Artificial Knowledge of Language

A Linguist's Perspective on Its Nature, Origins and Use

Edited by
José-Luis Mendívil-Giró
University of Zaragoza

Series in Language and Linguistics

VERNON PRESS

www.vernonpress.com

In the Americas:
Vernon Press
1000 N West Street, Suite 1200
Wilmington, Delaware, 19801
United States

In the rest of the world:
Vernon Press
C/Sancti Espiritu 17,
Malaga, 29006
Spain

Series in Language and Linguistics

Library of Congress Control Number: 2025945088

ISBN: 979-8-8819-0379-4

Cover design by Vernon Press with elements from Freepik.

Although Markov, Turing, von Neumann, or Chomsky never got a Nobel Prize, technologists developing their foundational ideas into Large Language Models surprisingly did: in 2024. The present book delves into the (for now) state-of-the-art in this saga, presenting it as dispassionately as possible through eight prismatic chapters. If readers wonder how momentous an epilogue and beyond can be, they only need to imagine the offspring of a next generation of such devices - albeit built beyond classical presuppositions in hardware, logic, or even modeling. The latter no longer seems like a chimerical possibility; whether that is for better or for worse, surviving the future may tell.

Dr. Juan Uriagereka
Professor, Linguistics & SLLC
University of Maryland

A comprehensive and wide-ranging overview of the issues (and non-issues) raised for linguistic theory, particularly Chomskyan linguistic theory, by the development of Large Language Models. The editor's introduction is particularly useful. Highly recommended for linguists, including computational linguists, of all persuasions, as well as computer scientists, philosophers and psychologists with an interest in language and AI.

Dr. Ian Roberts
Professor of Linguistics
University of Cambridge

Table of Contents

List of figures

List of tables

Preface

The objective of this volume is to gather linguistically informed opinions (from linguists or cognitive scientists with specialization in the study of language) on the nature, origins and use of the knowledge of language developed by the Artificial Intelligence systems called Large Language Models (LLMs).

The recent development and popularization of these systems has had a significant impact on the media, particularly the most well-known and widely used system, ChatGPT (developed by the company OpenAI). Similarly, this emergence has led to the proliferation of opinions regarding the relevance of LLMs beyond the practical (and typically commercial) purposes for which they have been designed, particularly in the fields of cognitive science and linguistic theory. Thus, it is not uncommon to find on social networks, blogs and popular magazines statements such as that LLMs have solved the problems that sciences such as linguistics aim to solve, that the success of LLMs in the generation of text can be considered a refutation of some notably influential theories of language (especially Noam Chomsky's approach and generative grammar, the tradition that arose from his ideas about language and the human mind), or the claim that LLMs are in fact scientific theories of language.

These statements appear to be founded upon the premise that the linguistic knowledge acquired by these systems is analogous to, or at least comparable to, that which is developed by human beings in a naturalistic manner. The objective of this book is to evaluate the extent to which this assumption is justified.

It is important to recall that Noam Chomsky, in his 1986 book *Knowledge of Language: Its Nature, Origins, and Use* (whose title serves as inspiration for that of the present volume), argued that the pivotal shift in the science of language that emerged with generative grammar and the so-called cognitive revolution in the 1950s was the realization that the object of linguistic inquiry, as part of the broader field of natural science, was not the observable linguistic behavior of speakers or the product of their language use, as was previously assumed, but rather the knowledge of language itself.

Consequently, the questions that the chapters gathered in this volume aim to answer, all of them related to the central issue of comparing natural and artificial knowledge of language, are the following:

- To what extent are LLMs models of human language?
- To what extent can it be asserted that, for example, ChatGPT has learned English (or any other language)?
- What, if any, are the differences between the knowledge of language possessed by LLMs and that of human beings?
- In what sense are LLMs relevant to our understanding of the processes by which humans acquire and use languages?
- Is there any component of the language faculty (semantics, syntax, pragmatics, morphology, phonology) in which LLMs' knowledge is more similar to that of humans than in others?
- What uses for language research can such systems have?
- What impact will the development of LLMs have on the field of linguistic theory as a scientific discipline?

The initial chapter presents an introduction to the fundamental operations of the so-called Large Language Models (especially ChatGPT) for readers lacking expertise in computational linguistics. Additionally, it offers an introduction to the structure and defining characteristics of the human faculty of language for readers lacking expertise in linguistics. Similarly, an initial examination of the concept of a model in the scientific context is provided, along with an overview of the remaining chapters in this volume, whose authors I sincerely thank for their willingness to get involved in the initiative to produce this book. I would also like to thank three anonymous reviewers who read the manuscript and provided valuable feedback that undoubtedly helped to improve it.

José-Luis Mendívil-Giró

Chapter 1

How do large language models work, and what are they a model of?

José-Luis Mendívil-Giró

University of Zaragoza

Abstract: The first chapter presents an introduction (adapted to readers who are not experts in computational linguistics) to the basic operation of the so-called Large Language Models (especially ChatGPT), and also an introduction (adapted to readers who are not experts in linguistics) to the structure of the human faculty of language and the properties that characterize it. The chapter introduces a crucial distinction between the internalist conception of language (which studies language as a system of knowledge) and the externalist conception (which studies language as a product or result of language use), and then shows that "language models" are models of the product of language use, not of the system of knowledge that underlies that use. Finally, an introduction to the content of the remaining chapters is provided.

Keywords: Knowledge of language, Linguistic Theory, Generative Grammar, Large Language Models, Artificial Intelligence

***

If we ask ChatGPT (the most popular of the latest generation of large language models) what it is and how it works, it writes—in a couple of seconds—the text we have in Figure 1.1.

Although the reader will have understood the definition—which is essentially correct—the wonderful thing is that the algorithm that wrote it does not understand a word of what it says. It also didn't know that humans would be reading its text, and it doesn't even know what a human is.

Figure 1.1: ChatGPT describes itself

You
Can you explain what is ChatGPT and how it works in 200 words?

ChatGPT
ChatGPT is a language model developed by OpenAI, built upon the GPT (Generative Pre-trained Transformer) architecture. Trained on diverse internet text, it excels in generating contextually relevant responses. The model's transformer architecture utilizes self-attention mechanisms for effective context understanding. Users interact with ChatGPT by providing prompts or questions, and it generates responses based on learned patterns. Although versatile, it has limitations, occasionally producing incorrect or biased content. It lacks real-world knowledge and may be sensitive to input phrasing. OpenAI emphasizes responsible usage, implementing safety measures, and welcomes user feedback for continuous improvement.

Nevertheless, it is clear that ChatGPT is an impressive and sophisticated engineering tool, as well as extraordinarily useful. Its potential, including its commercial use, is set to be enormous. Without a doubt, for better and worse, in natural language processing technology, there will be a before and after marked by this latest generation of large language models (LLMs). But surely there will be no such radical change, contrary to what numerous authors have suggested, with regard to our knowledge of human language and the human mind.

1. The human brain and the Chinese room

The Chinese Room is a renowned thought experiment proposed by John Searle to challenge the notion that machines are capable of understanding and processing information in a manner analogous to that of humans. In the *Gedankenexperiment*, Searle postulates a scenario in which an individual who is not proficient in the Chinese language (Searle himself) is confined within a room. The individual has a set of rules that enables him to manipulate Chinese symbols in response to specific inputs. To illustrate, if presented with a question in Chinese, the imaginary Searle will consult the rules and produce a response in Chinese, despite lacking a genuine comprehension of the language. Searle posits that, although the individual may produce responses that appear to be appropriate to native Chinese speakers, he does not actually comprehend the language. Similarly, Searle posits that a machine that emulates this form of symbol processing cannot possess a genuine comprehension of language or the world in the same manner as humans.

Of course, some philosophers have argued that Searle's thought experiment does not serve to reject a real knowledge capacity of Artificial Intelligence (AI), since it may well be that our own brains (the standard we use to define *knowledge*) are also Chinese rooms. We should then conclude that even if there were a real Chinese speaker inside the famous room, her brain would function like the Chinese room she is in. In fact, there is no doubt that, at a certain level, our "knowledge," whether of Searle, of Chinese, or of ourselves, will be supported by groups of cells and tissues that behave like the individual in the Chinese room, or like the algorithm that powers ChatGPT: moving chunks of incomprehensible information from one site to another. Nevertheless, even if this is the case, it does not necessarily follow that we cannot assert that we possess knowledge and that we are aware of our possession of that knowledge.

In any case, beyond the deep paradoxes and mysteries of consciousness and the very nature of knowledge, the relevant question now is whether LLMs such as ChatGPT have a knowledge of language that is analogous or comparable to that which human beings have. The many scientists who believe so are those who defend that ChatGPT is a useful model to better explain how it is possible that we can learn and use the languages of the environment—and even some, like Steven Piantadosi (2023), affirm that these systems are authentic theories of language, while scientists who believe not—for example, Chomsky, Roberts and Watumull (2023)—are those who reject that LLMs are relevant to better understanding our language capacity, even though they are useful as engineering products.

In order to understand the manner in which these inconsistencies arise, it is essential to initially examine the fundamental mechanisms underlying the operation of these LLMs. It is only then that an evaluation can be made as to whether the knowledge currently held about the processes of language acquisition and use in the human brain is comparable or equivalent to that observed in these AI models.

2. Inside the stochastic parrot: how does ChatGPT work?

Computational linguist Emily Bender (Bender et al. 2021) has popularized the name *stochastic parrots* for LLMs as it encapsulates two pivotal characteristics of their operation: the reliance on probability-based word prediction and the absence of comprehension.

And, in fact, what LLMs really do is calculate the probability that a graphic word appears, taking into account the graphic words that it already has chained together.[1]

I have purposely used the expression "graphic word" because, unlike what happens with human language, ChatGPT does not operate with authentic linguistic units (that is, morphemes and lexemes, entities with meaning), but with tokens, that is, sequences of very frequent combinations of letters or characters. Computational linguists had previously identified that the outcomes are enhanced when these tokens, which comprise approximately four characters in English, are utilized. Consequently, a single token is deemed to be analogous to approximately "three-quarters" of a graphic word in English (100 tokens = 75 graphic words). The tokens that ChatGPT manipulates are, in some ways, the written language equivalent of syllables in oral language: recurring sequences of basic units with no correlation to meaning. There is, apparently, no theoretical motivation behind this practice, but rather greater efficiency and flexibility in the task of predicting the next *word* (hereinafter understood as "frequent sequence of characters"), taking into account practically everything that has been written in English and was digitized.[2]

The key question is how to obtain the probabilities that the program uses to evaluate which word to write next to the one you just typed. Note that the aim is not to calculate the probability of the next word based on the last one used alone, but on all the words that have already been used. Otherwise, coherent and natural-looking text would not be generated, but rather a confusing and repetitive sequence. However, Wolfran's calculations indicate that the quantity of written English text in existence is insufficient to enable the calculation of the probabilities associated with groups of words. Given a lexicon of approximately 50,000 words, the potential combinations of three-word sequences exceed 60 trillion. For 20-word sequences (a typical length of written sentences), the number of possibilities is greater than the number of particles in the universe. Consequently, there is not—nor could there ever be—enough written text in the world for those probabilities to be calculated.

So what ChatGPT does to instantiate its basic ability to guess how to continue the piece of text it is given is to use a *model* (a huge mathematical function) that

[1] In this brief description of how current LLMs work, I focus on ChatGPT and I rely on the (much more detailed and technical) characterization by Stephen Wolfram (2023).
[2] For a detailed explanation of the tokenization process in ChatGPT, see Manova's chapter in this volume.

PAGES MISSING

FROM THIS FREE SAMPLE

About the contributors

Vincent J. **Carchidi** is a researcher in cognitive science and the philosophy of mind, with an interest in artificial intelligence. His work is interdisciplinary, most recently focusing on foundational issues in the generative linguistics tradition, including voluntary linguistic behavior. He likewise maintains an interest in moral cognition and human rights attitudes, focusing on leveraging the research strategies employed by generative linguistics for moral psychology. Carchidi's work appears in academic outlets including the *Human Rights Review*, *AI & Society*, *Biolinguistics*, and the *Cambridge University Press*. He also has an extensive publishing record on technology policy in outlets including *The Hill*, *National Interest*, *Military Strategy Magazine*, and *Defense One*, as well as with think tanks including the Middle East Institute, the Foreign Policy Research Institute, and the New Lines Institute.

Vittoria **Dentella** is a postdoctoral researcher in the Department of Brain and Behavioral Sciences at the University of Pavia. She holds a PhD in Cognitive Science and Language from Rovira i Virgili University in Tarragona, Spain, which was funded by a European MSCA grant. Her research focuses on the relationship between psycholinguistic theories of cognition and computation, with a focus on the role of computational models in advancing our understanding of human grammars.

Annika **Heuser** is a Ph.D. candidate in linguistics at the University of Pennsylvania, advised by Professor Charles Yang. She has a B.S. and M.Eng. from MIT in computer science and brain and cognitive sciences. She takes a computational approach, often using tools from natural language processing, in researching child language acquisition. More specifically, her research focuses on phonological development and how children learn discrete representations that differ from the continuous surface forms that they hear.

Jordan **Kodner** is a faculty member in the Department of Linguistics at Stony Brook University with additional appointments in Computer Science, the AI Innovation Institute, and the Institute for Advanced Computational Science. His work and teaching center on child language acquisition, language change, and the evaluation of natural language processing models from the perspective of linguistics and cognitive science.

Evelina **Leivada** is a psycholinguist, currently an ICREA Research Professor at Universitat Autònoma de Barcelona. Her research brings together both the cognitive and the sociolinguistic aspects of the ability to acquire and use language through investigating the effects of developmental trajectory on language and cognition. More recently, she has focused on Large Language Models and how they deviate or match human baselines. Supporting an interdisciplinary approach to language, Evelina is Associate Editor for Psycholinguistics in the Diamond Open Access journal *Biolinguistics*.

David J. **Lobina** is the author of *Recursion* (2017, Oxford University Press) as well as of an increasing number of articles on what may be termed the theoretical repercussions of generative linguistics, not least as a Monday Columnist at *3 Quarks Daily*. A former *Marie Curie* fellow at the University of Oxford and a former *Juan de la Cierva* fellow at the University of Barcelona, he has also worked at the universities of Milano-Bicocca and Rovira i Virgili. As a result, he specialises in theoretical linguistics, psycholinguistics, cognitive psychology, and in the philosophies of cognitive science and psychology. He is currently spending much time thinking about the present status of the language-of-thought hypothesis and is quite tempted to rewrite *The Language of Thought* all over again.

Miguel **López-Otal** is a Ph.D. candidate in Computer Science at the University of Zaragoza. He holds a degree in Spanish Philology and an M.Eng. in Computational Linguistics from the University of the Basque Country. He specializes in Transformer-based models and the analysis of their linguistic capabilities. His current research focuses on ways large language models could better support minority languages with limited text corpora.

Stela **Manova** holds a PhD in General Linguistics from the University of Vienna, where she served as a senior researcher and lecturer. Her academic background spans mathematics, computer science, and languages, and she has published extensively across theoretical, computational, and experimental linguistics. Her key works include *Understanding Morphological Rules* (Springer, 2011), *Affix Ordering Across Languages and Frameworks* (Oxford University Press, 2015), and *Diminutives Across Languages, Theoretical Frameworks and Linguistic Domains* (De Gruyter, 2024). She has edited and co-edited multiple special issues for the journals *Morphology* (Springer) and *Word Structure* (Edinburgh University Press), and contributed to leading handbooks and encyclopedias, including the forthcoming *Oxford Handbook of Iconicity in Language* (2026). Currently based in Vienna, Dr. Manova serves as CEO of MANOVA AI and Principal Investigator of *Gauss: AI Global*, where she applies her linguistic expertise to artificial intelligence research and development.

José-Luis **Mendívil-Giró** is a full professor of General Linguistics at the University of Zaragoza. His research has focused on the theory of grammar, the philosophy of linguistics, and the nature and extent of language change and diversity. He is the author of five books, including *On Biology, History and Culture in Human Language. A critical overview* (Equinox, 2014, with J.C. Moreno). His research articles have been published in journals such as *Folia Linguistica, Journal of Linguistics, Linguistics, Frontiers in Communication, Glossa, Biolinguistics,* and *Theoretical Linguistics.*

Dan **Milway** is a sessional lecturer of linguistics at the University of Toronto. His research focuses on syntactic theory and has published on the theory of grammatical agreement and on the nature of syntactic modification. He writes irregularly about linguistics, philosophy of science, and politics at his website milway.ca.

Elliot **Murphy** is a neuroscientist, linguist, philosopher and political economist at the Vivian L. Smith Department of Neurosurgery, McGovern Medical School, in Houston, Texas. His research concerns the neurobiological basis of syntax and semantics, using intracranial recordings from epilepsy and brain tumor patients. He is the author of numerous books, including *The Oscillatory Nature of Language*, and his research has appeared in *Nature Communications, PNAS, Journal of Neuroscience*, and *Progress in Neurobiology.*

Lucía **Pitarch** is a philologist (Spanish) and Ph.D. candidate in Computational Linguistics at the University of Zaragoza. She has collaborated in European projects such as *Nexus Linguarum*, focused on enhancing open and transparent multilingual resources, and *4DPicture*, a project paving the way towards improving communication in the oncological environment. Her focus is on computational semantics, with contributions on lexico-semantic relation extraction and computational metaphor processing at top conferences like ACL, ESWC or LREC.

Héctor Javier **Vázquez Martínez** is a Ph.D. researcher in Computational Linguistics at the University of Pennsylvania, with a B.S. and M.Eng. in Electrical Engineering and Computer Science from MIT. His work sits at the intersection of deep learning, natural language processing, and linguistics, specializing in modeling human speech and exploring how it can be represented and acquired by both minds and machines.

Charles **Yang** teaches linguistics, computer science and psychology and directs the Cognitive Science Program at the University of Pennsylvania. He is the author of several books on language learning, variation, and change, including *Knowledge and Learning in Natural Language* (Oxford University Press, 2002), *The Infinite Gift* (Scribner, 2006) and *The Price of Linguistic Productivity* (MIT Press, 2016), which won the *Leonard Bloomfield Award* from the Linguistic Society of America.

Index

A

B

E

F

G

H

Q

R

S

X

Y

Z