of *concepts about concepts*, especially regarding human metalinguistic knowledge. As already noted, transformers are trained and supervised by specific human agents according to specific cultural, racial, and corporate criteria that condition what these "stochastic parrots" can or cannot say. Moreover, the depletion of high-quality linguistic content to train LLMs by 2024 (Maslej et al. 2024: 52) announces an increment in data reuse in the form of synthetic data (LLM-generated text to train another LLM), which points to increased levels of AI language essentialism. Therefore, taking bot essentialism is often the result of *automation bias*, that is, excessive credibility on the part of users due to the bot's correctness and convincing text output (see *Federal Bureau of Information Security* 2023: 10). Consequently, automation bias can be defined as a cognitive process by which users of chatbots ascribe sentience and epistemic value to the responses of a language model in the belief that these are actual agents possessing awareness of nuance and a sense of appropriateness during a communicative event:

> On a nonacademic level, ChatGPT may *teach users various (meta)linguistic lessons*, ranging from basic language use (in terms of semantics and syntax, to start with) to more subtle and nuanced aspects, involving for instance politeness" (Dynel 2023: 122; my emphasis).

In the above passage, the author attributes *essential properties* to the chatbot by suggesting that its adjustments through training and enhanced processing capabilities create a form of ontological dependence (inherent in the linguistic interactions between human agents) between the user's needs and the prompt-driven event governing the generation of "polite" language. As previously discussed, propositions that lack gradation in describing events with agential participants oversimplify the hierarchy of those events through binary predictions of outcomes and relations (e.g., Cantlon and Piantadosi 2024). For instance, Dynel (2023: 122) states: "Since ChatGPT can adapt its responses to match various discourse styles and registers, users can witness and understand how the model adjusts its language based on the input provided." As we can see, the author makes a prediction that, nevertheless, imbues the machine with psychological content to support the assertion that chatbots have a sense of language in their interaction with humans. However, *politeness* is not an intrinsic property of the chatbot activated by its interaction with a human, but a feature introduced during transformer training by a human individual with specific cultural biases and interests (thereby resulting in statistical bias, a way of behaving). An essential aspect to consider regarding human perception of AI-generated language is the oversight of the layers of abstraction inherent in computational artifacts. These layers encompass, among others, the designer's intentions for the solution formulated to address a given problem. It is crucial to recognize that displaying politeness in behavior (the "*how*") does not inherently equate to possessing an intrinsic quality of being polite (the "*what*"). Consequently, suggesting a naturalistic alignment of intentions involving meaning negotiation and pragmatic adjustment becomes implausible when only one participant, the human agent, possesses awareness of their conscious states enabling them to plan beyond the immediate chatting event, which is determined by the prompt imposed by the human.

A misinterpretation of this fact has led to hasty conclusions regarding the metalinguistic and metapragmatic affordances of chatbots, which can only be interpreted as default responses vis-a-vis users' misuse of the tool, often for entertainment purposes. In fact, the statement "as an AI language model," used as a means to introduce a disclaimer regarding the limits of its training or its world knowledge (Dynel 2023), can be hardly interpreted as an indication of selfhood awareness for the production of context appropriate discourse during its interaction with a user. Regarding the nature of user evaluation of the chatbot performance, it is also a stretch to assert that most users are equipped with the appropriate metapragmatic, metalinguistic, or metadiscursive tools to assess the bot's competence beyond their own disruptive behavior in online interactions. This is of utmost importance when we acknowledge that the patterns of interaction among human agents on social media and chats are marked by an undue reaffirmation of their own beliefs

and prejudices through destabilizing and tendentious language usage. Since these agents are always seeking to vet gratuitously other people's capacities, knowledge base, and reactions, it is not an insight to assert that this proclivity could be extended to human–chatbot interaction as well. This "chatbot dummy effect" (my coinage), that is, the tendency to engage in abusive language use during interactions with chatbots, also brings to the fore the various ways linguistic (mis)use can be exerted in the near future on robots. A point worth considering here pertains to the appropriate distinction between terms and their application to describe particular interactions mediated by language.

Since transformers exhibit different representational structures, which can be aptly dubbed as a distinct computational semiosis possessing a surface human-like structure, but lacking experiential layers anchored in the physical world, the emergence of new types of conceptualizers that may eventually define new forms of interaction among biological and artificial agents possessing a symbolic contour and goals requires us to redefine our roles in an increasingly complex environment where our evolutionary success can no longer be taken for granted. The semiotic Self entails the reemergence of the human, a being that, in order to claim their humanity, does not need to "exclude [their] own strangeness or animality" (Zalloua 2021: 4).

References

Barron, Andrew B., Marta Halina & Colin Klein. 2023. Transitions in cognitive evolution. *Proceedings of the Royal Society. B* 290. 20230671.

Beuls, Katrien & Paul van Aecke. 2025. Construction grammar and artificial intelligence. In Mirjam Fried and Kiki Nikiforidou (eds.), *Preprint. To appear in the Cambridge Handbook of construction grammar*. Available at: arXiv.2309.00135

Birch, Jonathan, Simona Ginsburg & Jablonka Eva. 2020. Unlimited associative learning and the origins of consciousness: A primer and some predictions. *Biology and Philosophy* 3556.

Braidotti, Rosi. 2013. *The posthuman*. Cambridge, UK: Polity Press.

Brisard, Frank. 2023. Spectacle and sensationalism in construction grammar. *Constructions* 15(1). https://doi.org/10.24338/cons-536.

Cantlon, Jessica F. & Steven T. Piantadosi. 2024. Uniquely human intelligence arose from expanded information capacity. *Nature Reviews Psychology* 3. 275–293.

Cappelle, Bert. 2024. *Can construction grammar be proven wrong?* Cambridge: Cambridge University Press.

Clark, Andy. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36(3). 181–204.

Clarke, Bruce. 2008. *Posthuman metamorphosis: Narrative and systems*. New York: Fordham University Press.

Collins, Allan & M. & Ross Quillian. 1969. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior* 8(2). 240–247.

Coole, Diana & Samantha Frost (eds.). 2010. Introducing new materialisms, *New materialism: Ontology, agency and politics*, 1–43. Durham: Duke University Press.

Demuro, Eugenia & Laura Gurney. 2024. Artificial intelligence and the ethnographic encounter: Transhuman language ontologies, or what it means "to write like a human, think like a machine". *Language & Communication* 96. 1–12.

Denby, David. 2014. Essence and intrinsicality. In Robert Francescotti (ed.), *Companion to intrinsic properties*, 87–110. Berlin/Boston: Walter DeGruyter.

Depraetere, Ilse, Bert Cappelle & Martin Hilpert. 2023. Introduction. In Ilse Depraetere, Bert Cappelle, Martin Hilpert, Ludovic De Cuypere, Mathieu Dehouck, Pascal Denis, Susanne Flach, Natalia Grabar, Cyril Grandin, Thierry Hamon, Clemens Hufeld, Benoît Leclercq & Hans-Jörg Schmid (eds.), *Models of modals: From pragmatics and corpus linguistics to machine learning*, 1–13. Berlin-Boston: De Gruyter Mouton.

Dynel, Marta. 2023. Lessons in linguistics with ChatGPT: Metapragmatics, metacommunication, metadiscourse and metalanguage in human-AI interactions. *Language & Communication* 93. 107–124.

Federal Bureau of Information Security. 2023. *Generative AI models: Opportunities and risks for industry and authorities*. Available at: https://www.bsi.bund.de.

Fedorenko, Evelina, Anna Ivanova & Tamar Regev. 2024. The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience* 5. 289–312.

Fillmore, Charles J., Mary Catherine O'Connor & M. C. O'Connor. 1988. Regularity and idiomaticity in grammatical constructions, the case of let alone. *Language* 64(3). 501–538.

Frieder, Simon, Luca Pinchetti, Alexis Chevalier, Ryan-Rhys Griffiths, Tommaso Salvatori, Thomas Lukasiewicz, Philipp Christian Petersen & Julius Berner. 2023. Mathematical capabilities of ChatGPT. *arXiv preprint arXiv:2301.13867*. This is a preprint with no assigned issue, or page number.

Friston, Karl. 2005. A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 360(1456). 815–836.

Friston, Karl. 2009. The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences* 13(7). 293–301.

Friston, Karl. 2010. The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* 11(2). 127–138.

Gibson, James J. 1969. *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

Gibson, James J. 1977. *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goldberg, Adele E. 1995. *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

Goldberg, Adele E. 2006. *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldberg, Adele. E. 2019. *Explain me this: Creativity, competition, and the partial productivity of constructions*. Princeton, NJ: Princeton University Press.

Günther, Fritz, Luca Rinaldi & Marco Marelli. 2019. Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science* 14(6). 1006–1033.

Habermas, Jürgen. 1989. *Zwischenbetrachtungen im Prozeß der Aufklärung*. Frankfurt am Main: Suhrkamp Verlag.

Hagendorff, Thilo, Sarah Fabi & Michal Kosinski. 2023. Human-like intuitive behavior and reasoning biases emerged in large language models but disappeared in ChatGPT. *Nature Computational Science* 3. 833–838.

Hinton, G., J. McClelland & D. Rumelhart. 1986. Distributed representations. In D. E. Rumelhart & J. L. McClelland (eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Foundations*, Vol. 1, 77–109. Cambridge, Massachusetts: The MIT Press.

Imperial, Joseph M., Gail Forey & Harish Tayyar Madabushi. 2024. Standardize: Aligning language models with expert-defined standards for content generation. *arXiv:2402*. 12593.

Kim, Arum & Yushu Su. 2024. How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System* 122. 103256.

Kindermann, Dirk & Andrea Onofri. 2021. The fragmented mind: An introduction. In Cristina Borgoni, Dirk Kindermann & Andrea Onofri (eds.), *The fragmented mind*, 1–36. Oxford: Oxford University Press.

Landauer, Thomas K. & Susan T. Dumais. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review* 104. 211–240.

Lappin, Shalom. 2024. Assessing the strengths and weaknesses of large language models. *Journal of Logic, Language and Information* 33. 9–20.

Leclercq, Benoit & Cameron Morin. 2023. No equivalence: A new principle of no synonymy. *Constructions* 15(1). https://doi.org/10.24338/cons-535.

Marshall, Dan & Brian Weatherson. 2002. *Intrinsic vs. extrinsic properties*. Stanford Encyclopedia of Philosophy. Available at: https://plato.stanford.edu/entries/intrinsic-extrinsic/.

Maslej, Nestor, Loredana Fattorini, Raymond Perrault, Vanessa Parli, Anka Reuel, Erik Brynjolfsson, John Etchemendy, Katrina Ligett, Terah Lyons, James Manyika, Juan Carlos Niebles, Yoav Shoham, Wald Russell & Clark Jack. 2024. *The AI index 2024 annual report*. Stanford, CA: AI Index Steering Committee, Institute for Human-Centered AI, Stanford University.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado & Dean Jeffrey. 2013. Distributed representations of words and phrases and their compositionality. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani & K. Q. Weinberger (eds.), *Advances in neural information processing systems (NIPS)*, vol. 26, 3111–3119. Curran Associates.

Nov, Oded, Nina Singh & Devin Mann. 2023. Putting ChatGPT's medical advice to the (Turing) test. *arXiv*. https://doi.org/10.48550/arXiv.2301.10035.

Pearl, Judea. 1988. *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. San Mateo: Morgan Kaufmann Publishers.

Pepperell, Robert. 2003. *The posthuman condition*. Portland, OR: Intellect Books.

Rosendahl Thomsen, Mads. 2013. *The new human in literature: Posthuman visions of changes in body, mind and society after 1900*. London/New York: Bloomsbury Academic.

Samuel, Arthur G. 2020. Psycholinguists should resist the allure of linguistic units as perceptual units. *Journal of Memory and Language* 111. 104070.

Shumailov, Ilia, Zakhar Shumaylov, Yiren Zhao, Nicolas Papernot, Ross Anderson & Yarin Gal. 2024. AI models collapse when trained on recursively generated data. *Nature* 631. 755–759.

Silvennoinen, Olli O.. 2023. Is construction grammar cognitive? *Constructions* 15(1). https://doi.org/10.24338/cons-544.

Solaki, Anthia. 2022. The effort of reasoning: Modelling the inference steps of boundedly rational agents. *Journal of Logic, Language and Information* 31. 529–553.

Søgaard, Anders. 2024. Grounding the vector space of an octopus: Word meaning from raw text. *Minds and Machines* 33. 33–54.

Strachan, James W. A., Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, Michael S. A. Graziano & Cristina Becchio. 2023. Testing theory of mind in large language models and humans. *Nature Human Behaviour* 8. 1285–1295.

Torres-Martínez, Sergio. 2018a. Constructions as triads of form, function and agency: An agentive cognitive construction grammar analysis of English modals. *Cognitive Semantics* 4(1). 1–38.

Torres-Martínez, Sergio. 2018b. Exploring attachment patterns between multi-word verbs and argument structure constructions. *Lingua* 209. 21–43.

Torres-Martínez, Sergio. 2019. Taming English modals: How a construction grammar approach helps to understand modal verbs. *English Today* 35(2). 50–57.
Torres-Martínez, Sergio. 2020. On English modals, embodiment and argument structure: Response to Fong. *English Today* 38(2). 105–113.
Torres-Martínez, Sergio. 2021a. The cognition of caused-motion events in Spanish and German: An agentive cognitive construction grammar analysis. *Australian Journal of Linguistics* 41(1). 33–65.
Torres-Martínez, Sergio. 2021b. Complexes, rule-following, and language games: Wittgenstein's philosophical method and its relevance to semiotics. *Semiotica* 242. 63–100.
Torres-Martínez, Sergio. 2022a. Metaphors are embodied otherwise they would not be metaphors. *Linguistics Vanguard* 8(1). 185–196.
Torres-Martínez, Sergio. 2022b. The role of semiotics in the unification of Langue and Parole: An agentive cognitive construction grammar approach to English modals. *Semiotica* 244(1/4). 195–225.
Torres-Martínez, Sergio. 2022c. On the cognitive dimension of metaphors and their role in education: A response to Molina Rodelo (2021). *Revista Senderos Pedagógicos* 13. 113–123.
Torres-Martínez, Sergio. 2023a. A radical embodied characterization of German Modals. *Cognitive Semantics* 9(1). 132–168.
Torres-Martínez, Sergio. 2023b. The semiotics of motion encoding in early English: A cognitive semiotic analysis of phrasal verbs in old and middle English. *Semiotica* 251. 55–91.
Torres-Martínez, Sergio. 2023c. *Grammaire agentielle cognitive de constructions* : Explorations sémiotico-linguistiques des origines de la représentation incarnée. *Signata, Annales de Sémiotique* 14. https://doi.org/10.4000/signata.4551.
Torres-Martínez, Sergio. 2024a. Embodied human language models vs. large language models, or why Artificial Intelligence cannot explain the modal *be able to*. *Biosemiotics* 17. 185–209.
Torres-Martínez, Sergio. 2024b. Embodied essentialism in the reconstruction of *the animal sign* in robot animal design. *Biosystems* 238. 105178.
Torres-Martínez, Sergio. 2024c. Semiosic translation: A Bayesian-heuristic theory of translation and translating. *Language and Semiotic Studies* 10(2). 167–202.
Tuckute, Greta, Nancy Kanwisher & Evelina Fedorenko. 2024. Language in brains, minds and machines. *Annual Review of Neuroscience* 47. 277–301.
Vallentyne, Peter. 2014. Intrinsic properties defined. In Robert M. Francescotti (ed.), *Companion to intrinsic properties*, 31–40. Berlin/Boston: Walter DeGruyter.
Watson, Richard. 2023. Agency, goal-directed behavior, and part-whole relationships in biological systems. *Biological Theory* 19. 22–36.
Weissweiler, Leonie, Taiqi He, Naoki Otani, David R. Mortensen, Lori Levin & Hinrich Schütze. 2023a. Construction grammar provides unique insight into neural language models. In *Proceedings of the first international workshop on construction grammars and NLP (CxGs+NLP, GURT/SyntaxFest 2023)*, 85–95. Washington, D.C.: Association for Computational Linguistics. Available at: https://aclanthology.org/2023.cxgsnlp-1.10/.
Weissweiler, Leonie, Valentin Hofmann, Anjali Kantharuban, Anna Cai, Ritam Dutt, Amey Hengle, Anubha Kabra, Atharva Kulkarni, Abhishek Vijayakumar, Haofei Yu, Hinrich Schütze, Kemal Oflazer & David R. Mortensen. 2023b. Counting the bugs in ChatGPT's wugs: A multilingual investigation into the morphological capabilities of a large language model. In *Proceedings of the 2023 conference on empirical methods in natural language processing*, 6508–6524. Singapore: Association for Computational Linguistics.
Weissweiler, Leonie, Abdullatif Köksal & Hinrich Schütze. 2023c. Hybrid Human-LLM corpus construction and LLM evaluation for rare linguistic phenomena. *arXiv:2403*. 06965.

Wilson, Mark. 2023. *Imitation of rigor: An alternative history of analytic philosophy*. Oxford: Oxford University Press.
Yax, Nicolas, Hernán Anlló & Stefano Palminteri. 2024. Studying and improving reasoning in humans and machines. *Communications Psychology* 2(51). 51.
Youssef, Alaa, Samantha Stein, Justin Clapp & David Magnus. 2023. The importance of understanding language in large language models. *The American Journal of Bioethics* 23(10). 6–7.
Zalloua, Zahi. 2021. *Being posthuman: Ontologies of the future*. London/New York: Bloomsbury Academic.

Bionote

Sergio Torres-Martínez
Universidad de Antioquia, Cll. 67 #53 – 108, Medellín, Antioquia, Colombia
surtr_2000@yahoo.es
https://orcid.org/0000-0002-8823-1676

Professor Sergio Torres-Martínez investigates how language functions as a cognitive tool shaping human agency and meaning-making. As a cognitive linguist and semiotician, he examines language through the lens of embodied cognition – the theory that our physical experiences fundamentally structure our mental processes. His current research aims to bridge cognitive linguistics, semiotics, and the philosophy of mind as a means to explore how linguistic representation enables humans to reduce environmental uncertainty and expand their capacity for purposeful action.