

Diego A. Jiménez Sennrich

# Chomsky & Large Language Models: A response to Piantadosi (2023)

---

**Resumen:** *En un artículo reciente, Steven T. Piantadosi propone que «los modelos lingüísticos modernos refutan el enfoque de Chomsky acerca del lenguaje». El presente artículo pretende mostrar que el argumento de Piantadosi se basa en una mala comprensión tanto del enfoque de Chomsky sobre el lenguaje como, más ampliamente, de las restricciones a la adquisición del lenguaje a las que deberían atenerse las teorías lingüísticas serias. Como resultado, se muestra que bajo consideraciones de elección de teoría racional, el enfoque de Chomsky triunfa sobre el de Piantadosi.*

**Palabras clave:** *Grandes modelos de lenguaje, lingüística chomskiana, Programa minimanlista biolingüística, filosofía de la lingüística*

**Abstract:** *In a recent article, Steven T. Piantadosi proposes that «Modern language models refute Chomsky's approach to language». The present paper intends to show that Piantadosi's argument is based on a misunderstanding of both Chomsky's approach to language and, more broadly, language acquisition restrictions serious linguistic theories ought to commit to. As a result, it is shown that under rational theory choice considerations Chomsky's approach trumps Piantadosi's.*

**Keywords:** *Large Language Models, Chomskyan linguistics, Minimalist Program, Biolinguistics, Philosophy of linguistics*

## 0. Introduction

In a recent article, Steven T. Piantadosi argues that Large Language Models (LLMs) refute Chomskyan linguistics (CL). We posit that Piantadosi's argument is based on a misunderstanding of both CL and, more generally, language acquisition restrictions serious linguistics theories ought to commit to.

In section 1, we propose an interpretation of CL, more precisely, of the Basic Property of language, Merge, and the syntax-semantics relation. In section 2, we begin by highlighting the epistemic strengths of CL, according to rational theory choice (RTC)<sup>1</sup>. We then go on to present Chomsky's normative schema for any Learning Theory (LT, where any theory of language acquisitionLT). Following Chomsky's LT schema, we present a set of commitments any serious LT of language ought to commit to. In addition, we show that CL succeeds in doing so, whereas LLMs fail. We close section 2 with an example of interdisciplinary research into the Faculty of language based on CL principles. In section 3, we explain LLM pre-training and how it differs, substantially, from the linguistic experience of



any human language user. Finally, in section 4, we present Piantadosi's relevant arguments, and show that they indeed fail on the accounts we presented above. As result, we conclude that CL remains a stronger theory.

## 1. Chomskyan linguistics (CL)<sup>2</sup>

The theoretical commitments of CL are clearly stated in Berwick and Chomsky (2017), in the form of the Basic Property of language (BPL). Thus, (Ai)-(Aiii) holds true for any language, L.

- (A) i. L is a finite computational system.  
 ii. L yields an infinity of expressions.  
 iii. each expression of L has a definite interpretation in semantic-pragmatic and sensorimotor systems (Berwick & Chomsky 2017, 1).

BPL ought to be understood within the broader framework proposed by Hauser et al. (2002), according to which the vague concept «Faculty of language» is delineated into the concepts «Faculty of language—broad sense (FLB)» and «Faculty of language—narrow sense. (FLN)»

... FLB includes an internal computational system (FLN, below) combined with at least two other organism-internal systems, which we call «sensory-motor» and «conceptual-intentional». ... FLN is the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces. FLN is a component of FLB, and the mechanisms underlying it are some subset of those underlying FLB (1570-1571).

Both (Ai) and FLN are to be identified, in CL, with *Merge*.

Applied to two objects  $\alpha$  and  $\beta$ , Merge forms the new object K, eliminating  $\alpha$  and  $\beta$ . What is K? K must be constituted somehow from the two items  $\alpha$  and  $\beta$  ... The simplest object constructed from  $\alpha$  and  $\beta$  is the set  $\{\alpha, \beta\}$ , so we take K to involve at least this set, where

$\alpha$  and  $\beta$  the constituents of K. Does that suffice? Output conditions dictate otherwise; thus, verbal and nominal elements are interpreted differently at LF and behave differently in the phonological component. K must therefore at least (and we assume at most) be of the form  $\{\gamma, \{\alpha, \beta\}\}$ , where  $\gamma$  identifies the type to which K belongs, indicating its relevant properties. Call  $\gamma$  the *label* of K (Chomsky 1995, 243).

The former explanans is quite abstract. An example, though oversimplified, may help to elucidate the *explanandum*, «Merge». Given a lexicon consisting only of the lexical items {«John» and «runs»}, an instance of the operation Merge may yield the syntactic object  $K = \{\text{VP} \{\text{John, runs}\}\}$ , where «runs» is picked out as the projecting element, or *label*, of K, indicating the type of K, in this instance, VP, for Verb Phrase. It should be noted that i)  $K = \text{a bare phrase structure}$ , and ii) given that K is in a state  $S_1$ , it is the case that  $S_1$  holds for K in virtue of the interaction between Merge and the information specified for the lexical items of K by the Lexicon, namely, the category of the lexical items (this information being notoriously semantic).

That (Aii) holds true for any L is clear once the iterative property of Merge is shown. So, given a Lexicon consisting only of the lexical items {«Mary», «John», «knows» and «that»}, Merge may yield, after sufficient applications  $K_0 = \{\text{VP} \{\text{Mary, knows, that}\}\}$ . Another application of Merge, this time applying to  $K_0$ , may yield  $K_1 = \{\text{VP} \{\text{VP} \{\text{Mary, knows, that}\}\}, \{\text{VP} \{\text{John, knows, that}\}\}\}$ . A new syntactic object,  $K_2$ , may be produced in the obvious manner. More generally, any number of syntactic objects,  $K_n$ , may be produced.

Additionally, *chomskyano sensu, labelling* accounts for the interpretation of K at LF and PF, or, to use the terminology of Hauser et al. (2002), *labelling* accounts for the interpretation of K at the interface levels with the conceptual-intentional and sensorimotor systems, respectively.

Below, we formalize what we consider to be the most important elements implicated by our interpretation of CL thus far.

(B)

- i. For all (x), if Merge admits an element (x), then (x) has syntactic information.
- ii. Merge only admits elements of the Lexicon, i.e., lexical items.
- iii. Thus, lexical items have syntactic information. {i, ii, MPP}
- iv. For all (x), if (x) is an element of the Lexicon, then (x) has semantic information.
- v. Thus, lexical items have semantic information. {iv, ii, MPP}
- vi. Thus, lexical items have syntactic and semantic information. {iii, v, CONJ}
- vii. For all (y), if (y) is a bare phrase structure, (y) is composed of lexical items.
- viii. For all (y), if (y) is composed of lexical items, (y) has semantic information.
- ix. (y) is a bare phrase structure.
- x. (y) is composed of lexical items. {vii, ix, MPP}
- xi. (y) has semantic information. {viii, ix, MPP}

## 2. The case for CL

In addition to self-consistency,

- (C)
- i. CL explains linguistic phenomena other linguistic theories do not, such as *particle movement*<sup>3</sup>.
  - ii. CL fits the data of the critical period (CP) (and poverty of stimulus (PS)) of language acquisition (LA).
  - iii. Neurologists have been able to design fruitful experiments that add empirical confirmation to some fundamental assumptions of CL.

(Ci)-(Ciii) Are not equipollent. Success in (Ci) is a boon, but failure in (Ci) may only indicate some reformulation of a theory of FL. (Ciii) is clearly a lofty boon (indeed, generalizing, according to Kuhn (1977) it is a *desideratum* for theory choice) but research may carry on without it. But failure in (Cii) is a sufficient condition for rejecting a theory of FL<sup>4</sup>. Why?

To pursue the study of a given  $LT(O,D)$ <sup>5</sup> in a rational way, we will proceed through the following stages of inquiry:

1. Set the cognitive domain D.

2. Determine how O characterizes data in D «pretheoretically,» thus constructing what we may call «the experience of O in D.»
3. Determine the nature of the cognitive structure attained; that is, determine, as well as possible, what is learned by O in the domain D.
4. Determine  $LT(O,D)$ , the system that relates experience to what is learned (Chomsky 1975, 14).<sup>6</sup>

An immediate consequence of the former is that any theory, LT, ought to consider the «the experience of O in D» ( $=E(O,D)$ ), lest it fail to be a  $LT(O,D)$ . Evidently, given two organisms,  $O_0$  and  $O_1$ , and a domain, D,  $LT(O_0,D)$  and  $LT(O_1,D)$  might inform each other in each of three cases,

- (D)
- i.  $E(O_0,D)=E(O_1,D)$
  - ii.  $E(O_0,D)\approx E(O_1,D)$
  - iii.  $E(O_0,D)\leftrightarrow E(O_1,D)$

Presumably, trying to delineate when exactly one of (Di)-(Diii) holds for any two  $E(O_x,D)$ ,  $E(O_y,D)$  (Such that  $x\neq y$ ) invites vagueness. At any rate, such a task falls beyond the scope of this paper. We limit ourselves to arguing that such a relation clearly does not hold between human language and LLMs.

Consider CP; Friederici (2016) writes that,

In the language domain, Lenneberg (1967) was the first to propose a maturational model of language acquisition, suggesting a critical period for language acquisition that ends around the onset of puberty. Today researchers see the closure of the critical period, especially for syntax acquisition, either at or no later than age 6, whereas some claim that the critical period of native-like syntax acquisition is even earlier, around age 3 (145).

Consider the implications. Let «language user» =  $\mathcal{S}$ . Given a couple of  $\mathcal{S}$ ;  $\mathcal{S}_0^e$  and  $\mathcal{S}_1^e$ , where «e» indicates membership in the same linguistic environment, e, and the subindex numbers indicate «antecedence», in the obvious manner, any  $LT(H, L)^7$  needs to explain the following:

- (E) i.  $\mathcal{S}_1^e$  has a finite amount of time to yield FL (more precisely, 3 or 6 years).
- ii.  $\mathcal{S}_1^e$  has a finite amount of linguistic information to yield FL (from «i»)⁸.
- iii.  $\mathcal{S}_1^e$  is exposed to decrepit linguistic information (by  $\mathcal{S}_0^e$ , and, more generally, any  $\mathcal{S}^e$  that is in contact with  $\mathcal{S}_1^e$ )⁹.
- iv.  $\mathcal{S}_1^e$  has to acquire a skill, whatever it may be, that produces an infinite output of potentially novel expressions.
- v.  $\mathcal{S}_1^e$  has to acquire a skill, whatever it may be, that is not qualitatively dissimilar from that of  $\mathcal{S}_0^e$  (and, more generally, any  $\mathcal{S}^e$ ).

Proposing an innate, finite computational system is a parsimonious solution. Proposing an analogy, or equivalence, to LLMs, we will show, is to disregard the empirical conditions of any  $\mathcal{S}$ . In other words, it is to fail even to be an  $LT(H, L)$ . Further development of this point is left to the next section.

We now turn our attention to (Ciii).

CL not only fits CP and PS the best but has empirical confirmation. So, in *Impossible Languages* (2016), Moro describes many successful experiments based on the theoretical principles of CL. We are particularly interested in his

account of Musso et al. (2003). Because of its importance, we cite a rich part of Moro's report. For the sake of brevity, we exclude all that is not immediately relevant.

A group of twelve subjects who had been exposed to only one language over a lifetime was selected. ... In this case, the twelve people spoke German. They were taught a version of micro-Italian including only a limited set of nouns, verbs, and some basic function words such as articles, particles, auxiliaries, negations, and, of course, some syntactic rules. Some of these rules were actual rules of Italian—for example, we taught them that in Italian one can form a sentence without expressing its subject, unlike German (and French and English, among others); they were also taught how to construct an embedded sentence, a construction very different from the Italian matrix sentences. Then they were taught «impossible rules»: rules with rigid dependencies based on the position of words in the linear sequence, running against the specific recursive structure implemented in human syntax. What follows are three examples. The first rule was for constructing a negative sentence and required specific positioning within the sentence: insert the word *no* as the fourth word of the string. ... The subjects, of course, were not aware that these rules were based on two different major types (recursive and linear) and started to learn how to process the rules. The experiment consisted of resting the brain's reaction at different stages in the process of learning.

The results obtained by measuring the BOLD signal with an fMRI were very clear. We concentrated on the activity in Broca's area ... This activity was checked against the subject's ability to master the new rules in the new language. ... All in all, the experiment showed that the amount of blood in Broca's area augmented when the subjects increased their ability to apply rules based on recursive architecture, whereas it diminished when the subjects increased their ability to apply rules based on linear order ... (55-56).

### 3. Pre-training and LLMs

For the purposes of our paper, we are particularly interested in LLM pre-training, in order to showcase the difference between the linguistic experience of LLMs and any  $\mathfrak{S}$ .

Pre-training exposes an LLM to a given number of words, tokens (the data set), in a given number of determinate sequences. These sequences, though determinate, are shown incomplete to the LLM (lacking, for example, the last token of the sequence). The LLM is then tasked with completing the sequence. At the beginning of the pre-training process, the LLM will output a random, wrong token to complete the sequence. It is then corrected, updated. Alammr (2020) presents a paradigmatic example. So, given the sequence «a

robot must ...», and the expected output «obey», an LLM early in the pre-training process will output instead «troll», it is then corrected. How?

We can calculate the error, the differences between these words, we have ways to put that into a numeric value. After we calculate that error, we have a way of feeding it back to the model, updating the model, so that the next time it sees «a robot must ...», it's more likely to say «obey». We do this thousands, millions, tens of millions of times, on all the data that we have, and then we have a trained model. (Alammr 2020, 7, 11)

The LLM Piantadosi refers to is GPT-3. Consider the data set of GPT-3, *Figure 1* (Brown et al. 2020, 9)

| Dataset                 | Quantity (tokens) | Weight in training mix | Epochs elapsed when training for 300B tokens |
|-------------------------|-------------------|------------------------|--|
| Common Crawl (filtered) | 410 billion       | 60%                    | 0.44   |
| WebText2                | 19 billion        | 22%                    | 2.9  |
| Books1                  | 12 billion        | 8%                     | 1.9  |
| Books2                  | 55 billion        | 8%                     | 0.43   |
| Wikipedia               | 3 billion         | 3%                     | 3.4  |

Figure 1. GPT-3's data set.

Let us assume that the example in Alammr (2020) is an E(O,D), particularly «the experience of GPT-3 in the language domain», E(G,L). Moreover, Alammr (2020) proposes an LT(G,L). We will show that Piantadosi also proposes an LT(G,L), but not, as he purports, an LT(H,L). How? We argue that E(G,L) does not attain any of (D)i-(D)iii. Why?

- (F) i. The experience of language of GPT-3 is quantitatively incommensurable with the experience of language of any  $\mathfrak{S}$ .
- ii. Indeed, it is not even clear that the cognitive apparatus of any  $\mathfrak{S}$  would be able to apprehend something like E(G,L) throughout their life, let alone in CP<sup>10</sup>.
- iii. E(G,L) is not confronted with PS, since the data set is curated.

One could ask whether LT(G,L) has something to say to any LT(H,L); but to entertain the notion that LT(G,L) refutes any LT(H,L) is preposterous.

### 4. The case against LLMs

Piantadosi (2023)'s case for LLMs boils down to four points.

- (G) i. ... they are precise and formal enough accounts to be implemented in actual computational systems, unlike most parts of generative linguistics.
- ii. ... such models are able to make predictions.

iii. *Unlike generative linguistics*, these models show promise in being integrated with what we know about other fields.

iv. ... these models are empirically tested, especially as a theory of grammar (12-13). (the emphasis is ours)

(Gi). That LLMs are so precise and formal such as to be implemented in computational systems is clear enough (whether or not they could be implemented in the mind/brain is another matter, see note 10), to contrast LLMs with generative linguistics on this account is, however, rather odd (particularly if the desired effect is consistency revision, as is the case for Piantadosi). CL has a long history of formalization and precision; see, for example Chomsky (1953, 1957, 1965, 1993, 1995); and Chomsky & Schützenberger (1963). Suppose a model for each publication,  $M_1$ - $M_6$ , each is precise and formal enough to permit revision of consistency.

(Gii). Both LT(G,L) and CL are able to make predictions. It is useful to ask, however, whether both can make predictions of *human* language.

(Giii). Like CL. See Moro (2015, 2016) and Friederici (2016) for examples of CL-based interdisciplinary research on human language. That LLMs show promise is true enough, but substantial revision of LLM-based linguistic theories is in order. As it stands, they are not an LT(H,L).<sup>11</sup>

(Giv) The main point here is that, according to Piantadosi (2023) «Approaches from generative syntax are not competitive in any domain and arguably have avoided empirical tests of their core assumptions» (13). This is plainly false, see Moro (2015) and section 2 of this paper.

In addition to his case for LLMs, Piantadosi purports to refute the following key principles of CL (we reproduce only those we consider relevant). As before, our general approach is to deny LLMs LT(H,L) status, a sufficient condition to render Piantadosi's point moot. There are, however, particular counterarguments of interest (in the sense that they are informative) for some of (Hi)-(Hvi)

- (H) i. Syntax is integrated with semantics.  
ii. Probability and information are central.

iii. Learning succeeds in an unconstrained space.

iv. Representations are representationally complex, not minimal.

v. Hierarchical structure need not be innate.

(Hi). Piantadosi (2023)'s argument is as follows.

Modern large language models integrate syntax and semantics in the underlying representations: encoding words as vectors in a high-dimensional space, without an effort to separate out e.g. part of speech categories from semantic representations, or even predict at any level of analysis other than the literal word. Part of making these models work well was in determining how to encode semantic properties into vectors, and in fact initializing word vectors via encodings of distribution semantics ... (15)

In contrast «Chomsky and others have long emphasized the study of syntax as a separate entity, not only from the rest of cognition but from the rest of language» (15). Contrast Piantadosi's argument with section 2's talk of interface levels and the schematized interpretation of CL. The point is that, while syntax is independent, FLB necessitates interface levels, bringing the syntax-semantics relation, in CL, closer than reported by Piantadosi.

Moreover, that LLMs function by integrating syntax and semantics does not entail the same for human languages. Inversely, we have good reasons to deny it. Firstly, LLMs failure to attain LT(H,L) status. Secondly, CL epistemic strength (which we have been showing throughout this paper).

(Hiii). Piantadosi (2023) argues that «... modern language models succeed despite the fact that their underlying architecture for learning is relatively unconstrained. This is a clear victory for statistical learning theories of language ...» (18). Talk of relatively unconstrained, underlying architecture might be true enough, but two points are of note. Consider, on the one hand, the sheer magnitude of data sets, and the implications we have spelled out (the

LT(G,L)-LT(H,L) distinction). On the other hand, though the underlying architecture may be unconstrained, experience might be constrained, so Brown et al. (2017, 8) talk of the curation of process of the data sets for GPT-3

... we have found that unfiltered or lightly filtered versions of Common Crawl tend to have lower quality than more curated datasets. Therefore, we took 3 steps to improve the average quality of our datasets: (1) we downloaded and filtered a version of CommonCrawl based on similarity to a range of high-quality reference corpora, (2) we performed fuzzy deduplication at the document level, within and across datasets, to prevent redundancy and preserve the integrity of our held-out validation set as an accurate measure of overfitting, and (3) we also added known high-quality reference corpora to the training mix to augment CommonCrawl and increase its diversity.

(Hiv). That Representations are representationally minimal, not complex, is not gratuitous. It is a necessary condition for positing a plausible biolinguistic explication of language. As Chomsky (2017) puts it

Generative grammar sought, for the first time, to provide explicit accounts of languages—grammars—that would explain what we call the Basic Property of language ... When this problem was first addressed the task seemed overwhelming. Linguists scrambled to construct barely adequate grammar, and the results were so complex that it was clear at the time *that they could not be evolvable* (2). (the emphasis is ours)

Computational linguists like Piantadosi are now faced with such challenges, their solution cannot be to ignore them.

(Hv). Piantadosi tells us that «These models discover structure—including hierarchical structure—from their training ... These models certainly could learn rules based on linear, rather than hierarchical, structure, but the data strongly leads them towards hierarchical, structural generalization» (21). However, Piantadosi disregards

entirely the point of linguistic inquiry, to explicate how, under CP and PS restrictions, all  $\mathcal{G}$ s attain a FL that necessitates hierarchical structures. For such explications one must turn to CL.

## 5. Concluding remarks

Piantadosi (2023) paints a false, overly optimistic picture of the implication LLMs have for linguistic inquiry. He disregards essential problems serious linguistic theories ought to contend with, such as CP and PS, and, more generally, biolinguistic considerations. As a result, we argued, Piantadosi's proposal for an LLM-based linguistic theory fails to be an LT(H,L), and, by the same token, fails to be a serious contender for a biolinguistic explication of language. Under such conditions, LLMs may not refute CL.

Conversely, confrontation of CL by LLMs gave way to accentuate the epistemic robustness of CL. We showed, throughout, that CL is self-consistent, explanatory and descriptively rich, is better suited to the empirical data, and has become the basis for a number of interdisciplinary research projects (which, in turn, have produced serious evidence in favor some of the fundamental principles of CL).

## Notes

1. Though no schema of RTC is explicitly proposed, its presence in the background is more than obvious. N. b., this is a recurring theme of the paper, and, as such, not limited to section 2.
2. CL, as we understand the term, is a proper subset of Generative linguistics. The former is chosen throughout this paper because Piantadosi's critique by and large focuses on CL.
3. For examples see Akmajian (2010).
4. This is, indeed, the main point of our response against Piantadosi.
5. LT(O,D)=«the learning theory for the organism O in the domain D» (Chomsky 1975, 14).
6. The former passage is Philosophy of linguistics, if anything is.
7. LT(H,L)=«the learning theory for humans in the language domain».

8. Let us imagine a hyperbolic scenario of a child being exposed to a word every second, all day, from the moment they are born, until they are six years old. They would be exposed, in CP, to sixty million words.
9. That is, whatever number of expressions can be made up from sixty million words (a precise number is not of much interest), most of those expressions will not be perfect exemplars of *grammatical* expressions. As experience easily attests.
10. Consider what Chomsky (2012, cited in Piantadosi 2023, 20) has to say, «we cannot seriously propose that a child learns the values of  $10^9$  parameters in a childhood lasting only  $10^8$  seconds.» *N.b.*, we cannot because of i) cognitive restrictions (which Piantadosi argues against, unsatisfactorily), and ii) empirical data about CP, and the possible number of words and sequences that fit in CP.
11. Chomsky (2023) argues for a CL interpretation of work by David Poeppel.

## References

- Akmajian, Adrian., Demers, Richard., Farmer, Ann., Harnish, Robert. 2010. *Linguistics: An Introduction to Language and Communication*. Cambridge, MA: The MIT Press.
- Alammar, Jay. «The Narrated Transformer Language Model.» Online video clip. YouTube, October 26, 2020. <https://youtu.be/-QH8fRhqFHM>
- Berwick, Robert., Chomsky, Noam. 2017. *Why only Us: Language and Evolution*. Cambridge, MA: The MIT Press.
- Brown, Tom., Mann, Benjamin., Ryder, Nick., Subbiah, Melanie, Kaplan, Jared., Dhariwal, Prafulla., Neelakantan, Arvind., Shyam, Pranav., Sastry, Girish., Askell, Amanda., et al., 2020. «Language Models are Few-Shot Learners». *Adv. Neural Inf. Process. Syst.* 33: 1877–1901. <https://arxiv.org/abs/2005.14165>
- Chomsky, Noam. 1953. «Systems of syntactic analysis», *The Journal of Symbolic Logic* 18, no. 3: 242-256.
- Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: The MIT Press.
- Chomsky, Noam. 1975. *Reflections on Language*. New York: Random House.
- Chomsky, Noam. 1993. *Lectures on Government and Binding: The Pisa Lectures*. Berlin: Mouton De Gruyter.
- Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: The MIT Press.
- Chomsky, Noam. 2023. *The Secrets of Words*. Cambridge, MA: The MIT Press.
- Chomsky, Noam., Schützenberger, Marcel-Paul. 1963. «The algebraic theory of context free languages.» In *Computer programming and formal languages*, edited by Paul Braffort., David Hirschberg, 118–161. Amsterdam: North Holland Publishing.
- Friederici, Angela. D. 2016. *Language in Our Brain: The Origins of a Uniquely Human Capacity*. Cambridge, MA: The MIT Press.
- Hauser, Marc. D., Chomsky, Noam., Fitch, W. Tecumseh. 2002. «The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?» *Science* 298, no. 5598: 1569-1579.
- Kuhn, Thomas. 1977. *The Essential Tension: Selected Studies in Scientific Tradition and Change*. Chicago: The University of Chicago Press.
- Moro, Andrea. 2008. *The Boundaries of Babel*. Cambridge, MA: The MIT Press.
- Moro, Andrea. 2016. *Impossible Languages*. Cambridge, MA: The MIT Press.
- Piantadosi, Steven. T. 2023. «Modern language models refute Chomsky’s approach to language». Ms., available at <https://ling.auf.net/lingbuzz/007180>.

**Diego A. Jiménez Sennrich** ([diego.jimenez-sennrich@ucr.ac.cr](mailto:diego.jimenez-sennrich@ucr.ac.cr)) es estudiante del Bachillerato en Filosofía en la Universidad de Costa Rica.

Recibido: 29 de septiembre, 2023.

Aprobado: 6 de octubre, 2023.