

# How can Forensic Linguistics help law enforcement officers? Past, present and future applications. ¿Cómo puede la Lingüística Forense ayudar a las fuerzas de seguridad? Aplicaciones pasadas, presentes y futuras.

Miriam Jiménez Bernal

Schiller International University, Madrid, España.

Email de correspondencia: miriam.jbernal@gmail.com

## Resumen

*La Lingüística Forense es una disciplina relativamente novedosa que se enmarca en la Lingüística Aplicada. Sus orígenes tienen raíces en las teorías lingüísticas tradicionales, que ofrecen herramientas para el análisis e interpretación de textos y discursos en entornos legales, así como en campos relacionados, como la Literatura, donde los expertos han trabajado la atribución de autoría en textos antiguos, entre otras. En las últimas décadas, la Lingüística Forense ha evolucionado hasta convertirse en un instrumento fundamental para la resolución de casos legales, desde el análisis de testimonios y peritajes hasta la identificación de criminales a través de la lengua. Este artículo explora sus orígenes, su desarrollo dentro de la Lingüística Aplicada y su relevancia en el ámbito legal. Se presentan algunos casos notables que han demostrado el impacto de esta disciplina en la resolución de casos, resaltando cómo el conocimiento lingüístico puede ser de ayuda a las fuerzas del orden para identificar patrones, interpretar pruebas y aclarar acciones. Por último, se señalan algunas tendencias y retos futuros.*

## Palabras Clave

*Lingüística Forense, fuerzas de seguridad, análisis del discurso, Lingüística Aplicada, evidencia lingüística, ámbito legal*

## Abstract

*Forensic Linguistics is a relatively emerging discipline that has established itself as an essential field within Applied Linguistics. Its origin lies in traditional linguistic theories, which provide the necessary tools for the analysis and interpretation of texts and discourses in legal contexts, but also on related fields, such as Literature, where experts have developed methods to attribute the authorship of ancient texts, among others. Over the last decades, Forensic Linguistics has evolved to become a key tool in the resolution of legal cases, from the analysis of witness and expert statements to the identification of perpetrators through language. This article explores the origins of Forensic Linguistics, its development within Applied Linguistics and its relevance in the field of justice. It also discusses some outstanding cases that have demonstrated the impact of this discipline in solving crimes, highlighting how linguistic knowledge can be of great help to law enforcement agencies in identifying patterns, interpreting evidence and clarifying facts. Finally, some of the future trends and challenges are pointed out.*

## Keywords

*Forensic Linguistics, law enforcement, discourse analysis, Applied Linguistics, linguistic evidence, legal settings*

## I. INTRODUCTION

The interaction between Law and Language provides law enforcement officers and linguists with many opportunities to collaborate and will continue to do so with the support of digital technologies and the new challenges they bring. The applications of Forensic Linguistics methods have been well studied in the English-speaking world and are starting to call the attention of many other linguists around the world. However, the fact that the general audience uses a language to communicate makes it difficult for linguists to prove the relevance of their contributions in the legal field, since mastering a daily use of a language can be misconstrued as an expertise on the roots, strategies and mechanisms of that language. Our main aim in this article is to present law enforcement officers with some clues about the solid scientific grounds for this discipline, some relevant examples of how it has already helped to solve criminal cases and some future trends in which linguists can collaborate with them.

## II. LINGUISTIC THEORIES AND DISCIPLINES

Before we define the Forensic Linguistics areas and what it can do to help law enforcement officers around the world, we should place this discipline in context. Forensic Linguistics can be considered one of the applications of theoretical Linguistics to the field of legal studies and Forensic Sciences. This means that the methods and concepts that conform its basis come from the theoretical trends in the study of language, but also from other applied disciplines in this subject and some related ones. Thus, in order for us to understand how these theories, approaches and notions can be of use in the analysis of discourses and texts involved in the forensic or legal area, we will start by presenting a brief review of them.

### A. Structuralism

The first theoretical approach in the study of language is Structuralism. Saussure is considered the founder of this school, especially after the publication of his *Course de Linguistique Generale* in 1916. There are several structuralist schools, such as those of Geneva, Paris, Prague, and Copenhagen, although, as Foucault (1968) points out, these exhibit differences. In general, structuralists argue that language consists of interrelated structures and units. Grammatical categories are neither universal nor transferable; each language selects its own categories and determines how they relate. Thus, the value of grammatical units depends on their relation to other elements and their function within the sentence or discourse.

Relevant concepts for Forensic Linguistics come from authors such as Jakobson, Bloomfield, and Coseriu, among others. One of these concepts is the distinction between the signifier (sounds and letters) and the signified (meaning, structure imposed on thought). In forensic linguistic analysis, both the topics (content) and their expression (form) are studied. Another key contribution was the distinction between syntagma (the combination of linguistic elements) and paradigm (the relations between an element and those that could occupy its place in a sentence). Finally, Barthes (in Alonso and Fernández Rodríguez, 2006) advocates for discourse analysis as a tool to identify the organizing principles and social relations present in discourse, a perspective with applications in Forensic Linguistics.

### B. Generativism

The next important theoretical trend in linguistics was Chomsky's Generativism, which has evolved in response to debates and criticisms. Generativism introduced the concept of linguistic competence, which is crucial for understanding how native

speakers acquire and use language. Along with this, innatism (the idea that humans are biologically predisposed to learn language) and linguistic universals (common features across all languages) have shaped linguistic models (Chomsky, 1997). Linguistic competence refers to implicit knowledge of grammar, syntax, phonology, and lexicon, distinct from performance, which refers to actual language use. This concept has been integrated into Forensic Linguistics, where competence can relate to factors like education, linguistic origin, or age of the author.

Another key distinction is between grammaticality and acceptability. Grammaticality refers to sentences that follow formal rules, while acceptability involves a sentence's appropriateness in a specific context, influenced by pragmatic factors. When writing, people reveal their linguistic competence, including whether they are native speakers, through their lexical and syntactic choices. Unusual grammatical or lexical structures can indicate a person's native language or their competence in a foreign language. This also helps identify the author's sociocultural background and ethnic identity.

### C. *Text Linguistics*

Text Linguistics studies texts as communicative occurrences, defined by Beaugrande and Dressler (1981) through seven textuality criteria: cohesion, coherence, intentionality, acceptability, informativeness, situationality, and intertextuality. Some of these are particularly relevant to Forensic Linguistics. Intentionality refers to the sender's implicit attitude, while acceptability relates to the importance of the text for the receiver. Informativeness concerns the receiver's expectations about the text's content, and situationality refers to the context that makes a text relevant. Finally, intertextuality addresses the relationship between the text and other prior texts.

Coherence and cohesion are crucial in Text Linguistics (Waller, 2015). Coherence is linked to context, while cohesion deals with internal text structure. In Forensic Linguistics, mechanisms of cohesion—such as vocatives, imperatives, and interrogatives—play a vital role in interpreting texts, their veracity, and the author's identity. Cohesion connects surface elements, while coherence connects concepts within the text. This textual universe may involve a single text or a set of interconnected texts. Coherence includes relational categories like causality, purpose, or time. In Forensic Linguistics, text and discourse analysis helps understand deeper meanings, perceptions of truth, responsibility, and focus on statements, which is valuable in deception detection, witness statement analysis, and bias studies in police interrogations.

### D. *Applying theoretical Linguistics to practice*

One of the most significant applications of Linguistics has been in language learning and acquisition. In this field, both Neurolinguistics and Psycholinguistics have undergone a remarkable development in recent decades. These disciplines are concerned with the study of the cognitive and neurological processes involved in language use, focusing on how the brain processes, produces and understands language. In turn, theories of second language learning have provided frameworks for better understanding how people acquire languages in addition to their native language, exploring both the teaching methodologies and the cognitive processes involved in learning a foreign language. Today, we know that the brain plays a fundamental role in language acquisition, comprehension and production, and scientific studies attempt to identify general processes that can be applied to all human beings. However, we must remember that, when analysing language performances, we have to consider their variability, since even the same person will not always present the same linguistic behaviours in all situations, as these may be influenced by a number of factors, such as the social environment, the emotional state, or the particularities of the communication context (Caplan, 2018).

Psycholinguistics and Neurolinguistics have contributed greatly to the development of language acquisition models, but Sociolinguistics has also been interested in issues related to second language acquisition since, if the social context can be considered one of the factors influencing the acquisition process, then this discipline has much to say. As for the methodology of sociolinguistic studies, while quantitative and qualitative approaches have traditionally been seen as separate methodologies, in recent years there has been a growing tendency, previously uncommon, to combine them in the same study. This integration of both approaches has proven to be especially useful in the analysis of conversational phenomena, which have long been the main object of study of Sociolinguistics. The most studied linguistic variation over time has been that which refers to oral productions, so Phonetics is one of the most documented components in this field.

However, in the written language, and although “to account for a linguistic phenomenon we need different types of information from different components of grammar” (Akmajian et al., 2001), we will focus mainly on variation related to Semantics and Morphosyntax. The concept of variation is valuable because it allows us to obtain relevant information about the sender of the message. Although the notion of idiolect (the individual variant of language) is closer to the idea of style and we cannot overlook the fact that both dialect and sociolect (social variants) significantly influence the linguistic choices of individuals, contributing to shape their way of speaking. Afarli and Maehlum (2014) emphasize the fact that languages do not evolve and change in a uniform way, but according to the dynamics of the speakers and the interaction of various contextual and social factors... If we talk about linguistic variation, then it is also necessary to take into consideration the importance of formal and informal register. Informal register takes place in social contexts where there is no protocol, where a relaxed atmosphere prevails and speech can be spontaneous and not subject to censorship (Akmajian et al., 2001).

Another remarkable linguistic application and discipline is Computational Linguistics, which has experienced many advancements in recent decades (Moreno Sandoval, 2019), a progress that has gone hand in hand with advances in the field of Computer Science. This discipline has not only facilitated the creation of computer languages that have given rise to software and applications, as well as the development of digitized corpora, but has also played a key role in the optimization of web search engines. Joint efforts between linguists and computer scientists have resulted in more efficient information search tools, while contributing to the development of models that simulate human language use by systematizing linguistic processes. This advance has allowed progress in both quantitative and qualitative analysis in the Social Sciences.

It should be noted that empiricism in Linguistics is closely linked to research practices based on the observation of real data and phenomena. In the framework of Structuralism and Pragmatics, this empirical observation implies a detailed description of the components and characteristics of language. Corpus Linguistics further deepens this empirical approach by elaborating linguistic corpora that include both oral and written data, with the aim of studying a wide variety of linguistic features, from concordances to close word combinations or most frequent usages. The corpora generated by ethnographic techniques and collected in the linguistic field are also of great value for other applications of this discipline, such as the identification of common errors in second language learning or in sociolinguistic studies. Sociolinguistics studies the variation of the different linguistic levels (phonetic-phonological, morphosyntactic, lexical, semantic, etc.), variation that depends on factors such as ethnicity, gender, age... The impact of these factors on language is closely related to the identity of individuals and their linguistic behaviour, and it would be important in Forensic settings for, for example, linguistic profiling or attribution of authorship of texts, some of the cases that a Forensic linguist may have to deal with.

Finally, regarding Pragmatics, it is both a linguistic component and the discipline studying it. It is dedicated to the study of those extralinguistic factors that influence communication, so that the context, the intentionality of the speaker and the relationship between the interlocutors become relevant. Some of the main theories that can be found are Grice's principle of cooperation or Austin and Searle's speech act theory. Concerning the theory of speech acts by Austin (1962) and Searle

(1970), we can say that it relates intentionality with the form of the message, describing the association produced between the will (hidden or not) of the sender and the response of the interlocutor, and starting from the idea that “to speak” is “to do”. We can say that, in terms of the three levels established by the theory, the perlocutionary level is the intended effect, and the intention, the illocutionary level, while the locutionary level would be the message as it is expressed.

Another major contribution from Pragmatics, after Grice’s conversational principle, is the concept of implicature or inference, which refers to the reinterpretation of the message by the receiver when one of the maxims is not fulfilled. According to various authors (Yule, 2010), this suggests the existence of an implicit meaning that is not expressed explicitly or literally in the text, whether oral or written, but is derived from it and from the rules that are followed or transgressed, in such a way that it must be interpreted by the receiver of the message. For this reason, semantics alone is not sufficient to analyse texts, especially when it comes to evidence in legal cases that could result in a custodial conviction or a more serious sentence, depending on the legal context. Moreover, these implicatures can originate from various contextual elements, such as situation, place, communication channel, time or subject matter, and are based on the shared knowledge between the interlocutors, as well as on their presuppositions and common information.

### III. ORIGINS AND AREAS OF STUDY

Forensic Linguistics is a discipline that explores the relationship between language and law, answering legal questions about the use of language. The term was introduced by Svartvik in the late 1960s, and according to Ramirez (2017), it can be understood as the application of linguistics to judicial issues. In 1997, David Crystal incorporated this term in his encyclopaedia, linking it to stylistics. The author also mentions historical antecedents that reflect the interest in topics today related to Forensic Linguistics, such as the ability of the eighteenth-century blind policeman John Fielding to identify criminal voices, the study of authorship in the Bible, and the copyright regulation in the United Kingdom in 1710, aimed at protecting authors from unauthorized copying.

According to Chaski (1999), modern forensic linguistics emerged after the “Daubert decision”, which established the criteria for the admission of scientific evidence in this field. The linguistic tools used in the discipline range from the analysis of syntactic and lexical patterns to theories such as speech acts or theoretical semantics. Common areas of application include speaker identification, plagiarism detection, text authorship attribution, legal and judicial discourse analysis, threat assessment and criminal intent, as well as translation and court interpretation and linguistic profiling. Gibbons and Turell (2008) group these studies into three main areas: legal language, judicial language and language as evidence.

The origins of Forensic Linguistics can be traced back to a 1927 case in the United States. A ransom note addressed to a man named Duncan McLure was written by someone who misspelled Duncan's last name in a revealing way. While most people spelled it "McClure," Duncan uniquely used "McLure." This subtle linguistic detail exposed the author of the ransom note as a member of Duncan’s own family rather than a stranger.

Forensic linguistics gained further attention in the mid-20th century in the United States, largely due to lexical ambiguities in the Miranda warning. The Miranda warning informs individuals of their legal rights, and police officers typically recite it during criminal investigations. Concerns arose about whether the language of the warning was universally understandable across different regions and populations. In 1966, its wording was standardized in English to address these issues. Today, similar concerns persist, particularly when dealing with individuals who are not native English speakers.

In the United Kingdom, the field of forensic linguistics began to gain prominence as distrust in police statements grew.

Investigations revealed that police often failed to accurately represent the statements of suspects or witnesses. Essential linguistic details—such as pauses, hesitations, and minor corrections—were frequently omitted, casting doubt on the authenticity and completeness of these records.

Nowadays, with the increase of crimes committed through the Internet and other electronic media, the need arises to analyse digital texts in order to identify their perpetrators. Traditional methods, such as handwriting analysis, are not useful in these cases, due to the lack of relevant data in the digital environment. Therefore, the investigation in these cases must incorporate linguistic and discursive elements that allow examining both the form and the content of the message. Forensic Linguistics plays a crucial role here, although it faces challenges and biases that hinder its full acceptance.

In the English-speaking world, forensic linguists are called to testify in cases such as trademark disputes and analysis of law enforcement recordings. In the former, linguists assess whether similarities between trademarks are likely to cause confusion among consumers, or whether, on the contrary, the differences are sufficient to avoid confusion. In the second case, experts in discourse analysis, such as Shuy (2016), try to unravel the intentions and thoughts underlying a person's words, as well as to identify linguistic features that allow the individual to be recognized. In both types of cases, computer tools and linguistic corpora are of great help in performing these analyses.

Authorship attribution in electronic or printed texts, excluding handwritten ones, is based on aspects such as punctuation, grammar, spelling and “stylistic deviations”. These elements have been key in numerous cases in recent decades. However, Spanish courts have not yet adopted these linguistic tests with the same regularity as in other countries, such as the Anglo-Saxon ones, where Forensic Linguistics is considered an objective scientific discipline. Examples of studies, such as that of Basim (2012), demonstrate the usefulness of these tools for the attribution of authorship in suicide notes and other relevant texts.

Linguistic profiling, closely related to the attribution of authorship of texts, involves a discourse analysis to identify distinctive features of an author's writing. In contrast to authorship attribution, which is based on the comparison between dubious and indubitable texts, the aim of linguistic profiling is to extract specific features of the author's identity from his or her writing. This includes, for example, the analysis of vocabulary, syntax and other linguistic elements in texts such as threat notes, in order to infer social characteristics of the author, such as gender, age, educational level or geographical origin. This type of analysis is part of Sociolinguistics, which has developed numerous studies on how language use varies according to the gender, age or dialect of the speakers.

In recent years, a number of textual analysis software programs have emerged, allowing the creation of linguistic profiles from different approaches. Tools focused on lexical, syntactic and other aspects analysis go beyond traditional lexicography (Kilgarriff and Rundell, 2002) and benefit from advances in big data and data analysis, facilitated by text mining on the Internet and social networks. In the Hispanic domain, the PRESEEA project (Crespo, 2017) has gathered a sociolinguistic corpus representing speech variation with the aim of creating automatic sociolinguistic profiles. This project has proven useful in detecting the gender and background of speakers, although there are still limitations in the accuracy of variables such as age and educational level. Despite these difficulties, tools that analyse syntactic and lexical structures remain crucial for forensic linguists.

The attribution of authorship of texts, which originated in studies of classical literary texts, such as the Bible, seeks to determine whether the author attributed to a text is really the one who wrote it. This type of analysis has been fundamental in stylometry, a field that studies the unique linguistic characteristics of an author to identify his or her writing style (Frías, 2009; Celma and Ruiz, 2021). An early example is found in the work of Chaski, who analysed the syntax of several texts to

demonstrate that the supposed author of a suicide note was not the real author, using statistical procedures to measure the complexity of syntactic structures.

Stylometry, key to authorship attribution, refers to the grammatical and lexical choices an author makes when writing, which is associated with the concept of idiolect - the particular way in which an individual uses language. Elements such as the diversity of vocabulary, the number of words used, and the frequency of certain pronouns and particles are distinctive features of individual style. Lexical, syntactic and morphological choices reflect not only the author's personality and social status, but also his or her time in life, since the same person may write differently at different stages of his or her life, influenced by personal events.

In the study of idiolect and style, it is crucial to take into account the marked and unmarked linguistic forms. The foregrounded form is the one that stands out and attracts attention, while the backgrounded form is the most common and goes unnoticed. The choice between one or the other reflects the author's personal preferences and is indicative of his or her linguistic traits. In the field of Forensic Linguistics, authorship attribution focuses on identifying these idiolectal features, such as the use of marked forms, through both quantitative and qualitative techniques, and comparing these features in different texts to determine authorship (Cicres and Gavalda, 2014).

The sociolinguistic studies initiated by Labov are closely linked to the work of forensic linguists, especially with regard to the analysis of linguistic variation. In recent years, individual variation within social groups, or idiolectal variation, has become increasingly relevant as an area of study (Cicres and Gavalda, 2014). In addition, the degree of influence of factors such as textual genre on linguistic choice has begun to be discussed, comparing it with the impact of the communicative situation and the speaker's personal preferences.

One of the best-known cases in which stylistic or idiolectal trait analysis was used to identify a criminal was that of Theodore Kaczynski, known as UNABOMBER. For years, Kaczynski sent bombs to universities and airlines, causing more than twenty injuries and three deaths. In 1995, the perpetrator sent a 35,000-word manifesto explaining his motives and demanding its publication. The hope of identifying the perpetrator was fulfilled when a couple recognized a close relative (brother and brother-in-law, respectively) in the style and content of the manifesto, prompting the authorities to investigate further. Based on an analysis of Kaczynski's previous texts, his arrest was finally secured. FBI investigators stated: that the linguistic analysis, combined with other evidence, provided the basis for a search warrant (*FBI 100, 2008*).

In more recent times, the most mediatic cases in the field of Forensic Linguistics have been those related to kidnappings or murders, in which law enforcement agencies have documents (text messages, suicide letters, etc.) and several possible perpetrators. An example of this is the case described by Coulthard concerning the abduction and murder of a 15-year-old girl by her uncle, with whom she was in a relationship. The man sent several messages from his cell phone in order to create the false impression that the girl was still alive and had disappeared of her own free will. When the phone was turned off, police were unable to identify the location from which the messages were sent. This forced investigators to resort to analysing the content and form of the messages for clues to the perpetrator. They then sought the advice of a forensic linguist.

Analysis of the messages made it possible to identify several distinctive features of the young woman's idiolect, such as her choice of vocabulary, linguistic turns of phrase, use of abbreviations, punctuation and the use of upper- and lower-case letters. By comparing these features with those of the texts in question, it was concluded that it was very likely that the young woman was not the author of these messages. In addition, when the texts were checked against the family member's regular messages, further coincidences were discovered in terms of abbreviations, capitalization, punctuation and certain

specific terms. This analysis led the forensic linguist to conclude that the messages were probably written by the uncle, who attempted to conceal his identity by impersonating the young woman but did so ineffectively.

What has been discussed above suggests that, in the collective imagination outside the academic and professional field of linguistics, there are several factors that influence or determine the way in which individuals express themselves linguistically, either through actual characteristics or socially accepted stereotypes. Furthermore, this understanding of linguistic factors allows speakers to “disguise” the way they communicate. This tactic has been used for a variety of purposes: from law enforcement officers pretending to be potential victims in order to catch sex offenders, to criminals attempting to confuse their victims or investigators through false identities, as is the case in the documents analysed in these studies. However, it has been shown that, even with the right linguistic skills, creating a perfect “disguise” is extremely difficult, as it usually fails to hide all the necessary features or commits inconsistencies that allow the author of the message to be identified.

A clear example of this difficulty can be found in one of the most famous cases, that of the devil strip, solved by Roger Shuy. The linguist, in analysing a kidnapping note, suggested that the authorities focus on a well-educated suspect from Akron, Ohio. Shuy noted inconsistencies in vocabulary usage: the author misused simple terms but correctly used more complex terms. He also observed proper handling of punctuation and grammatical structures. This led him to the conclusion that the author was trying to simulate a lower level of education than he actually possessed. The use of the expression devil strip, a term characteristic of the Akron linguistic variant, was instrumental in identifying the suspect.

Linguistic analysis not only focuses on lexical choices or the correct use of grammar and punctuation; the use of rhetorical strategies is also considered. An example of this is the case solved by Wald and Shuy, in which ironic repetition and the peculiar use of abbreviations were the common elements between two groups of texts: one of an alleged stalker and the other of an alleged serial killer. This pattern allowed the researchers to conclude that both texts belonged to the same author. Although these cases are relatively frequent in the English-speaking world, in Spain linguistic analysis has been limited mainly to those aspects that can be measured by means of computer and digital tools, under the premise that knowledge of language is general and that only those aspects that can be measured objectively, by means of percentages of correctness and significance, can be considered scientific.

Along the same lines, stylometric techniques and computer applications have driven the development of a quantitative approach to authorship attribution and profiling, aligning with the most accepted standards in the scientific community. Several researchers have attempted to apply these tools to identify distinctive features in texts, in order to differentiate authors or groups of authors based on their linguistic style (Estival et al., 2007). These studies employ methodologies that include corpus data collection and subsequent processing using programs such as the Text Attribution Tool (TAT).

Speaker identification in the analysis of oral texts has also addressed the challenge of voice “disguise”, an issue that, as in other fields, faces the difficulty of hiding or modifying voice traces. This discipline has been among the first to receive recognition both in the scientific community and in the legal field in Spain, since there are software programs that give a more “objective” and “scientific” character to the results obtained. However, in this case, the decisive characteristics are not only the words used, but also aspects such as tone of voice and pronunciation. As with written linguistic variations, some features of social groups are accentuated to create false identities, while others are hidden, but the result remains the same: not all features are conscious, so that some remain evident and do not disappear, allowing the speaker to be identified.



IV. HOW IS FORENSIC LINGUISTICS HELPING LAW ENFORCEMENT OFFICERS AROUND THE WORLD?

Forensic Linguistics can contribute to our understanding of the law, of the power dynamics exerted through language in legal settings and to how veracity is constructed in texts and discourses, for instance. Its possibilities and applications to the academic study of language are wide, but it transcends theory and is, primarily, an applied discipline. Traditionally, Forensic Linguistics has been associated with three main areas: understanding the language of the law, understanding the language used in legal settings and procedures, and providing linguistic evidence. Following the Daubert standard, a methodology is valid whether the technique or theory can be and has been tested, whether it has been subjected to publication and peer reviewed, if the known or potential error rate are identified, whether there are standards controlling its operation and whether it has attracted wide acceptance in a relevant scientific community. It would be safe to say, then, that Forensic Linguistic methods are to be considered valid, although the cases can be so varied that establishing common methods for the different languages and evidence is still a work in progress.

Forensic Linguistic analysis, methods and conclusions can help law enforcement officers in a wide range of cases (threats, extortion, hate speech, fraud...) and areas (witness statements, police interviews, legal translation, interpreting in trials...), including, but not limited to, the following ones:

Table 1 . How Forensic linguists help law enforcement officers

Moment	Contribution
During the investigation	Providing scientific knowledge to support cases.
	Studying linguistic evidence.
	Improving the rapport with victims.
	Narrowing down the number of suspects.
Before the investigation	Providing deeper understanding of the law.
	Providing training for interviews and infiltrations or impersonations in cases.
	Improving the understanding of the linguistic strategies used in linguistic crimes and related felonies, so that they can create awareness campaigns.
	Identifying and avoiding abuse and bias in interviews in legal and judicial settings.

Regarding the provision of scientific knowledge to support cases, the study of linguistic evidence in cases and narrowing down the suspects, for instance, a quick overview of some of the most famous ones will offer an idea of the relevance of linguistic evidence and the work of Forensic linguists.

A. The Bentley case

One of the most widely recognized cases where Forensic linguistics played a crucial role is the Bentley case. Derek Bentley, a 19-year-old with developmental disabilities, was convicted of murdering a police officer during a robbery in 1952 and executed for the crime in 1953. Following years of protests by his family, the case was reopened in 1993. Forensic linguist Malcolm Coulthard was part of the investigative team for the reopened case. The primary evidence used against Bentley

was a statement he allegedly made after his arrest. However, Coulthard argued that the statement had been altered by the police. At the time of the original trial, the police had sworn under oath that the statement was an exact, unaltered record of Bentley's monologue. Based on this evidence, Bentley was found guilty and subsequently executed.

Coulthard reanalysed the statement to determine whether it was genuinely Bentley's own words, free from police interference. During his analysis, Coulthard identified several distinct lexical choices and phrases that were more characteristic of a police officer's language than that of an average person, especially someone with limited verbal communication skills like Bentley. Additionally, Coulthard noted that Bentley's statement included several denials of events. In a natural narrative, speakers typically focus on what happened, rather than emphasizing what did not happen. This suggested that the statement was more likely the result of police questioning rather than an uninterrupted monologue by Bentley. Further evidence of police interference came from the use of the word *then*. In everyday language, *then* is often used at the beginning of a sentence to describe sequential actions, such as "Then I..." when referring to oneself. However, in Bentley's statement, *then* frequently followed the pronoun *I*, as in "I then ...," a construction that is much less common in natural speech but often found in police reports. Coulthard concluded that the authorities had manipulated the statement and were at least partial authors of the document.

To substantiate his findings, Coulthard relied on linguistic corpora specifically created for the case. One corpus contained witness statements, while the other comprised police statements. By comparing the frequency of *I then* across the two corpora, Coulthard found that the construction occurred approximately ten times more often in police statements than in witness accounts. This statistical evidence was sufficient to discredit the validity of Bentley's alleged confession. Ultimately, Bentley was posthumously pardoned.

### *B. The Miranda Warning and linguistic ambiguity*

While not tied to a specific case, the development of the Miranda warning in 1966 was a pivotal moment in the intersection of linguistics and law in the US. The Supreme Court decision in *Miranda v. Arizona* led to the requirement that individuals in police custody be informed of their legal rights. Concerns soon arose about whether the language used in Miranda warnings was universally comprehensible. Linguists were called upon to analyse the wording and ensure it was accessible to a broad audience, laying the foundation for modern forensic linguistics.

### *C. Letters to an actress*

In this particular case, several anonymous emails and letters were sent to an actress, claiming she was under government surveillance. At the time, the actress was already involved in a separate criminal investigation. When examining the letters, the police suspected that the actress herself might have written them to divert attention from the ongoing case. To verify this, a forensic linguist was brought in to analyse the texts. The first task was to determine whether the anonymous texts had been written by multiple authors or a single individual. The forensic linguist began by analysing the writing style in the letters, searching for distinctive features. One notable finding was a strong sense of cohesion among the anonymous texts. For instance, the author frequently referred back to statements made in earlier letters, suggesting they were written by the same individual. Additionally, the recurring phrase "your every move is being monitored" stood out. The linguist searched for this phrase in various linguistic corpora and search engines and found it to be extremely uncommon. The rarity of this formulation made it unlikely that multiple authors would independently use the same phrase, further supporting the idea

of a single author.

Another peculiar detail was the repeated use of the word trace in connection with car. The forensic linguist examined its usage across several corpora and found it to be exceptionally rare. Typically, trace is used to describe objects that are lost and being searched for, whereas track is the appropriate term for following the movements of an object. This misuse suggested an intended meaning of track, further highlighting the unique linguistic patterns of the author. The next question was whether the actress herself had written the anonymous letters as a tactic to distract the police. To answer this, the linguist analysed known letters written by the actress, particularly those in which she referred to the anonymous letters. A critical discovery was the same confusion between track and trace, which appeared in both the anonymous texts and the actress's own writing. This linguistic overlap provided compelling evidence that the actress was indeed the author of the anonymous letters. Further analysis revealed additional uncommon expressions that appeared across both sets of texts, reinforcing the forensic linguist's conclusion.

#### *D. The JonBenét Ramsey case*

In this high-profile case involving the murder of six-year-old JonBenét Ramsey, a ransom note left at the crime scene became a key piece of evidence. Forensic linguists were asked to analyse the language of the note to determine whether it was consistent with the writing style of the Ramseys or someone outside the household. While the analysis provided insights into the peculiarities of the note's language, no definitive conclusions were reached about its authorship.

### **V. CURRENT TRENDS AND FUTURE CHALLENGES**

Forensic Linguistics, on the one side, has been impacted by digital technologies and the new crimes that have spread thanks to their use. The current trends in Forensic Linguistics cover, thus, a wide range of linguistic evidence in crimes, from hate speech to scam. Also, researchers are developing new methods with the help of digital tools to be able to learn from huge amounts of data, find patterns and standardized certain analysis methods as much as possible, in spite of the wide variety of cases and their specificities. This is, in fact, one of the main challenges that Forensic linguists are facing, the development of more precise and case adjusted, but standardized, methods of analysis.

Globalization and digital technologies also bring a challenge in terms of how to interpret texts and discourses in multicultural (or intercultural) and plurilingual settings, how to distinguish between human (although computer mediated) interactions and those created by Artificial Intelligence applications, how to detect and counteract hate speech escalation or how to deal with the new forms of hypertextuality, while keeping ethical and privacy standards.

Probably, one of the main strengths that could help Forensic linguists face the future is the collaboration with other experts, such as criminologists, computer scientists or psychologists, to improve the collection and interpretation of data. Hopefully, this could also lead to the expansion of the collaboration to other fields, enriching all the disciplines involved. Undoubtedly, law enforcement officers could benefit from any advances in the area and, also, contribute to them with their empirical understanding and experience.

## VI. CONCLUSIONS

Forensic Linguistics is a discipline embedded in Applied Linguistics whose more and more solid methods are rooted into theoretical Linguistic concepts, qualitative and quantitative scientific research methods and the combined application of the notions and findings from several disciplines related to language, culture and communication, mainly. In this article, we have presented some of those roots, as well as the ways in which Forensic Linguistics has helped, can help and will continue to help law enforcement officers when facing cases where linguistic evidence is involved.

## VII. REFERENCES

- Afarli, T., Maehlum, B. (2014). *The Sociolinguistics of Grammar*. John Benjamins Publishing House.
- Akmajian, A., Demers, R.A., Farmer, A.K., Harnish, R. (2001). *Linguistics. An introduction to language and communication*. The MIT Press.
- Alonso, L.E., Fernández Rodríguez, C.J. (2006). *Roland Barthes y el análisis del discurso*. EMPIRIA. Revista de metodología de las Ciencias Sociales, 12, pp. 11-35.
- Austin, John L. (1962). *How to Do Things with Words*. Oxford University Press.
- Basim, Y.J. (2012). *Author attribution in suicide notes: evidence from applied linguistics*. Comparative Legilinguistics, 10, 7-22.
- Beaugrande, R. y Dressler, W. (1981). *Introduction to text linguistics*. Routledge.
- Caplan, S. (2018). *Word learning as category formation*. Proceedings of the Society for Computation in Linguistics, 1(1), 174-177.
- Celma, M.P., Ruiz, C. (2021). *Una aproximación a la escritura de Miguel Delibes desde la estilometría*. Tonos digital: Revista de estudios filológicos, 41, 1-13.
- Chaski, C.E. (1999). *Linguistic authentication and reliability*. National Institute of Justice.
- Chomsky, N. (1997). *Language and problems of knowledge*. Teorema, 16(2), 5-33.
- Cicres, J. Gavalda, N. (2014). *La lingüística forense: la llengua com a evidencia*. Revista de Llengua i Dret, 61, 60-71.
- Crespo, M. (2017). *PRESEEA y su aporte a la creación de perfiles lingüísticos en Lingüística Forense*. Linred: Lingüística en la Red, 15.
- Crystal, D. (2001). *Language and the internet*. Cambridge University Press.
- Estival, D., Gaustad, T., Hutchinson, B., Radford, W. (2007). *Author profiling for English and Arabic emails*. Proceedings of the Australasian Language Technology Workshop 2007, pages 21-30.
- FBI (2020). *Internet Crime Report 2020*.
- Foucault, M. (1968). *Las palabras y las cosas. Una arqueología de las ciencias humanas*. Siglo XXI.
- Frias, A. (2009) *Distribución de frecuencias de la longitud de las palabras en español: aspectos diacrónicos y de estilometría*. A survey of corpus-based research, 756-770.
- Gibbons, J., Turell, M.T. (eds.). (2008). *Dimensions of Forensic Linguistics*. John Benjamins.
- Kilgariff, A., Rundell, M. (2016). *Lexical profiling software and its lexicographic applications – a case study*. Euralex proceedings, 807-818.
- Moreno Sandoval, A. (2019). *Lenguas y computación*. Síntesis.
- Ramírez, M. (2017). *Antecedentes de la lingüística forense: ¿desde cuándo se estudia el lenguaje como evidencia?* Pragmalingüística, 25, 525-539.
- Searle, J.R. (1970). *Speech acts: an essay in the philosophy of language*. Cambridge University Press.
- Shuy, R.W. (2016). *The language of fraud cases*. Oxford University Press.
- Waller, S. (2015). *Cohesion is still not coherence, so what is?* English teaching in China, 6, 31-35.
- Yule, G. (2010). *The study of language*. Cambridge University Press.