Original research article / Originalus mokslinis straipsnis
Vol. 1 No. 10 (2025): Gruodis

Artificial Intelligence as a New Lens on Linguistics

[1] Fidan Ismayilova

Accepted: 11.20.2025
Published: 12.03.2025
https://doi.org/10.69760/portuni.0110003

Abstract. This article explores how recent advances in artificial intelligence (AI) are reshaping linguistics by providing new tools, perspectives, and research paradigms. We review the deepening correlation between AI and linguistic science, from historical roots in early computational linguistics to modern neural models. Key methods such as natural language processing (NLP), machine learning, and deep learning have revolutionized corpus analysis, language acquisition studies, and semantic modeling (Groenewald et al., 2024; Incelli, 2025). Researchers have demonstrated that AI can induce human-interpretable grammatical rules from data (Zewe, 2022; Stanford University, 2024). This synergy is bidirectional: linguistic theory offers formal foundations for AI models, while AI opens new horizons for linguistic inquiry (Portelance & Jasbi, 2025; Shormani, 2025). We discuss applications (e.g., corpus exploration, language learning, large language models) and challenges (data bias, ethical concerns) of this AI-driven lens on language (Anthony, 2024; Incelli, 2025). Overall, AI is seen not as replacing human insight but as a "completely new lens" for understanding language structures and use (Shormani, 2025; Portelance & Jasbi, 2025).

Keywords: *artificial intelligence; linguistics; natural language processing; language models; computational linguistics; corpus linguistics.*

1. Introduction

Artificial intelligence and linguistics have become deeply entwined in recent decades. A comprehensive scientometric review finds a "strong correlation between linguistics and artificial intelligence, best manifested by deep learning language models" (Shormani, 2025). Publication counts in AI–linguistics research exploded after the 1990s, reflecting topics ranging from NLP tool development to generative systems such as ChatGPT (Zewe, 2022). Despite this rapid expansion, scholars note that the connection is still "largely overlooked" in some academic communities, particularly those rooted in traditional theoretical linguistics (Incelli, 2025).

This article surveys how AI provides a new perspective on language—effectively a "new lens" for linguistics (Shormani, 2025)—by automating analysis, revealing novel patterns, and testing linguistic theory at a scale never before possible. We begin with the historical background of AI and language science, then examine AI methods and their linguistic insights, discuss theoretical implications, and consider emerging practical and ethical issues.

Computational approaches to language trace back to Alan Turing's 1950 question, *"Can machines think?"* (Stanford University, 2024), which laid the foundations for computational intelligence. Early symbolic AI researchers treated language as a rule-governed formal system; linguists like Winograd and Rosenberg emphasized that "language is one of the most complex and unique of human activities, and understanding its structure may lead to a better theory of how our minds work" (Groenewald et al., 2024).

[1] İsmayılova, F. Master Student, Nakhchivan State University. Email: fidan.ismayilova@ndu.edu.az. ORCID: https://orcid.org/0009-0003-9242-5857.

These early ideas gave rise to computational linguistics, which used computers to model human language through grammars, lexicons, and symbolic operations. Early NLP relied heavily on hand-crafted rules and linguistic expertise, but as Manning and others observed, the field eventually shifted toward data-driven methods: "rather than handwrite out grammars and rules… we get [computers] to learn from language data" (Anthony, 2024). Today, AI systems learn linguistic structure from massive corpora using statistical, machine-learning, and neural-network approaches, radically expanding the empirical base of linguistic analysis.

2. AI Methods in Linguistic Analysis

AI techniques have transformed how linguists process and interpret language data. Modern natural language processing (NLP) tools—including part-of-speech taggers, syntactic parsers, named-entity recognizers, and sentiment analyzers—now operate with "unprecedented accuracy and efficiency" (Anthony, 2024). These tools automate the analysis of massive corpora, performing tasks that once required extensive manual annotation by experts.

In supervised approaches, machine-learning algorithms build predictive models of language use by training on labeled examples (Groenewald et al., 2024). For example, a sentiment classifier trained on online product reviews can evaluate emotional tone or categorize feedback without any explicit linguistic rules. Unsupervised AI methods—such as clustering, topic modeling, and vector-space embeddings—can discover latent patterns across corpora, revealing new syntactic constructions, semantic relations, or discourse trends (Incelli, 2025). AI systems can cluster contexts of a word such as *bank* into categories like "river bank" vs. "financial bank," automatically distinguishing meanings based on usage.

Thus, AI not only streamlines routine annotation but also uncovers previously unrecognized linguistic features—emerging dialect terms, evolving slang, gendered discourse shifts, and semantic change—hidden within large datasets (Incelli, 2025; Groenewald et al., 2024).

In the study of language acquisition and cognitive processes, AI offers powerful simulation tools. Deep neural networks trained on child-directed or learner-directed input can model how grammatical knowledge might emerge. Groenewald et al. (2024) note that AI models "play a pivotal role in modeling language development processes" by simulating key cognitive mechanisms. Researchers have used neural networks to predict missing words in syntactic dependencies and induce agreement, case-marking, and other grammatical rules from corpora, mirroring phenomena described by theoretical linguistics.

In psycholinguistics and second-language research, machine-learning models analyze learners' utterances, predict developmental patterns, and detect systematic errors. These insights "provide insights into the mechanisms underlying language acquisition" (Groenewald et al., 2024). In sum, AI enables a scale of computational experimentation impossible in traditional linguistics, yielding quantitative tests of linguistic hypotheses and offering new forms of evidence for long-standing theoretical debates.

3. Theoretical Foundations and Synergy

At a deeper level, AI and linguistics inform each other's theories. Linguistic science (phonology, syntax, semantics, pragmatics, etc.) has long supplied formal models that AI repurposes. As Shormani observes, linguistics "provides AI with the theoretical foundations necessary for programming, training and working of language models" (Shormani, 2025). Chomskyan ideas about grammar and complexity, for example, underpin formalisms in natural language processing. Conversely, modern AI offers new conceptual frameworks for linguistics. Shormani argues the AI–language correlation is "an integrative phenomenon that unveils and enhances our understanding of how human language and AI interact" (Shormani, 2025). On one hand,

linguistics teaches AI about universal patterns; on the other, AI yields "new approaches, methods and tools for linguistics and linguistic inquiry" (Groenewald et al., 2024).

A striking illustration is the parallel between Chomsky's generative grammar and today's neural language models. Portelance and Jasbi (2025) argue that large language models (LLMs) like GPT are formal generative systems in the Chomskian sense. They note that LLMs generate grammatical sentences from learned rules ("formal generative models") and can serve the same scientific goals Chomsky had in mind: explaining how humans acquire and use language (Portelance & Jasbi, 2025). In turn, generative linguistic theory provides a yardstick for AI: concepts like Universal Grammar and the ease of learnability can guide evaluation of AI language proficiency. Thus, generative linguists are beginning to see neural networks not as threats to theory, but as complementary tools that can "reinforce its basic tenets" and even suggest new insights (Portelance & Jasbi, 2025). In short, linguistics offers a rich lens for understanding AI models, just as AI offers an empirical "window" into language learning and grammar (Anthony, 2024).

Another cross-cutting domain is semantics and meaning. Word embeddings and contextual representations in neural models capture subtle semantic relations across languages. Researchers find that vector representations automatically encode analogies and entailments that were once thought to require explicit symbolic rules. As Gretchen McShane and colleagues note, neural nets "learn to encode words and sentences as vectors… then transform them through arithmetic operations" to mimic meaning (Zewe, 2022). These developments prompt linguists to reconsider classical semantic categories: meaning might emerge as much from distributional patterns as from formal definitions. Ethnographic and anthropological linguists are also exploring AI's "societal lens" – using NLP to track how language use reflects cultural change and power dynamics.

4. Applications and Case Studies

AI techniques are already embedded in many linguistic subfields. In corpus linguistics, AI-driven tools allow linguists to analyze data at unprecedented speed. For example, large pre-trained models can be fine-tuned to serve as concordancers that retrieve and cluster word occurrences semantically, as Laurence Anthony demonstrated for corpus research (Anthony, 2024; Incelli, 2025). In sociolinguistics, network analysis on social media text can chart linguistic change in near real time. As one review notes, AI enables "near-instantaneous processing of language data, enabling linguists to track live changes in language use, evolving slang, and even political or cultural shifts" (Groenewald et al., 2024). In language learning and education, adaptive AI tutors and chatbots (e.g., intelligent tutoring systems, GPT-based writing assistants) are being studied as tools that can model learner language and provide feedback. Ongoing research, for instance, analyzes ChatGPT's generated responses as if they were learner output to investigate how AI's language differs from human production (e.g. sentence complexity, error patterns). Preliminary findings suggest AI-generated text can serve as training data or test items in language assessment, though care is needed since it may lack authentic idiosyncrasies.

A notable case is the rise of chatbots and virtual assistants. Linguists examine how interacting with LLM-powered chatbots shapes communication. One recent study argues that constant AI interaction demands a new kind of "critical interactional competence" for humans, as we learn to interpret and guide AI dialogue (Stanford University, 2024). Others use corpora of AI–human chat logs to analyze politeness strategies or conversational markers. Through these cases, AI acts as a mirror reflecting human communication norms back to us, offering a "completely new lens" to observe pragmatics and discourse patterns (Shormani, 2025).

AI is also used in language documentation and typology. For endangered languages with little data, AI can generate synthetic corpora or help annotate recordings. As one study notes, AI can produce hypothetical word forms or provide translations, assisting field linguists when data are scarce (Groenewald et al., 2024). Moreover, computational models have been shown to discover systematic phonological or morphological rules from small

datasets (Anthony, 2024), potentially accelerating grammar description. An MIT-led project demonstrated an AI that learned the affixation rule in Serbo-Croatian gender formation, inferring the -a pattern without prior knowledge (Zewe, 2022). This suggests AI can sometimes emulate human-like hypothesis generation in formal linguistics.

5. Challenges and Future Directions

While AI offers powerful new capabilities, it also raises challenges that linguists must address. A major issue is data bias and representativeness. Most AI language models are trained predominantly on major languages (especially English) and on certain genres (web text, news). This leads to skewed coverage: "current AI models have been primarily trained on large-scale corpora of standard English and other dominant languages," neglecting thousands of underrepresented languages and dialects (Incelli, 2025). Linguists emphasize that future AI datasets must be more diverse and inclusive, otherwise we risk erasing minority language phenomena. In corpus creation, this means curating multilingual and non-standard corpora so AI tools serve all speech communities (Incelli, 2025).

Another concern is interpretability and oversight. Deep learning models often operate as "black boxes," making it hard to trace why a certain prediction is made. For linguists, this opacity conflicts with the desire to explain language rules. Some cognitive scientists worry that extremely good AI (LLMs) may actually obscure human-like language processes (Stanford University, 2024). As McEnery and Hardie caution, deploying AI corpus tools requires a balance: "the power of these tools must be matched by a commitment to linguistic insight," ensuring that technology does not "overshadow the interpretive essence" of the field (Groenewald et al., 2024).

Practical challenges include the need for large, annotated corpora (for supervised learning) and the computational resources to train models. Groenewald et al. note that dependence on annotated data can be a limitation: creating high-quality labels is labor-intensive, and models trained on biased annotations can perpetuate those biases (Groenewald et al., 2024). Ethical issues loom large: protecting user privacy in language data, avoiding the injection of stereotypes by AI, and the question of authorship when AI generates text. The field of AI ethics in linguistics is emerging to address these.

Looking forward, AI is expected to become even more central to linguistics research. We anticipate tighter integration of AI into linguistic theorizing: for instance, using neural models to propose and test new grammatical hypotheses. Generative models may become standard tools for typological generalizations (e.g. probing what universals LMs discover across languages; Portelance & Jasbi, 2025). Advances in neural-symbolic methods might bridge rule-based and statistical approaches, aligning with how children learn language. In application, we will likely see smarter linguistic assistants for transcribing and analyzing speech (including minority languages). Overall, AI is poised to not just speed up old tasks, but fundamentally reshape how we understand language. As Laurence Anthony puts it, LLMs provide "a completely new lens through which we understand and analyze language" (Anthony, 2024).

6. Conclusion

Artificial intelligence has ushered in a new era for linguistics, enriching it with scalable methods and fresh perspectives. By analyzing vast data with machine learning, AI highlights patterns and correlations beyond human reach (Groenewald et al., 2024; Incelli, 2025). At the same time, linguistic theory grounds AI in a scientific understanding of language. The two fields now form an integrative phenomenon – as one review concludes, their interaction "unveils and enhances our understanding of how human language and AI interact" (Shormani, 2025). In practice, AI tools assist with corpus analysis, language teaching, and even theory testing. But alongside these benefits come important cautions about bias, diversity, and interpretability (Groenewald et

al., 2024; Anthony, 2024). Embracing AI as a lens on linguistics means harnessing its power while retaining critical linguistic insight. If done thoughtfully, this synergy promises to uncover new facets of human language – its structure, use, and evolution – that were previously hidden.

References

Anthony, L. (2024, October 28). *Breaking new ground: AI-enhanced concordance analysis*. Reading Concordances in the 21st Century. Retrieved from https://blog.bham.ac.uk/rc21/2024/10/28/laurence-anthony-breaking-new-ground-ai-enhanced-concordance-analysis/

Groenewald, E. S., Pallavi, S., Rani, S., Singla, P., Howard, E., & Groenewald, C. A. (2024). Artificial Intelligence in linguistics research: Applications in language acquisition and analysis. *Naturalista Campano, 28*(1), 1253–1262.

Incelli, E. (2025). Exploring the future of corpus linguistics: Innovations in AI and social impact. *International Journal of Mass Communication, 3*, 3–10.

Portelance, E., & Jasbi, M. (2025). *On the compatibility of generative AI and generative linguistics*. arXiv. https://arxiv.org/abs/2411.10533v2

Shormani, M. Q. (2025). What fifty-one years of linguistics and artificial intelligence research tell us about their correlation: A scientometric analysis. *Artificial Intelligence Review, 58*, 379. https://doi.org/10.1007/s10462-025-11332-5

Stanford University. (2024, April 18). *The linguist who lay the foundation for generative AI*. Stanford News. Retrieved from https://news.stanford.edu/2024/04/18/laying-foundation-todays-generative-ai/

Zewe, A. (2022, August 30). *AI that can learn the patterns of human language*. MIT News. Retrieved from https://bcs.mit.edu/news/ai-can-learn-patterns-human-language/