

Authorship Attribution Using Sociolinguistic Profiling When Considering Civil and Criminal Cases

¹ **Diana A. Sokolova**

PhD Candidate of Department Language Theory and German languages, Yaroslavl State Pedagogical University
Master of Laws, Forensic Linguist of Interregional Institute of Forensic Expertise LLC
IAFLL's member, Yaroslavl, Russia
diana.falcone.exp@gmail.com
ORCID: 0009-0005-8038-4327

¹ **Alexander F. Sokolov**

PhD in Law, Head of Department of the Criminal Procedure and Forensic Research, Yaroslavl State University
Yaroslavl, Russia
afsokolov@mail.ru
ORCID: 0009-0001-6014-3893

Abstract—This article is devoted to one of the possibilities for identifying the author of an oral or written text - sociolinguistic profiling. Sociolinguistic profiling is utilized as a forensic linguistics technique to identify individuals through language patterns, particularly in criminal cases. It examines how social factors influence language use. This study aims to showcase the significance of linguistic profiling for attributing authorship in texts and emphasizes the necessity for its continuous enhancement while considering its strengths and weaknesses. The study employs semantic-syntactic, lexical-semantic, linguopragmatic, logical, presupposition, authorization, and content analysis methods to investigate linguistic profiling. The research highlights the relevance of sociolinguistic profiling in authorship attribution and underscores the importance of ongoing refinement of the technique, considering its limitations. This study emphasizes the practical application of linguistic profiling in legal settings and underscores the impact of social factors on language use, contributing to the field of forensic linguistics. Data collection involves collecting oral and written texts from criminal and civil court cases to analyze language patterns for authorship attribution. The collected data is analyzed using various linguistic analysis methods to identify individual characteristics and patterns that can aid in authorship attribution. The study addresses the effectiveness of sociolinguistic profiling in identifying authors of texts and explores the impact of social factors on language use in legal contexts. In spite of advantages challenges in linguistics profiling have spurred debates and controversies in academic circles, legal environments, and the public sphere. So, this research highlights the significance of sociolinguistic profiling in authorship attribution and emphasizes the need for further development of this method considering its strengths and weaknesses.

Keywords—authorship attribution, detection of identifying, dialect, features, forensic linguistics, social influence, sociolinguistics, unique speech characteristics.

I. INTRODUCTION

Sociolinguistic profiling (or just 'Linguistic Profiling') refers to a methodological approach that British, American and Australian forensic linguists use to identify distinctive features of a particular person's language use. This approach is not only about defining who a person might be in his professional activities, but also applies to his background, biography, and sometimes – to his intentions or motivation.

The field of forensic linguistics is vast, and its application is important in various fields. It should be noted that in England, which is the progenitor in the field of forensic linguistic research, there is one of the few higher educational institutions in the world that trains students in the field of "Forensic Linguistic Examination" this is Aston University in Birmingham.

Judicial proceedings, criminal prosecutions and even private organizations may request information obtained as a result of linguistic profiling. When a threat is made, a ransom note is sent, or an anonymous statement is provided; the language used can provide valuable information about the sender.

By analyzing the text, experts can presumably identify the person, or at least create his accurate language portrait.

II. ANALYSIS PROCEDURES

One of the main techniques of linguistic profiling is the study by an expert of the vocabulary and syntax used. By analyzing words and phrases, one can gain deep insight into a person's past and life experience. Regional specific features of the language, the use of specific jargon or slang may indicate the place of residence or sociocultural context of the author[1].

Besides, the sentence structure and syntactic features may indicate the level of education and literary preferences of the author.

Advances in digital communications mean that forensic linguists now have to consider typing patterns and Internet slang. Speech identifiers on the Internet can be disguised or changed with the click of a button. But the language remains unchanged, allowing experts to track communication with one person or a group of people.

Mother-of-two Julie Turner, a resident of Yorkshire, disappeared after a date with her boyfriend Howard Simmerson. Her acquaintance Darren received strange SMS messages and a letter, outwardly indicating Julie's depressed state and her desire to temporarily hide and "arrange her thoughts." Police initiated an investigation and discovered unusual use of phrases in these messages which were verbatim identical to phrases used during the interrogation by her boyfriend Simmerson, including the phrase "to arrange her thoughts", which indicated Julie's disappearance may have been faked, and that the investigation was deliberately misled. Simmerson was charged with killing Julie Turner by shooting her in her head. He put her corpse in a barrel and filled it with an oil, which he buried in a field. Simmerson was sentenced to life imprisonment[6].

However, it is important to consider the factors that limit linguistic profiling. People travel, cultures merge, and languages evolve. A person may have acquired language habits from different regions due to long journeys or multicultural upbringing. Situations like this can sometimes blur the picture and make profiling a challenging task.

Moreover, the influence of popular mass media cannot be underestimated. TV shows, movies, books and podcasts for any taste introduce people to a variety of the language patterns and styles. A person may unconsciously borrow phrases or linguistic behavioral pattern from a character he admires or otherwise relate to, which makes the profiling process even more complicated[2].

Linguistic profiling is a dynamic and complex discipline of forensic linguistics. Despite the fact that it offers powerful tools for identification and profiling, it nevertheless requires a differentiated approach that balances the valid data provided by a language with an understanding of its changing and evolving nature.

Dialectal variations and its consequences

Phonetics is the science of studying sounds, and it also plays an important role. If there is audio recording in a case, forensic linguists can analyze the accent, intonation and pronunciation of the speaker to determine the speaker's background or other influence of surrounding factors. A particular accent or unique way of pronouncing certain words can significantly narrow a person's geographic or cultural background[4].

Dialectal variations represent one of the most noticeable features of a language. These differences arise due to geographical, historical and sociocultural factors and play an important role in the field of forensic linguistics. Each region, whether it be a country, a city or even a specific area, can demonstrate its own unique linguistic characteristics. For forensic linguists, these regional fingerprints represent both opportunities and challenges in terms of linguistic profiling.

At a basic level, dialects manifest themselves in vocabulary differences. Words and phrases that are common in one region may be misunderstood in another. For example, in one English-speaking region a soft drink may be called "pop", while in another it may be called "soda" or "cola"; in English-speaker calls one vegetable as "aubergine", but American-speaker calls this vegetable as "eggplant". These small discrepancies, when analyzed in a larger context, can provide significant insight into a person's geographical origin.

Pronunciation and accent are other distinctive features of the dialectal variations. Different regions have various ways of pronouncing certain words, which can be crucial when analyzing the recorded speech. Rhythm, intonation and stress can often be as revealing as the words themselves. For trained professionals, these differences become invaluable markers that help forensic linguists narrow down inferred information about a person or determine his recent location[5].

Grammar and syntax also have regional differences. The order of words in a sentence, the preference for certain verb forms or even the use of certain prepositions may vary due to dialect influence[5].

After the terrorist attacks on the Trade Center and the Pentagon in the United States on September 11, 2001, several high-ranking officials received letters disguised as harmless messages from schoolchildren, but a threatening letter was inside these envelopes, and it contained anthrax spores. Of course, the investigation initially linked the letters to Al-Qaeda's

attack on America. One of the first questions asked was whether the author was a native speaker of English or any language that used the Latin alphabet, or whether he was a native speaker of one of the languages that used the Arabic script as a written language - Arabic, Farsi, Urdu and Punjabi which were spoken in Pakistan. Despite the fact that the letter was deliberately devoid of punctuation and contained grammatical errors, experts came to the conclusion that the text was composed by a native English speaker of American origin[6].

However, understanding the dialectal variations is not just about identifying linguistic differences. It is also about recognizing the reasons for these differences. Migration flows, historical events and cultural exchanges contribute to the evolution of dialects. These underlying factors can provide a deeper and more detailed understanding of an individual or a society in question.

For the first time ever, linguistic data was presented in an Australian court case in 1959. T.G.H. Strehlow, the Lecturer in Linguistics at the University of Adelaide, presented evidence in the South Australian court case against an Aboriginal Rupert Max Stuart. His research revealed the linguistic variations that were ahead of their time. Strehlow described how Stuart first spoke Pidgin English and afterwards the more complex Northern Territory English. His research resonated with studies of the English language of Australia's Northern Territories, which drew attention to the linguistic characteristics of the Aboriginal people. With the development of legal norms and development of the Aboriginal rights in Australia, linguists began to actively study a language in a legal context. For example, Diana Eads became known for her research into Aboriginal English and communication styles among them[7]. Her paper helped in one of the cases where, using linguistic analysis, it was proven that the accused could not use the style of speech attributed to him in the interrogation record[8].

However, as in the case with all linguistic profiling tools, relying solely on the dialectal variations has its drawbacks. The modern world is becoming increasingly interconnected. Thanks to the ease of travel and ubiquity of the Internet, people are exposed to a variety of dialects from an early age. A person may use words from multiple dialects, influenced by travel, migration, mass media or personal relationships. Forensic linguists must therefore exercise caution and avoid jumping to conclusions based solely on the dialect data.

In addition, the social stigma associated with certain dialects may lead people to consciously change their speech patterns, especially in written communication. It should be acknowledged that this fact may be crucial for ensuring an impartial and accurate analysis.

Social factors influencing language

A language, at its core, is a living, constantly evolving matter, continuously shaped by the society in which it exists. The relationship between the social factors and language is intertwined, and each of them influences each other in a variety of ways. When studying a language through the lens of forensic linguistics, understanding these social influences becomes a major concern.

It should be noted that a person's language is formed by his thoughts, and a person's thinking often depends on the level of his education. People with a higher education can demonstrate more extensive vocabulary, use detached sentence structures and be fluent in scientific and formal business styles. They can also easily switch from a formal tone to an informal one, depending on the current situation. On the other hand, people with primary or secondary education can only use spoken language, which is influenced by the way of communication they use in everyday speech.

Socioeconomic status is also another key factor that influences the vocabulary of someone or other. The individuals who, according to the universal stratification of societies, belong to the highest socio-economic status (large owners, successful entrepreneurs), most often have a variety of experiences of interacting with the world, such as traveling abroad or attending cultural events, which directly influence their language and thinking. In their speech, they may mention the names of places, events or even brands, which will be identified as markers that indicate their lifestyle. Such a language will be radically different from the language and thinking of a person who is at the lower level of the universal stratification of societies (marginal, lumpen), formed by an environment where the priority is a matter of survival.

The next marker that influences the vocabulary of a language is cultural background and ethnicity. It has a significant impact on the speech production, since a person, who grew up under the influence of different cultures (for example, European and Arabic) or in a bilingual environment, may use words and phrases from different languages (for example, French and Arabic) in his everyday speech.

These linguistic markers can be explicit (express), for instance, using foreign words mixed with the language of the surrounding society, or implicit (hidden), for instance, following a sentence structure that is more common in another language[9].

Moreover, the social groups and subgroups also form their own “internal” language. A person belonging to a certain subculture, for example, to the programming community, will use various slang words and expressions that are established in this community. The way a person speaks on the current topics and the terminology he uses may indicate his belonging to a particular subculture.

The influence of exposure to the mass media and popular culture on vocabulary has also been noted. The books a person reads, the movies he watches, his favorite music, his hobbies - all this contributes to his linguistic composition. The youth slang, catchphrases from YouTube or Shorts videos, even memes and viral videos can penetrate a person’s everyday language, indicating his habits in choosing the media resources he consumes[10].

However, there are exceptions in any theory, so relying solely on these social indicators when constructing linguistic profiling may be a hasty decision for professional forensic linguists. Stereotypes based on the social factors can be a potential trap. For example, the assumption that a person who expresses his thoughts competently and concisely belongs to the highest level of the society, or that the use of certain slang phrases indicates a lack of education, can also lead to inaccuracies in the conclusion. In this regard, the conclusions of an expert in cases of linguistic profiling are most often of a probable (rather than categorical) nature, one way or the other, helping the investigation or court to narrow down the circle of suspects.

Let us summarize that social factors definitely influence the thinking, language and composition of vocabulary, but in forensic linguistics it is carefully taken into account. A comprehensive understanding of the subject matter requires a balance between acknowledging these influences and avoiding overarching assumptions.

Gender, age and other factors that determine speech patterns

The rich picture of human communication is formed from many determinants, among which gender, age and personal experience are some of the most important. These elements provide access to a person’s mental health and biography, proving invaluable for forensic linguistics.

Everyone knows that social norms, expectations and social roles influence the way men and women use language. Historically, these differences were more pronounced: men and women had completely different spheres of discourse. In modern European society, although some of these distinctions are blurred, subtle linguistic markers still remain. In Muslim countries and countries with the predominance of a conventional view of family relationships, such as the UAE, Russia, these linguistic markers remain pronounced. For example, studies have shown that women, on average, may use more evasive phrases such as “I think” or “Maybe,” while men may use more assertive, dominant phrases. A teenager’s vocabulary is often replete with the terms and phrases that may be incomprehensible to an adult. For example, the Digital Age has given rise to a host of new terms, from “selfies” to “tweets”, that may be second nature to a young person but unfamiliar to older generations.

Age also influences topics of discussion, references and even preferred means of communication. Young people may be more comfortable using digital communication platforms, while older generations may lean towards more traditional means.

Personal experience, covering a wide range of factors, from profession to hobby, leave an indelible mark on the language. A musician may use metaphors related to music, and an athlete may use analogies related to sports. Someone who has traveled extensively may exhibit linguistic influences from different regions, creating a unique blend in his communication. However, while these determinant factors offer a wealth of information, it is associated with challenges. A major concern among forensic linguists is the danger of overgeneralization. The fact that someone uses a certain term does not necessarily place him in a certain age group.

Likewise, relying solely on the European-specific gender norms can be misguided, especially in an era where conversations about gender are evolving and becoming more sensitive in the countries where there is a strong influence of LGBTQ+ being a part of the politics in England, the US and Australia.

Moreover, external influences such as popular culture can cause a theory to fail. An elderly person may pick up modern slang through interaction with the mass media or close communication with younger family members.

Challenges and Controversies in Linguistic Profiling

Linguistic profiling, as an tool in forensic linguistics, presents a multitude of challenges and has been the subject of various controversies. The endeavour to unravel the mysteries of an individual’s background, intentions, or identity through language is undoubtedly a complex one, fraught with both pitfalls and possibilities.

One of the foremost challenges is the dynamic nature of language. Language isn't static; it evolves and adapts with time, influenced by cultural shifts, technological advancements, and societal changes. This fluidity can make it difficult for forensic linguists to pin down specific linguistic traits to a particular group or individual.

Next, there's the challenge of the vast variability within groups. Even within seemingly homogenous groups, such as those defined by age, region, or gender, there's a wide range of linguistic diversity. Two individuals from the same town, of the same age and gender, might have markedly different ways of expressing themselves due to personal experiences, education, and exposure to diverse cultures and media.

The rapid growth of digital communication presents both opportunities and challenges. While it provides forensic linguists with a wealth of data, the informal and ever-evolving nature of online language, replete with emojis, abbreviations, and internet-specific jargon, adds another layer of complexity to the analysis[11].

Lastly, there's the challenge posed by deliberate obfuscation. Individuals, aware that their communication might be subjected to scrutiny, might deliberately alter their language patterns, adopt a different dialect, or use language tools to mask their identity.

II. CONCLUSION

These issues have stimulated debate and controversies in the academic circles, legal environment and public sphere. Reviewers have debated the reliability of linguistic profiling, pointing to cases where it has led to incorrect conclusions. Proponents, on the other hand, highlight its potential, arguing for the improved methodologies and a more detailed approach.

Although linguistic profiling represents a powerful instrument in the forensic linguistics arsenal, it is not devoid of issues and controversies. This field continues to evolve in an effort to solve these issues, improve methodologies and ensure the preservation of language as a reliable witness in the revelation of truth.

REFERENCES

- [1] Coulthard M., Johnson A., Wright D. (2017) *An Introduction to Forensic Linguistics. Language in Evidence. Second Edition.* – Routledge. London and New York. 270 p.
- [2] Olsson J. (2013) *Wordcrime: Solving Crime Through Forensic Linguistics/ Bloomsbury.* 208 p.
- [3] Burrows J. (1987) *Computation into Criticism: A Study of Jane Austen's Novels and Experiment in Method.* Oxford: Clarendon Press. 270 p.
- [4] Canter. D., Chester J. (1997) Investigation into the claim of weighted Cusum in authorship attribution studies. *International Journal of Speech, Language and the Law (formerly Forensic Linguistics)*, IV (2). 252-61.
- [5] Chaski C. (1998) A Daubert-Inspired Assessment of Current Techniques for Language-Based Author Identification. US National Institute of Justice. 2001a. 'Language based author identification techniques *International Journal of Speech, Language and the Law (formerly Forensic Linguistics)*, VIII (1). 1-65..
- [6] Olsson J. (2008) *Forensic Linguistics: Second Edition.* – Continuum. 256 p.
- [7] Eads D. (1994) Forensic linguistics in Australia: An overview. *International Journal of Speech Language and the Law*, 1(2). p. 113-132.
- [8] Eads D., Fraser H., Heydon G. (2023) *Forensic linguistics in Australia: Origins, Progress and Prospects. Elements in Forensic Linguistics// Cambridge University Press.* 88 p.
- [9] Kreuz R. (2023) *Linguistic Fingerprints. How Languages Creates and Reveals Identity.* 241 p.
- [10] Willinsky J. (1994) *Empire of Words: The Reign of the OED,* Princeton University Press. 268 p.
- [11] Picornell I., Perkins R., Coulthard M. (2022) *Methodologies and Challenges in Forensic Linguistic Casework/ Wiley Blackwell.* 191 p.