***M. Holovchenko**, BA student*
***N. Baranivska**, assistant lecturer*
Zhytomyr Polytechnic State University

APPLIED ASPECTS OF FORENSIC LINGUISTICS IN THE 21ST CENTURY

Linguistics is a fundamental and multifaceted discipline that helps us analyze the structure of language, which consists of many different aspects. It goes far beyond the scope of grammatical rules and lexical composition. The discipline combines the study of language as a system with its history, function, and role in communication. Equally important is how language influences our perception of the world around us.

In modern times, linguistics is often viewed as an applied tool rather than a theoretical science. The applied aspect allows for the resolution of numerous practical problems in a wide variety of fields—from AI development to legal expertise. In this context, forensic linguistics stands out, which researchers define as " The branch of forensic science which deals with the application of linguistic knowledge… to the legal system is known as forensic linguistics.". Its relevance stems from the fact that any speech act leaves a distinctive mark of the writer or speaker. Thus, language becomes an observational tool capable of answering questions that go beyond mere philological interest.

Forensic linguistics emerged as a distinct discipline in the second half of the 20th century. A key milestone in its development was Jan Svartvik's work, published in 1968. In his book, the author analyzed the testimony in the Timothy Evans case. He identified inconsistencies in the texts of the confessions, which led him to doubt their veracity. This marked the beginning of the use of linguistic methods in judicial practice [1, pp. 5–6].

In the early stages, progress in this field was slow due to a lack of methodology and recognition of criminal linguistics as a distinct discipline. However, the situation has improved over time—methods for text analysis have been developed, and the number of specialists and experts testifying in court has increased [1, pp. 5-6].

Toward the end of the 20th century, linguistics began to develop rapidly. Professional organizations and academic journals emerged. This helped establish it as a distinct academic discipline.

Forensic linguistics relies on a set of methods that enable the analysis of text as a source of evidentiary information. Here, the linguist acts as a "detective" who reconstructs the author's identity and interprets linguistic traces.

One of the main methods is author profiling. It is based on the concept of idiolect—a person's individual linguistic signature. This method determines the characteristics of the author of an anonymous text: age, gender, level of education, and region of residence.

Stylometric analysis uses mathematical statistics to examine word frequency, sentence length, and the use of function words or specific terms, while also paying attention to spelling and punctuation.

Linguistic expert analysis involves analyzing a text from the perspective of its legal content. The linguist identifies signs of extremism, defamation (slander), aggression, and manipulation. They also interpret hidden meanings, communicative strategies, and the context of statements.

Thus, the combination of all these methods makes the linguist a specialist capable of extracting information and meaning from minimal data.

At the core of forensic linguistic analysis lies the theory that every native speaker has a unique, individual version of the language—an idiolect. The authors assert: "every native speaker has their own distinct and individual version of the language … that this idiolect will manifest itself through distinctive and idiosyncratic choices in speech and writing" [1, pp. 161].

This concept implies the possibility of "linguistic fingerprinting," where a person's linguistic traces can serve as identification on par with a signature. Uniqueness here manifests as the "co-selection" of a multitude of linguistic patterns. Their combination in a single text becomes individualizing evidence, as in the Unabomber case [1, pp. 162–163].

Linguistic traceology (the search for traces) includes:

- Systematic errors: If the author has internalized a rule that differs from the standard (for example, spelling "anenome" instead of "anemone"), they will consistently repeat this error, which becomes an important marker of authorship [1, p. 171].
- Dialectal features: Rare dialectalisms or accents can indicate a person's origin. In the "Yorkshire Ripper" case, expert Stanley Ellis used an audio recording to pinpoint the author's residence to within a few hundred meters, based on specific features of the Sunderland dialect [1, p. 5].
- Professional habits: Sometimes a "linguistic trace" of someone else's register becomes evidence. For example, in the Derek Bentley case, analysis showed that his confession had been fabricated by the police, as the word "then" frequently appeared in positions typical of police reports rather than spontaneous speech [1, pp. 178–179].

Traditional forensic linguistics has long relied on the analysis of idiolect—a person's unique "linguistic fingerprint," which manifests itself in their specific choice of words and grammatical constructions. However, in the age of digital communications, criminals are increasingly exploiting the anonymity of the internet, believing that messaging apps and the dark web reliably conceal their identities. The current stage of the discipline's development is shifting from the analysis of classical texts to the study of the "digital footprint," where the fight against cybercrime utilizing artificial intelligence comes to the forefront.

Linguistic analysis enables sociolinguistic profiling, providing investigators with "sociodemographic information (age, gender, origin) that helps guide the investigation" [3, p. 165]. In closed chat rooms and on the dark web, a unique set of abbreviations and slang forms an individual style that experts use to confirm or refute authorship [3, p. 165].

The AI Problem: Humans vs. Neural Networks

- A New Challenge: The emergence of large language models (LLMs) and generative AI "has created new difficulties: cybercrime has increased... and generative AI makes it harder for forensic linguists to attribute text authorship to a specific person" [3, p. 165].
- AI Detection: Studies show that texts generated by AI (e.g., for phishing) can often be identified by their "excessive correctness" and the absence of idiolectal features characteristic of humans. Experiments have shown that "generative AI tends to produce more monotonous and predictable texts" compared to the varied nature of human speech [3, p. 170].
- The Future of Evidence: Linguistics is adapting by incorporating NLP (natural language processing) methods to detect "AI-generated malicious texts." Now the

linguist's task is to determine whether a text is the product of human idiolect or the result of model generation, which is becoming critically important for combating disinformation and cyberattacks.

REFERENCES

1. Coulthard M., Johnson A. An Introduction to Forensic Linguistics: Language in Evidence. London : Routledge, 2007. 250 p.
2. Forensic Linguistics: Language as Evidence in Legal Investigations. *LIFS: Legal Institute of Forensic Science*. 2019.
3. Sousa-Silva R. Fighting Cyber-malice: A Forensic Linguistics Approach to Detecting AI-generated Malicious Texts. *Proceedings of the 1st Workshop on Natural Language Processing for Analysis of Interpretability, Correctness and Safety (NLPAICS 2024)*. 2024. P. 165–174.