UDC 81:004.8:001.8
DOI https://doi.org/10.24919/2308-4863/101-18

Nataliia YEFREMOVA,
orcid.org/0000-0003-1476-1831
Ph.D. in Philology, Associate Professor,
Associate Professor at the Conversational English Department
Lesya Ukrainka Volyn National University
(Lutsk, Ukraine) yefremova05@gmail.com

Valentina BOICHUK,
orcid.org/0000-0002-9870-2195
Ph.D. in Philology,
Senior Lecturer at the Department of Philology
The Municipal Higher Educational Institution "Lutsk Pedagogical College"
of the Volyn Regional Council
(Lutsk, Ukraine) boichukvalia@gmail.com

Larysa RYS,
orcid.org/0000-0003-0775-9629
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Department of German Philology
Lesya Ukrainka Volyn National University
(Lutsk, Ukraine) rys.larysa@vnu.edu.ua

METHODOLOGY OF LINGUISTICS: THE IMPACT OF ARTIFICIAL INTELLIGENCE ON HYPOTHESIS-MAKING

The methodology of linguistics has historically been shaped by the interplay between observation, theory construction, data selection, and explanatory adequacy. From philological comparison and structural analysis to generative formalization and corpus-based description, linguistic inquiry has repeatedly redefined how hypotheses are formed, tested, and revised. The rise of artificial intelligence, especially machine learning and large language models, introduces a new methodological moment. Aritficial intelligence systems can process massive multilingual corpora, detect latent patterns, generate candidate generalizations, simulate judgments, and support exploratory analysis at unprecedented scale. At the same time, they raise serious epistemological concerns: opacity, bias, overfitting to surface distributions, conflation of correlation with explanation, and the risk of outsourcing theoretical reasoning to statistical systems.

This article examines how aritficial intelligence is reshaping linguistic hypothesis-making. It argues that artificial intelligence should be understood neither as a replacement for linguistic theory nor as a neutral computational aid, but as a transformative methodological instrument that changes the pace, scale, and form of hypothesis generation. The article situates hypothesis-making within major traditions of linguistic methodology. It then defines basic methodological functions of aritficial intelligence in linguistic hypothesis-making and analyzes its specific contributions to hypothesis generation across subfields of morphology, syntax, semantics, pragmatics, historical linguistics, and sociolinguistics.

The discussion highlights how artificial intelligence can expand the empirical horizon of linguistics by revealing non-obvious regularities and enabling more dynamic interactions between data and theory. It also shows that AI-generated hypotheses remain dependent on data quality, annotation regimes, modeling choices, and human interpretation. The conclusion is that the future of linguistic methodology lies in a reflexive partnership between computational systems and theoretically informed researchers. In that partnership, aritficial intelligence can accelerate discovery, though explanatory accountability should remain a human scholarly responsibility.

Key words: *research, methodology, hypothesis-making, theoretical interpretation, linguistics, aritficial intelligence.*

Наталія ЄФРЕМОВА,
orcid.org/0000-0003-1476-1831
кандидат філологічних наук, доцент,
доцент кафедри практики англійської мови
Волинського національного університету імені Лесі Українки
(Луцьк, Україна) yefremova05@gmail.com

Валентина БОЙЧУК,
orcid.org/0000-0002-9870-2195
кандидат філологічних наук,
старший викладач кафедри філології
Комунального закладу вищої освіти «Луцький педагогічний коледж» Волинської обласної ради
(Луцьк, Україна) boichukvalia@gmail.com

Лариса РИСЬ,
orcid.org/0000-0003-0775-9629
кандидат філологічних наук, доцент,
доцент кафедри німецької філології
Волинського національного університету імені Лесі Українки
(Луцьк, Україна) rys.larysa@vnu.edu.ua

МЕТОДОЛОГІЯ ЛІНГВІСТИЧНИХ ДОСЛІДЖЕНЬ: ВПЛИВ ШТУЧНОГО ІНТЕЛЕКТУ НА ФОРМУВАННЯ ГІПОТЕЗИ

Історично методологія лінгвістики формувалася під впливом взаємодії між спостереженням, відбором даних, побудовою теорій та адекватністю пояснень. Від порівняння мовних фактів і структурного аналізу до генеративної формалізації та опису на основі корпусів, лінгвістичні дослідження неодноразово переосмислювали те, як формуються, перевіряються та переглядаються гіпотези. Поява штучного інтелекту, особливо машинного навчання та великих мовних моделей, відкриває новий методологічний етап. Системи штучного інтелекту здатні обробляти великі багатомовні корпуси, виявляти приховані закономірності, генерувати можливі узагальнення, моделювати судження та проводити дослідження у значних масштабах. Водночас, вони викликають серйозні епістемологічні проблеми: непрозорість, упередженість, плутання кореляції з поясненням, а також ризик делегування теоретичного міркування статистичним системам.

У статті продемонстровано, як штучний інтелект змінює процес формування лінгвістичних гіпотез. Стверджується, що штучний інтелект слід розуміти як трансформативний методологічний інструмент, що змінює темп, масштаб та форму генерування гіпотез. Визначено основні методологічні функції штучного інтелекту у формуванні лінгвістичних гіпотез: виявлення закономірностей, узагальнення на різних рівнях, виявлення аномалій, моделювання та інтерактивне дослідження. Проаналізовано внесок ШІ у генерування гіпотез у морфології, синтаксисі, семантиці, прагматиці, історичній лінгвістиці та соціолінгвістиці.

Увагу зосереджено на тому, як штучний інтелект може розширити емпіричний горизонт лінгвістики, виявляючи неочевидні закономірності та забезпечуючи більш динамічну взаємодію між даними та теорією. Однак, гіпотези, згенеровані штучним інтелектом, залишаються залежними від якості даних, вибору моделей та людської інтерпретації. Зроблено висновок, що майбутнє лінгвістичної методології полягає у рефлексивному партнерстві між штучним інтелектом та дослідником, що спирається на теоретичні знання. У рамках такого партнерства штучний інтелект може прискорити процес нових відкриттів, хоча відповідальність за пояснення має залишатися науковим обов'язком дослідника.

Ключові слова: *дослідження, методологія, формування гіпотези, теоретична інтерпретація, лінгвістика, штучний інтелект.*

Formulation of the problem. Linguistics, as the scientific study of language, depends on the process of hypothesis-making. Though linguistic hypotheses may concern any field of linguistics, methodology matters because it determines how researchers move from observed linguistic data to claims about underlying regularities, mechanisms, or representations (Saussure, 2011).

The history of linguistics can be treated as a succession of methodological shifts in hypothesis-making. Comparative philology sought systematic correspondences across languages to infer genealogical relationships (Campbell, 2013: 1–14). Structuralism prioritized distributional evidence and formal oppositions. Generative linguistics emphasized abstract mental representations and native-speaker judgments (Chomsky, 1965: 63–107). Functionalist and usage-based traditions focused on discourse, frequency, and communicative pressures (Bybee, 2010: 194–122). Corpus linguistics introduced systematic, large-scale attestation as a central evidentiary base (McEnery, Hardie, 2012: 1–25). Each methodological turn expanded some forms of evidence while contesting others.

Artificial intelligence now adds a new layer to this historical development. It alters the conditions under which hypotheses emerge. Contemporary AI systems can cluster constructions, predict missing forms, identify latent dimensions of variation, model sequential dependencies, compare multilingual structures, and generate plausible explanatory suggestions from heterogeneous data (Jurafsky, Martin, 2023; Manning et al., 2020). These capacities have immediate implications for methodology. They change both how linguists test and generate hypotheses.

This is particularly important because hypothesis-making is often the least formalized stage of research. While methods for testing claims may be standardized, the generation of candidate explanations often relies on expert intuition, deep familiarity with data, and sensitivity to anomalies. AI systems increasingly participate in this exploratory stage by surfacing patterns that humans might overlook (Bender et al., 2021: 610–611; Zhang et al., 2020). This participation is ambiguous as AI can reveal regularities, but it does not automatically offer a linguistically meaningful generalization. Thus, the methodological challenge is not whether AI can produce hypotheses, but how those hypotheses should be evaluated, interpreted, and integrated into linguistic theory.

This research aims to investigate the impact of AI on linguistic hypothesis-making within the broader methodology of linguistics. The specific **research objectives** are as follows: 1) to outline the role of hypothesis formation in linguistic science; 2) to define basic methodological functions of AI in linguistic hypothesis-making; 3) to examine the ways AI contributes to hypothesis generation across subfields; 4) to determine the epistemological risks that threaten the integrity of hypothesis-making; 5) to propose principles for responsible methodological integration.

Presenting the main research material. Methodology is a framework that defines what counts as data, what kinds of explanations are legitimate, and how evidence should be used to justify claims. In linguistics, methodological debates have often revolved around several points of tension: competence versus performance, introspection versus observation, qualitative versus quantitative analysis, formal explanation versus functional interpretation, and universality versus diversity (Chomsky, 1965).

Hypothesis-making comes first among these tensions. In linguistics it typically begins with an observed pattern or problem. The task is then to formulate a claim that organizes the data and predicts new observations. This process usually includes several stages. First, the linguist identifies a phenomenon and delimits relevant data. Then, a preliminary generalization is proposed. This generalization is tested against further evidence, including exceptions and edge cases. The hypothesis is then revised or replaced in light of broader explanatory demands. These may include cross-linguistic applicability, cognitive plausibility, diachronic consistency, or formal economy (Croft, 2001: 9–14).

Traditional hypothesis formation often relies on a combination of trained intuition and disciplined observation. Linguists are skilled at detecting contrasts, identifying distributional gaps, and recognizing theoretically salient structures. This expertise may be constrained by human limitations, as researchers are subject to confirmation bias, theoretical precommitments, and selective attention. For this reason, methodological innovations have often aimed to widen the evidentiary base and reduce arbitrariness (McEnery, Hardie, 2012: 225–236).

By processing large and complex datasets, AI can extend human pattern recognition. It can expose recurring structures across millions of tokens, identify variables correlated across social or linguistic dimensions, and generate candidate groupings without prior labeling (Jurafsky, Martin, 2023). In methodological terms, AI has a strong effect on the discovery context of linguistics, though its implications also extend to justification and explanation.

Unlike earlier computational methods that often required explicit feature engineering and narrowly specified tasks, modern AI can learn representations from raw or minimally processed data. Neural systems can encode contextual similarity, model long-distance dependencies, and detect structures not directly specified by the analyst (Manning et al., 2020). This ability makes AI especially relevant to the early stages of hypothesis formation, where the problem is often not testing a known pattern, but discovering whether a pattern exists at all.

AI contributes to linguistic hypothesis-making through several distinct functions: pattern discovery, generalization across scale, anomaly detection, simulation, and interactive exploration.

At the pattern discovery stage, AI systems help researchers detect hidden regularities that are difficult to identify manually. Studies on corpus and parallel-corpus analysis emphasize AI's ability to extract frequency distributions, identify clusters, detect structural patterns, and uncover hidden correspondences across languages or scripts, including in multilingual and low-resource settings (Micallef, 2026).

AI also supports generalization across scale. A human linguist may detect a pattern in one corpus or one language variety, but AI can test whether comparable structures appear across hundreds of datasets or languages. This makes it easier to formulate broader hypotheses about universals, areal effects, or typological tendencies. A pattern that once seemed idiosyncratic may turn out to be widespread, or vice versa (Haspelmath, 2007: 124–130).

AI is useful for anomaly detection. Linguistic discovery often begins with exceptions. A model trained on typical distributional behavior can flag tokens, constructions, or contexts that diverge from expectations. These anomalies may reveal grammatical change, register-specific uses, annotation problems, or previously undescribed categories.

As far as the simulation function is concerned, language models can approximate distributional expectations and generate continuations under controlled conditions. While such outputs are not equivalent to native-speaker competence, they can help researchers explore possible generalizations and design better hypotheses. If a model consistently prefers one structure over another across contexts, the pattern may suggest a latent constraint worth investigating (Linzen, Baroni, 2021: 200–213).

AI facilitates interactive exploration. Modern computational environments allow researchers to query corpora, visualize clusters, compare variants, and inspect model outputs iteratively. This creates a more dynamic methodology in which hypotheses can be formed, challenged, and reformulated in rapid cycles. The researcher is no longer limited to linear analysis but can move recursively between observation and abstraction.

AI as a discovery tool contributes to hypothesis generation across various subfields of Linguistics. Morphology and syntax are fertile areas for AI-assisted hypothesis-making because they involve complex combinatorial structure. Parsing systems, sequence models, and grammatical induction methods can detect recurrent patterns in word formation and sentence structure across large corpora. This can help linguists formulate hypotheses about constituent order, agreement, case marking, productivity, and constructional alternation (Manning et al., 2020).

Large language models create new possibilities in syntax research. They can be probed using carefully designed prompts to test sensitivity to agreement, island constraints, filler-gap dependencies, and word order restrictions. Although these models do not possess human grammar in any straightforward sense, their successes and failures provide comparative evidence about what kinds of structural generalizations can emerge from distributional learning. This, in turn, may inspire new hypotheses about the relationship between input, representation, and abstraction in human language acquisition (Linzen, Baroni, 2021: 200–213).

In semantics and pragmatics, AI helps researchers examine meaning through contextual usage at scale. Distributional semantic models reveal fine-grained distinctions in lexical meaning, metaphorical extension, semantic shift, and collocational framing. These patterns can generate hypotheses about polysemy, semantic prosody, evaluative meaning, and conceptual structure (Boleda, 2020).

Pragmatics also benefits from AI-based discourse analysis. Models can classify speech acts, detect stance, trace politeness strategies, and analyze turn-taking patterns in conversational data. This enables linguists to ask new questions about how pragmatic functions vary across genres, communities, or platforms.

However, semantics and pragmatics illustrate especially clearly the limits of AI. Models may capture associations without representing inferential structure, speaker intention, or social indexicality in a theoretically robust way. Therefore, AI is most productive here when used as a generator of candidate regularities rather than as a final arbiter of semantic explanation (Bender, Koller, 2020: 5193).

Historical linguistics has always relied on inferential reconstruction, making it especially compatible with AI-driven pattern discovery. Digitized historical corpora, handwritten text recognition, and diachronic language models allow researchers to trace lexical, morphological, and syntactic change across long time spans. AI can detect emerging trends, transitional forms, and shifts in collocational patterns that would be difficult to identify manually. Such tools are particularly useful for hypothesis-making about grammaticalization, semantic shift, analogical leveling, and diffusion pathways. Hypotheses about change must therefore remain sensitive to archival limitations and the non-random nature of historical evidence (Kytö, 2011).

Sociolinguistics has been transformed by computational scale. AI can process massive datasets from speech corpora, social media, messaging platforms, and transcribed interactions, allowing researchers to model variation with highly detailed resolution. Clustering, classification, and regression-enhanced machine learning can identify patterns linking linguistic variables to age, gender, geography, ethnicity, social network structure, and communicative context. These techniques support new hypothesis-making in several ways. They can reveal emerging variables before they are recognized in traditional sociolinguistic interviews and uncover nonlinear interactions among social factors. They can also help identify style shifts across situations and platforms.

Yet sociolinguistic methodology requires reflexivity because AI models may reproduce social bias or treat socially constructed categories as fixed natural facts. If a model infers gender from language use, for example, it may encode stereotypes rather than linguistic realities. Consequently, AI-generated hypotheses about social meaning must be evaluated critically, with attention to ethics, representation, and the politics of categorization (Hovy, Spruit, 2016: 593–596).

The benefits of AI are substantial, but they come with epistemological and methodological risks: opacity, data dependence, correlational overreach, automation bias in scholarship, the erosion of negative evidence and anomaly sensitivity. Regarding opacity, many AI systems, especially deep neural networks, operate through internal representations that are difficult to interpret. A model may distinguish categories accurately without making transparent what cues it uses. This creates a problem for linguistics, which is not only interested in prediction but in explanation. If a model identifies a pattern, researchers still need to know whether that pattern corresponds to a theoretically meaningful generalization (Bender et al., 2021: 615–620).

When it comes to data dependence, AI systems learn from available data, and linguistic data are never neutral. Corpora reflect genre, literacy, platform conventions, transcription standards, colonial histories, and unequal language resources. High-resource languages receive more attention, better tools, and cleaner datasets. Low-resource and marginalized varieties may be misrepresented or excluded. As a result, AI can skew the very landscape from which hypotheses are drawn (Bird, 2010).

Concerning correlational overreach, AI is excellent at finding statistical associations, but not every association supports a substantive linguistic explanation. A model may detect that two forms co-occur frequently, but this does not prove a grammatical dependency, semantic feature, or cognitive constraint. Without theoretical discipline, researchers may mistake predictive success for explanatory adequacy.

Automation bias in scholarship takes place when there is a temptation to treat AI-generated outputs as objective because they are computational. This can reduce critical scrutiny and lead scholars to defer too readily to model suggestions. Yet every model embodies assumptions: tokenization schemes, annotation choices, hyperparameter settings, training objectives, and sampling decisions. AI does not remove interpretation; it relocates and sometimes obscures it.

As far as the erosion of negative evidence and anomaly sensitivity are concerned, some AI systems smooth over irregularity in pursuit of optimization. But in linguistics, the anomalous case is often theoretically decisive. If researchers rely too heavily on averaged patterns, they may miss exactly the forms that challenge and refine hypotheses.

A productive methodology for AI-assisted linguistics must be reflexive, which means explicit awareness of how tools shape observation, inference, and theoretical choice. Initially, AI should be treated as a hypothesis generator, not a self-sufficient explanatory framework. Its outputs are starting points for analysis, not conclusions immune to critique. It should be asked what data the model saw, what distinctions it encodes, and whether the discovered pattern is robust across methods.

In addition, linguistic research should combine triangulation across evidence types. AI-derived patterns should be compared with corpus frequencies, elicitation results, speaker judgments, experimental findings, and cross-linguistic parallels where relevant. A hypothesis gains credibility when it survives multiple forms of scrutiny.

Moreover, researchers should prioritize interpretability and transparency where possible. Interpretable methods can help clarify what cues drive classification or clustering. Even when using complex models, scholars should document preprocessing steps, dataset composition, evaluation criteria, and uncertainty.

Furthermore, a distinction between description, prediction, and explanation should be maintained. AI may improve all three, but they are not identical. A model may predict omitted subjects accurately without explaining the grammatical representation of pro-drop. Likewise, it may cluster politeness markers without explaining the social logic of face management. Linguistic methodology must preserve this distinction if theory is to remain meaningful.

Finally, AI use in linguistics should be ethically accountable. This includes attention to speaker communities, consent, representational harms, and the extraction of language data from vulnerable populations. Methodology is not only epistemic, but it is also social.

AI improves linguistic inquiry most when it augments rather than replaces human expertise. The ideal relation may be indicated as collaborative: AI offers scale, consistency, and pattern detection, while humans supply theoretical framing, interpretive judgment, and epistemic control (Karjus, 2023; Strashko et al., 2024: 247–256).

Conclusions. Artificial intelligence is reshaping the methodology of linguistics by transforming the conditions under which hypotheses are made. It expands the scale of observable evidence, accelerates exploratory analysis, identifies latent patterns, and enables new forms of comparison across languages, modalities, and time depths. The impact of AI is epistemological as it changes how linguists encounter data, what patterns become visible, and how quickly generalizations can be proposed. This creates genuine opportunities for discovery, especially in complex or large-scale domains. At the same time, AI introduces new risks and should not be regarded as a substitute for linguistic reasoning. Hypotheses generated by AI require theoretical framing, empirical triangulation, and critical evaluation by researchers who understand language as a structured, social, cognitive, and historical phenomenon. The future of linguistic methodology therefore lies in a partnership in which AI enhances discovery but does not displace scholarly judgment.

BIBLIOGRAPHY

1. Bender E. M., Koller A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 5185–5198. URL: https://aclanthology.org/2020.acl-main.463.pdf
2. Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021. P. 610–623. URL: https://s10251.pcdn.co/pdf/2021-bender-parrots.pdf
3. Bird S. A Scalable Method for Preserving Oral Literature from Small Languages. The Role of Digital Libraries in a Time of Global Change. 2010. P. 5–14. DOI:10.1007/978-3-642-13654-2_2
4. Boleda G. Distributional Semantics and Linguistic Theory. Annual Review of Linguistics. 2020. Vol. 6. P. 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303
5. Bybee J. Language, Usage and Cognition. Cambridge: Cambridge University Press, 2010. 246 p. URL: https://books.google.com.ua/books/about/Language_Usage_and_Cognition.html?id=Htwy_iPG114C&redir_esc=y
6. Campbell L. Historical Linguistics: An Introduction. 3rd ed. Edinburgh: Edinburgh University Press, 2013. 512 p. URL: https://books.google.com.ua/books/about/Historical_Linguistics.html?id=6pskDQAAQBAJ&redir_esc=y
7. Chomsky N. Aspects of the Theory of Syntax. Cambridge, MA: MIT Press, 1965. 251p. URL: https: //books. google.com.ua/books/about/ Aspects_of_the_Theory_of_Syntax.html?hl=ru&id=iOpsAAAAIAAJ&redir_esc=y
8. Croft W. Radical Construction Grammar: Syntactic Theory in Typological Perspective. Oxford: Oxford University Press, 2001. 416 p. URL: https://books.google.com.ua/books/about/Radical_Construction_Grammar.html?id=3OwDr0gb9CcC&redir_esc=y
9. Haspelmath M. Pre-established categories don’t exist: Consequences for language description and typology. Linguistic Typology. 2007. Vol. 11, no. 1. P. 119–132. DOI: 10.1515/LINGTY.2007.011
10. Hovy D., Spruit S. L. The Social Impact of Natural Language Processing. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics. 2016. P. 591–598. DOI: 10.18653/v1/P16-2096
11. Jurafsky D., Martin J. H. Speech and Language Processing. 3rd ed. draft. Stanford: Stanford University, 2023. URL: https://web.stanford.edu/~jurafsky/slp3/
12. Karjus A. Machine-assisted quantitizing designs: augmenting humanities and social sciences with artificial intelligence. Humanities and Social Sciences Communications. 2026. Vol. 12. DOI: 10.1057/s41599-025-04503-w
13. Kytö M. Corpora and historical linguistics. RBLA. 2011. Vol. 11, Issue 2. P. 417–457. DOI: 10.1590/S1984-63982011000200007
14. Linzen T., Baroni M. Syntactic Structure from Deep Learning. Annual Review of Linguistics. 2021. Vol. 7. P. 195–217. URL: https://marcobaroni.org/publications/linzen_baroni_2021_annual_reviews_linguistics.pdf
15. Manning C. D, Clark K, Hewitt J, Khandelwal U, Levy O. Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences.* 2020. 117(48):30046-30054. DOI: 10.1073/pnas.1907367117
16. McEnery T., Hardie A. Corpus Linguistics: Method, Theory and Practice. Cambridge: Cambridge University Press, 2012. 294 p. URL: https://nlp.csie.org/~sound/Corpus%20Linguistics%20Method%20Theory%20and%20Practice.pdf
17. Micallef L. Artificial Intelligence in Linguistics: Modeling Universal Phonological Systems for Sustainable Communication. Veredas do Direito. 2026. Vol. 23. e235126. DOI: 10.18623/rvd.v23.5126
18. Saussure F. de. Course in General Linguistics / trans. W. Baskin. New York: Columbia University Press, 2011. 336 p. URL: https://www.scribd.com/document/824950487/Saussure-Course-In-General-Linguistics
19. Strashko I., Melnyk I., Kozak V., Torchynska N., Dyiak O. Linguistic Analysis of Texts in Philological Research: The Use of Salesforce Einstein Artificial Intelligence. Forum for Linguistic Studies. 2024. Vol 6. Issue 3. P. 246–259 DOI: 10.30564/fls.v6i3.6601
20. Zhang Y., Xiong F., Xie Y., Fan X., Gu H. The Impact of Artificial Intelligence and Blockchain on the Accounting Profession. IEEE Access. 2020. Vol. 8. P. 110461–110477. DOI: 10.1109/ACCESS.2020.3000505

REFERENCES

1. Bender E. M., Koller A. (2020). Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198. URL: https://aclanthology.org/2020.acl-main.463.pdf

2. Bender E. M., Gebru T., McMillan-Major A., Shmitchell S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

3. Bird S. (2010). A Scalable Method for Preserving Oral Literature from Small Languages. *The Role of Digital Libraries in a Time of Global Change,* 5–14. https://doi.org/10.1007/978-3-642-13654-2_2

4. Boleda G. (2020). Distributional Semantics and Linguistic Theory. *Annual Review of Linguistics,* 6, 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303

5. Bybee J. (2010). *Language, Usage and Cognition*. Cambridge University Press.URL:https://books.google.com.ua/books/about/Language_Usage_and_Cognition.html?id=Htwy_iPG114C&redir_esc=y

6. Campbell L. (2013). *Historical Linguistics: An Introduction* (3rd ed.). Edinburgh University Press. URL: https: //books.google.com.ua/ books/about/ Historical_Linguistics. html?id=6pskDQAAQBAJ&redir_esc=y

7. Chomsky N. (1965). *Aspects of the Theory of Syntax*. MIT Press. URL: https: //books. google.com.ua/ books/ about/Aspects_of_the_Theory_of_Syntax. html?hl=ru&id=iOpsAAAAIAAJ&redir_esc=y

8. Croft W. (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press. URL: https://books.google.com.ua/books/about/Radical_Construction_Grammar.html?id=3OwDr0gb9CcC&redir_esc=y

9. Haspelmath M. (2007). Pre-established categories don’t exist: Consequences for language description and typology. *Linguistic Typology,* 11(1), 119–132. https://doi.org/10.1515/LINGTY.2007.011

10. Hovy D., Spruit S. L. (2016). The Social Impact of Natural Language Processing. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 591–598. https://doi.org/10.18653/v1/P16-2096

11. Jurafsky D., Martin J. H. (2023). *Speech and Language Processing* (3rd ed. draft). Stanford University. URL: https://web.stanford.edu/~jurafsky/slp3/

12. Karjus A. (2026). Machine-assisted quantitizing designs: Augmenting humanities and social sciences with artificial intelligence. *Humanities and Social Sciences Communications,* 12. https://doi.org/10.1057/s41599-025-04503-w

13. Kytö M. (2011). Corpora and historical linguistics. *RBLA,* 11(2), 417–457. https://doi.org/10.1590/S1984-63982011000200007

14. Linzen T., Baroni M. (2021). Syntactic Structure from Deep Learning. *Annual Review of Linguistics,* 7, 195–217. URL: https://marcobaroni.org/publications/linzen_baroni_2021_annual_reviews_linguistics.pdf

15. Manning C. D., Clark K., Hewitt J., Khandelwal U., Levy O. (2020). Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences,* 117(48), 30046–30054. https://doi.org/10.1073/pnas.1907367117

16. McEnery T., Hardie A. (2012). *Corpus Linguistics: Method, Theory and Practice*. Cambridge University Press. URL: https://nlp.csie.org/~sound/Corpus%20Linguistics%20Method%20Theory%20and%20Practice.pdf

17. Micallef L. (2026). Artificial Intelligence in Linguistics: Modeling Universal Phonological Systems for Sustainable Communication. *Veredas do Direito,* 23, e235126. https://doi.org/10.18623/rvd.v23.5126

18. Saussure F. de. (2011). *Course in General Linguistics* (W. Baskin, Trans.). Columbia University Press. URL: https://www.scribd.com/document/824950487/Saussure-Course-In-General-Linguistics

19. Strashko I., Melnyk I., Kozak V., Torchynska N., Dyiak O. (2024). Linguistic Analysis of Texts in Philological Research: The Use of Salesforce Einstein Artificial Intelligence. *Forum for Linguistic Studies*. 6(3), 246–259. https://doi.org/10.30564/fls.v6i3.6601

20. Zhang Y., Xiong F., Xie Y., Fan X., Gu H. (2020). The Impact of Artificial Intelligence and Blockchain on the Accounting Profession. *IEEE Access,* 8, 110461–110477. https://doi.org/10.1109/ACCESS.2020.3000505

Дата першого надходження статті до видання: 22.07.2026
Дата прийняття статті до друку після рецензування: 13.08.2026
Дата публікації (оприлюднення) статті: 02.09.2026