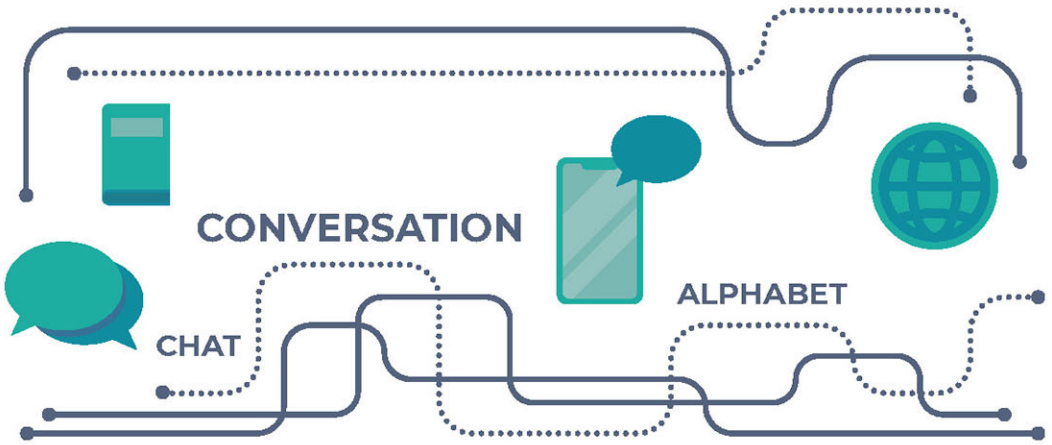


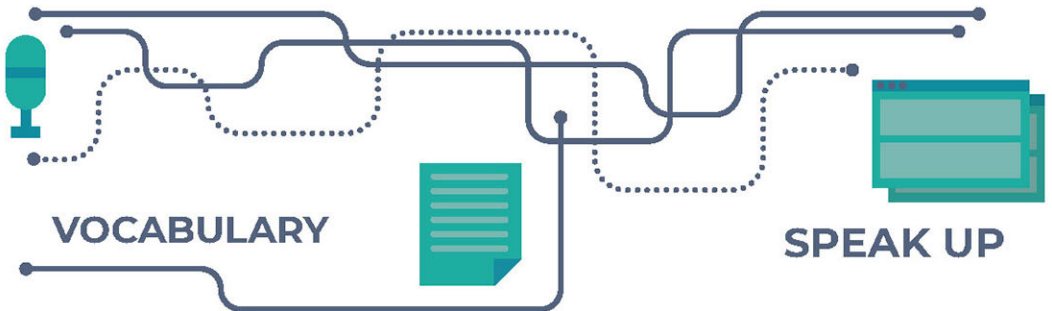
David J. Lobina
Rovira i Virgili University



ARTIFICIAL KNOWLEDGE OF LANGUAGE

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

Edited by
José-Luis Mendívil-Giró



VERNON PRESS

SERIES IN LANGUAGE AND LINGUISTICS

David J. Lobina
Rovira i Virgili University

David J. Lobina
Rovira i Virgili University

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

David J. Lobina
Rovira i Virgili University

Copyright © 2026 by the authors.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior permission of Vernon Art and Science Inc.

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

Product and company names mentioned in this work are the trademarks of their respective owners. While every care has been taken in preparing this work, neither the authors nor Vernon Art and Science Inc. may be held responsible for any loss or damage caused or alleged to be caused directly or indirectly by the information contained in it.

Cover design by Vernon Press with elements from Freepik.

Every effort has been made to trace all copyright holders, but if any have been inadvertently overlooked the publisher will be pleased to include any necessary credits in any subsequent reprint or edition.

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	vii
	List of tables	ix
	Preface	xi
Chapter 1	How do large language models work, and what are they a model of?	1
	José-Luis Mendívil-Giró <i>University of Zaragoza</i>	
Chapter 2	The quo vadis of the relationship between language and large language models	21
	Evelina Leivada <i>Autonomous University of Barcelona</i>	
	Vittoria Dentella <i>University of Pavia</i>	
	Elliot Murphy <i>University of Texas</i>	
Chapter 3	Language models as function approximators of text data: disrupting comprehension through and adversarial attack	45
	David J. Lobina <i>Rovira i Virgili University</i>	
Chapter 4	On modern language models, impossible languages, and anti-science	65
	Dan Milway <i>University of Toronto</i>	

Chapter 5	What kind of linguistic knowledge is encoded in large language models?	83
	Miguel López-Otal <i>University of Zaragoza</i>	
	Lucía Pitarch <i>University of Zaragoza</i>	
Chapter 6	ChatGPT and linguistic theory, with a focus on morphology	109
	Stela Manova <i>Manova AI</i>	
Chapter 7	Evaluating the existence proof: LLMs as cognitive models of language acquisition	147
	Héctor Javier Vázquez Martínez <i>University of Pennsylvania</i>	
	Annika Heuser <i>University of Pennsylvania</i>	
	Charles Yang <i>University of Pennsylvania</i>	
	Jordan Kodner <i>Stony Brook University</i>	
Chapter 8	The creative aspect of human language use: How modern language models fare with an old idea	187
	Vincent J. Carchidi <i>Independent Researcher</i>	
	About the contributors	217
	Index	221

List of figures

Figure 1.1:	ChatGPT describes itself	2
Figure 2.1:	A stepwise risk assessment of LLMs	24
Figure 2.2:	ChatGPT fails to consistently detect attraction errors (October 2023). It also produces agreement errors of the same type in its output. Note that more recent models of ChatGPT get this correct (July 2025 testing), but see Murphy et al. (2025a) for more acute tests.	26
Figure 2.3:	An LLM failure to judge a semantically coherent and grammatical sentence of English as such.	27
Figure 2.4:	ChaptGPT (o3-mini-high) test of structural ambiguity (March 2025)	28
Figure 2.5:	Two LLM failures to generate factually correct content (October 2023)	29
Figure 2.6:	An LLM failure to distinguish between possible and impossible language.	31
Figure 2.7:	The circle of experimentation, taking language as the domain of study	35
Figure 3.1:	The Different Kinds of Artificial Intelligence	48
Figure 3.2:	Rough outline of a deep neural network	48
Figure 3.3:	Adversarial attack on a visual input	56
Figure 3.4:	How an LLM works	57
Figure 3.5:	What an LLM sees	57
Figure 3.6:	What an LLM sees under attack	58
Figure 3.7:	Affix attack on an LLM	58
Figure 3.8:	Two-step attack on an LLM	59
Figure 6.1:	-er (token as Text)	112
Figure 6.2:	-er (token ID)	113
Figure 6.3:	Tokens that are not linguistic units	113
Figure 6.4:	<i>er-</i> (token as text)	114
Figure 6.5:	<i>er-</i> (token ID)	115
Figure 6.6:	Token #1 in the List	115
Figure 6.7:	Token ID [0] in the Tokenizer	116
Figure 6.8:	'linguistic research' (Text)	118
Figure 6.9:	'linguistic research' (IDs)	118

Figure 6.10:	‘crosslinguistic’ (Text)	119
Figure 6.11:	‘crosslinguistic’ (IDs)	119
Figure 6.12:	‘cross-linguistic’ (Text)	120
Figure 6.13:	‘cross-linguistic’ (IDs)	120
Figure 6.14:	Native speakers’ accuracy of recognition of the 60 suffix combinations tested in the experiment (only statistically significant results).	132
Figure 6.15:	‘read’ is a single token	135
Figure 6.16:	‘reads’ is a single token	136
Figure 6.17:	‘organization’ is a single token	136
Figure 6.18:	‘organizations’ is a single token	137
Figure 6.19:	‘illegible’	138
Figure 6.20:	‘illogical’	139
Figure 6.21:	‘illiberal’	139
Figure 6.22:	‘illegal’	140
Figure 7.1:	LM performance on the LI-Adger dataset. Human performance is marked by the vertical line. Baby=BabyBERTa, CHI=AO-CHILDES, News=AO-NEWSELA, Wiki=Wikipedia-1.	173
Figure 7.2:	Correlation matrices of human judgments and LM output means (top) and standard deviations (bottom) on each sentence type on the LI-Adger dataset. Baby=BabyBERTa, CHI=AO-CHILDES, News=AO-NEWSELA, Wiki=Wikipedia-1.	174

List of tables

Table 6.1:	Combinability of the English suffix <i>-ist</i>	129
Table 6.2:	Combinability of the Polish suffix <i>-arz</i>	131
Table 7.1:	Summary performance for 5-grams relative to BabyBERTa on Zorro and BLiMP. The number of paradigms in which a 5-gram model outperforms BabyBERTa and the overall average accuracy across paradigms are reported. Either = either 5-word or 5-tag outperformed BabyBERTa on the entire paradigm. Oracle = sentence pairs were marked correct if either 5-word or 5-tag made the correct prediction.	161
Table 7.2:	Word and tag-level 5-gram models trained on ao-childes plus 5-Gram Oracle and Simple Linear Rule Oracle for Zorro. 5-Gram and Simple Rule scores that are greater than BabyBERTa_ao-childes scores are bolded.	162
Table 7.3:	Word and tag-level 5-gram models trained on AO-CHILDES plus 5-Gram Oracle and Simple Linear Rule Oracle for BLiMP (ANAPHOR AGR through ELLIPSIS). 5-Gram and Simple Rule scores that are greater than BabyBERTa_AO-CHILDES scores are bolded.	163
Table 7.4:	Word and tag-level 5-gram models trained on AO-CHILDES plus 5-Gram Oracle and Simple Linear Rule Oracle for BLiMP (FILLER-GAP through SUBJECT-VERB AGR). 5-Gram and Simple Rule scores that are greater than BabyBERTa_AO-CHILDES scores are bolded.	164
Table 7.5:	Simple Linear Rule descriptions for Zorro. Rules that require sentences within a minimal pair to be compared are marked with an asterisk (in_question_with_aux).	167
Table 7.6:	Linear Rule descriptions for BLiMP (ANAPHOR AGR through ELLIPSIS). Rules that require sentences to be compared are marked with an asterisk.	168

Table 7.7: Linear Rule descriptions for BLiMP (Filler-gap through Subject-verb agr). Rules that require sentences to be compared are marked with an asterisk.	169
--	-----

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 3

Language models as function approximators of text data: disrupting comprehension through and adversarial attack

David J. Lobina

Rovira i Virgili University

[A] computer program that succeeded in generating sentences of a language would be, in itself, of no scientific interest unless it also shed some light on the kinds of structural features that distinguish languages from arbitrary, recursively enumerable sets.

Chomsky (1963, 360)

Abstract: Large Language Models such as ChatGPT are examples of Machine Learning models (ML), and as such, they are best regarded as “function approximators,” or vast correlation machines. That is, given a dataset, however big, of pairs of inputs and outputs, a ML network will construct a function that best describes (approximates) the distribution in the data, and from whence it can generate similar input-output pairs if prompted. In the case of ChatGPT, the dataset is composed of text—the products of the language faculty. In this sense, ChatGPT does not model a natural language per se, but the statistical distribution of text, the latter treated as recorded, static data. *A fortiori*, the claim that some of the results emanating from the development of Large Language Models bear any consequence for the study of language, especially regarding linguistic competence and acquisitional studies, appear to be the case of committing a category mistake.