

**How to Cite:**

Petrova, I., Sezonov, V., Perlin, S., Sezonova, O., & Pidlybna, A. (2021). Linguistic document research technologies (forensic and procedural aspects). *Linguistics and Culture Review*, 5(S4), 1464-1482. <https://doi.org/10.21744/lingcure.v5nS4.1767>

## **Linguistic Document Research Technologies (Forensic and Procedural Aspects)**

**Iryna Petrova**

National Scientific Center “Hon. Prof. M. S. Bokarius Forensic Science Institute”,  
Kharkiv, Ukraine

**Viktor Sezonov**

Kharkiv Scientific Research Forensic Center of the Ministry of Internal Affairs of  
Ukraine, Kharkiv, Ukraine

**Stanislav Perlin**

Kharkiv Scientific Research Forensic Center of the Ministry of Internal Affairs of  
Ukraine, Kharkiv, Ukraine

**Olga Sezonova**

Kharkiv National University of Internal Affairs, Kharkiv, Ukraine

**Alina Pidlybna**

Kharkiv National University of Internal Affairs, Kharkiv, Ukraine

**Abstract**--Linguistic examination is a modern type of forensic examination, which is directly related to the study of the products of speech activity in order to establish facts (circumstances) of evidentiary significance in a particular (criminal, civil and administrative) case. Although linguistic knowledge is often useful, author identification has not yet reached a significant level of reliability. The main purpose of the study is to investigate the current state of legal regulation of document linguistic examination in Ukraine and to disclose the criminal and procedural aspects of linguistic technologies of document examination. The research process involved using a combination of general scientific, philosophical and specialised scientific methods, which are in line with the aims and objectives set by the authors. The philosophical basis of the methodological paradigm of the study is the dialectic as a philosophical learning method. The author's conclusion provided a conceptual understanding of the implementation of document linguistic examination in Ukraine. The review of implementation and evolution of legal regulation in the outlined area demonstrates that Ukraine in directions of legal regulations reform attempts to be

oriented towards implementation of progressive international approaches to forensic linguistic examination, paying attention to the urgent need to develop methodological approaches to examination of physical products containing destructive information due to the growing threats of virtual space. The relevant legal mechanisms suggested by the scholars can be generalized, adapted and implemented into the Ukrainian legislation, which regulates the procedure of conducting linguistic examination of documents. The study suggests the gradual introduction of positive foreign practices of adapting innovative technologies of linguistic research to the domestic realities of today.

**Keywords**--document examination, expert technology, forensic evaluation, linguistic examination, procedural.

## Introduction

Forensic linguistic examination has become one of the practices that is developing rapidly within the judicial process in the third millennium, and this trend is due to the dramatic increase in multimodal communication among individuals and organisations (Seyari & Bagheri, 2019). Forensic linguistics is an interdisciplinary field of applied linguistics that mainly analyses language in two directions (Ahmed, 2021). Firstly, from the perspective of offence, it may be useful in the investigation of crimes. An integral part of forensic linguistics in this field is the identification of the author of written or spoken texts. Forensic linguistics can thus be called the interface between language and law (Ahmed & Arcelus-Ulibarrena, 2021). The second area of focus is the research and use of language for judicial procedures. Forensic linguistics is a subdivision of linguistics which is particularly closely related to professional and institutional interaction in legal contexts, it is the study of language in a legal framework, with texts, both oral and written, forming the basis for the study, analysis and measurement of language (Coulthard et al., 2016). Linguistic expertise, as a field of knowledge that studies the understanding of language use in legal proceedings and develops guidelines for the provision of linguistic expert opinions (Umiyati, 2020), complements legal analysis by applying rigorous, scientifically accepted principles of language analysis to legal evidence such as emails, text messages, contracts, letters, confessions and recorded speech (Leonard et al., 2017). Nowadays, the need to simplify the processes of mechanical processing of arrays of information, including its systematisation, search automation, translation into other languages, annotation, has stimulated the emergence of computer linguistics; the development of optimal models of mass management and has confirmed the need to recognise ways of influencing the recipients' minds, which, in turn, leads to the justification of psycholinguistic approaches as an explanatory tool of science.

Linguists now apply their knowledge in areas such as: identification of the author of a written document or the speaker of an audio document, police interrogation practices, contract disputes, legal discourse, defamation, trademark infringement, copyright disputes, discrimination, commercial warning messages and various types of criminal charges such as bribery, extortion, money laundering, threats

and fraud. Practically all such cases involve written or oral evidence, which makes their linguistic analysis extremely relevant. In contract disputes, the meanings of individual words and phrases (as well as syntactic relations) can form disputes. In plagiarism cases, which are a subset of copyright analysis, the question arises whether the defendant has taken text or content from an author or company document (such as a novel, a court opinion, a screenplay or a patent applications) to another document without proper citation (Pfefferli, 1983; Biedermann et al., 2011). In copyright cases, linguistic issues may include not only simple borrowing of words but also copied discourse structure, such as a sequence of themes. In the relevant branch of the law, trademark infringement cases regularly involve linguistic similarities between a junior and a senior trademark (for example, phonological analysis can demonstrate that they sound similar, and semantic and pragmatic analysis can ascertain similar meanings). Even in cases of liability for copying a certain product, linguists will help to confirm important information, for instance by revealing that the product contained insufficient, incomprehensible information. Other types of cases in linguistic analysis can become key ones – these are cases of discrimination and defamation, where the defendant's speech can be subject to revision, for example by its meaning in context. In today's realities, "hate speech" becomes a type of malicious online content that directly attacks or promotes hatred towards a group or individual member based on their actual or assumed aspects of identity, such as ethnicity, religion and sexual orientation. With hate speech on the Internet, there is a growing interest in its automatic detection as a natural language processing task (Yin & Zubiaga, 2021).

The information obtained as a result of linguistic examination of documents must be used in the course of procedural evidence, so it must be materialised in a procedure and form defined by law – the means of evidence. Therefore, the issue of studying different sources of evidentiary information, especially new ones related to electronic means of evidence, is always relevant (Kalamaiko, 2016). Courts recognise the legitimacy of the field of forensic linguistics and entrust experts to give opinions and, in some cases, evidence. This position applies to a wide range of cases and situations. Nevertheless, although forensic linguistics has been widely introduced into judicial practice, it is still an underused tool; at the same time it can be applied to almost any case in which speech may be considered as evidence, and this certainly covers many more cases than those where it is actively tested today (Leonard et al., 2017).

The social demand for conducting forensic linguistic examinations as required by the current Ukrainian procedural legislation objectively required professional expert and linguistic support for court proceedings. Since the number of information and documentary disputes has increased, involving not only citizens, legal persons, but also the media and Internet publications, an objective need has emerged to transfer this examination to the level of a well-tested technological process, which should be carried out using unique, scientifically validated methods and reproduced linguistic technologies (Yaroshchuk et al., 2020). Expert technology in this sense is an information model that includes the task of expert investigation, methods and techniques for solving these tasks in order to establish the factual data in the process of forensic investigation (Korma, 2017). However, specialists still lack a unified, systematically organised and empirically grounded base of diagnostic and identification features necessary to make fundamental

knowledge in linguistics the basis for the development of tested technologies for solving practical forensic tasks.

The conclusions of an expert are recognised as evidence in the course of investigating a crime. They are independent of the procedural status of the person who initiated the examination, have no pre-determined force or advantages over other evidence, and are subject to verification, procedural and forensic assessment by the prosecution, the defence and the court (Shcherbakovskyi, 2013). Thus, 82 percent of the interviewed investigators and 74 percent of the judges consider expert opinion to be a stronger source of evidence than other evidence (Shcherbakovsky, 2015). The investigator, prosecutor and court may recognise the findings of the expert examination as admissible evidence if the requirements of the criminal procedure law and regulations for the appointment of the expert examination, the research methodology of the expert examination have been complied with and the findings contain reliable evidentiary information. On the basis of the aforementioned, it is highly relevant to consider key issues concerning the technology of linguistic research of documents. In order to achieve the objective in view, it is necessary to define the following tasks: 1) to consider the main points of the state of modern legal regulation regarding the implementation of speech linguistic examination in Ukraine (forensic and procedural aspects); 2) to consider a number of lawyers' recommendations regarding the implementation of technologies for linguistic examination of documents with the aim of their possible implementation in the relevant procedures in Ukraine.

### **Materials and Methods**

The methodological basis of the study was a set of subject-based principles, approaches and methods of knowledge. A combination of general scientific, philosophical and special methods of science was used in the research process, which corresponds to the purposes and objectives of the study. Thus, general scientific methods such as analysis and synthesis, induction and deduction, comparison, abstraction, etc. were used in the process of writing the paper.

Conducting a gradual historical and legal analysis of the selected topic is impossible without taking into account the transformations that have occurred not only with the object of study, but also with all the processes and phenomena related to it. The mentioned above, primarily, enables to identify and take into account all factors and conditions that have determined the evolution of technologies for linguistic research of texts in Ukraine, especially in the context of forensic and procedural aspects, and so the historical and legal method is used to determine the stages, analogues and determinants of international cooperation in the field of improving the legal regulation of the relevant linguistic studies; the problem-chronological method enabled the research text to be structured, and the empirical analysis facilitated comparison of historical facts and competent expert characteristics.

The leading method of scientific research was the dialectical method, based on two principles: 1) determinism (the principle of conditionality and interrelationship between different phenomena) 2) historicism (any phenomenon is not static, but develops in a temporal dimension). The method chosen made it

possible to focus scientific research on transformations in the technology of linguistic document research through the prism of societal development and qualitative changes in the contemporary realities of legal consciousness and law enforcement. This method also enabled the identification of the state, directions and prospects of development of scientific research and legislative developments in the field of legal regulation of the conduct and significance of document forensic linguistic examination.

The methodology of the study is also based on a systematic approach, due to the specific nature of the paper and related to the use of general scientific and specifically scientific methods. The historical method is used to investigate the formation and development of complex forensic linguistic studies, the stages of their formation and the research technology. The formal-logical method used when working on the study made it possible to determine the logic of forensic expert linguistic research, its specificity, the availability of special knowledge among experts and the specifics of implementing expert conclusions in this field. Statistical and comparative legal methods were used to identify areas of improvement in the legal framework and the practice of applying the law regulations governing these issues by forensic experts and the persons (authorities) who order linguistic research of documents.

The methods of grammatical examination, interpretation and comparison of individual categories and legal provisions were also applied, helping to identify gaps and other shortcomings in the current Ukrainian legislation and providing suggestions for its improvement. The method of legal analysis was used when studying the relevant provisions of forensic science, criminal and civil procedural legislation, departmental regulations on the issues of expert support of justice and preliminary investigation; the forecasting method – when developing suggestions for improving the current legislation in the field of linguistic research of documents, the practice of its application.

Induction enabled general conclusions to be drawn on trends in the development of document linguistic research technology by studying specific statistical data. The deductive method was applied when dealing with legal provisions that directly regulate the admission of the results of document linguistic research as proper and admissible evidence in various proceedings. The specific sociological method used in the study of law enforcement materials made it possible to assess the real implementation of legal regulations and basic technological aspects of conducting research with texts in the life of society, the efficiency of legal regulation in the studied area, the impact of the results of linguistic research on law enforcement practice and the state of law and order, on the qualitative characteristics of legal culture and legal consciousness. As part of the analysis of legislative recognition of the results of linguistic research on documents, the following ways of interpreting the law were taken into account: logical, systematic, grammatical, historical and functional.

Nowadays, with globalisation and increasing legal ties between different states, the importance of the comparative method is constantly growing, as it holds great methodological potential. For this reason, the study used the special legal method of knowledge – the comparative law method (comparative method). In particular,

macro comparisons were used in the context of the significance of testing the latest foreign technologies of linguistic research of documents in Ukraine.

## Results

In Ukraine, the commissioning and conducting of forensic examinations and expert studies (Law of Ukraine No. 4038-XII, 1994) is regulated by the relevant departmental instructions under which linguistic examination of speech is a component of forensic expertise and is carried out by forensic experts. This forensic expertise is performed by state specialised institutions (Law of Ukraine No. 4038-XII, 1994), which include research institutions of forensic expertise of the Ministry of Justice of Ukraine; research institutions of forensic expertise, forensic medical and forensic psychiatric institutions of the Ministry of Health of Ukraine; expert services of the Ministry of Internal Affairs of Ukraine, Ministry of Defence of Ukraine, Security Service of Ukraine and State Border Guard Service of Ukraine. According to the Scientific and Methodological Guidelines on the issues of preparation and commissioning of forensic examinations and expert studies (Order of the Ministry of Justice of Ukraine No. 53/5, 1998), the linguistic examination of speech includes linguistic examination of written and oral speech, with the objects of examination being the products of human speech activity reflected in written and oral form, fixed in the (audio, video) recordings. It is worth noting that since 2017 Ukraine has introduced the legal institution of electronic evidence in the Economic Procedural Code of Ukraine, the Civil Procedural Code of Ukraine (2012) and the Code of Administrative Procedure of Ukraine (2017).

The modern division of linguistic examination according to the complex "object – subject – expert technology" framework is as follows: naming expertise (examination of names); linguistic examination of regulatory legal acts and documents; linguistic examination of advertising texts; extremist materials; defamatory materials (Chang et al., 2008; Wen, 2010). Linguistic examination is most often commissioned in cases of bribery, extortion, threats, drug trafficking, etc. (Nikishin, 2018). International linguists suggest the following tasks to be applied in the linguistic examination of speech (The Arts and Humanities Research..., 2010): Identification of the author, includes determining whether a particular person has said or written something, and which is based on an analysis of their idiolect or particular patterns of language use (vocabulary, word combinations, pronunciation, spelling, grammar); forensic stylistics, which subjects written or spoken materials (or both) to scientific analysis in order to determine and assess content, meaning, identify the speaker or identify the author when plagiarism is detected; discourse analysis concerns analysis of written, spoken or sign language as well as any meaningful semiotic event; linguistic dialectology – the methodological study of dialects based on anthropological information; forensic phonetics deals with producing an accurate transcription of what is said, which can reveal information on the social and regional origin of the speaker; forensic transcription in the form of written documents as well as video and audio recordings; within-author variations – ways in which texts by the same author differ from one another (Madea & Preuß, 2009; Bataille et al., 1999).

The object of expert linguistic research are written materials of various types: newspaper and magazine publications; speech scripts; oral speech scripts of participants in legally significant communication situations; books; legally binding documents (contracts, receipts), case materials for which linguistic examination is conducted (witness statements, applications, minutes, etc.), Internet communication texts (forums, blogs, social networks). The subject of the examination is the facts and circumstances of the case established on the basis of an examination of the regularities in the existence and use of a natural or artificial language by its speakers (Gow Jr & Im, 2004; Waxman et al., 1997).

The tasks of linguistic examination include: scientific interpretation and clarification of the meaning and origin of a word, word combination, phraseology or other linguistic unit; interpretation of the primary and secondary (connotative) meaning of a linguistic unit or a linguistic unit actualised in context; investigation of a text (fragment) to identify its semantic orientation, modality of sentences, expressiveness and emotionality of linguistic units, their formal and grammatical characteristics and semantics, the specifics of the stylistic means and techniques used. The tasks mentioned are specified according to the field of application of special linguistic knowledge in different case categories. Unlike the examination of audio speech, which identifies the acoustic characteristics of language utterances, including identifying the author of a speech message, the written text is analysed according to a range of features: grammatical, semantic, pragmatic, which involves the use of traditional methods and approaches grounded in linguistic science. Linguistic examination of the text establishes the relationship between the person, the level of language, communicative writing competence, characteristics, purpose and effects of language; it makes full use of all the achievements of basic science: the concept of language identity, typology of texts, literary language and spontaneous speech, the correlation of written and spoken forms of expression, norms and development in the process of communication, etc. The result of a linguistic examination is the opinion of an expert, the main part of which is structured as a response to court enquiries or an information dispute party and contains conclusions on the disputed expressions of a conflictogenic text or on the text as a whole (Bogoslavska, 2020).

The Ukrainian procedural legislation regulates the activity of experts by types of examination, which include primary, additional, repeated, commission and complex examinations. The relevant expert investigations are carried out with the help of special knowledge and the use of methods of forensic science and forensic examination (Ullman, 2004). The basis for conducting an examination should be: a procedural document appointing the examination, or an agreement with an expert or an expert institution, with mandatory references to the relevant articles of law; a list of issues to be solved; and the objects to be examined (Ullman & Pierpont, 2005). The results of expert examinations and expert studies shall be provided in a written document – expert opinion, and the period of expert studies shall not exceed 90 calendar days. Linguistic experts should not draw conclusions about legal issues. The competence of the expert linguist does not include the legal assessment of the texts examined, i.e. the expert does not draw conclusions about the presence of insults, slander, humiliation of the honour and dignity of a person (group of persons), about obscene word forms, etc., and does not classify materials as extremist, does not establish the extremism signs.

Regardless of how strongly the linguist is convinced that the defendant is innocent, he or she should limit his or her opinion to stating the degree of probability. Moreover, forensic linguists must always remain impartial, as they implement the law in the role of experts and cannot, under any circumstances, take the side of either the defence or the prosecution. In this case, the expert needs to find out what form the information is in and provide a statement or value judgement.

The organisation of conducting forensic examinations and registration of their results shall be carried out in the manner prescribed by the Criminal Procedural Code, the Civil Procedural Code, the Economic Procedural Code of Ukraine, the Code of Ukraine on Administrative Offences, the Code of Administrative Court Procedure of Ukraine, the Laws of Ukraine "On Forensic Examination" and "On Enforcement Proceedings". Thus, for instance, with regard to criminal proceedings in accordance with the requirements of Article 94 of the Code of Criminal Procedure of Ukraine (2012), the conclusions of the expert as a source of evidence for the crime under investigation are subject to procedural assessment from the viewpoint of relevance, admissibility and credibility (Peniro & Cyntas, 2019). According to part 2 of Article 84 of the Code of Criminal Procedure of Ukraine (2012), the procedural sources of evidence are testimonies, material evidence, documents, and expert opinions (Gabidullina et al., 2021). The documents submitted for linguistic expertise serve as physical evidence in accordance with Article 98 of the Code of Criminal Procedure of Ukraine (2012). In order to clarify the validity of the facts stated in the expert opinion, it is necessary to analyse the course and results of the expert examination. The investigator, prosecutor and court may recognise the findings of an expert examination as reliable evidence if the expert has conducted a full examination of the object within the limits of his or her task and has provided reasonable and objective written conclusions the evidential value of which directly or indirectly confirms the facts, circumstances relevant to the investigation of the crime.

Linguistic examination of written speech in Ukraine is divided into two types – authorship and semantic-textual investigations. Linguistic examination of spoken language consists of identification and diagnostic studies of a person's oral speech and semantic studies of spoken language (Suryasa, 2019). In accordance with the tasks set, the linguistic expert uses standard questions, algorithms and methods of examination. With regard to authorship examination, linguists study, for example, grammatical structures in which changes in punctuation schemes between texts can indicate different authors; semantics, which investigates how meaning is formed, for instance when understanding written text; sociolinguistics, which analyses differences in the use of language among different social groups. The procedure for analysing the linguistic features of a text, as adopted in linguistic expertology, includes the following aspects of studying its features: lexical-phraseological characteristic features, grammatical and stylistic features. When features are present at any one or more levels of analysis, the question of their relevance is decided: the mass, the persistence that distinguishes the abilities. A text of at least 100 words in the form of free, conditionally free and experimental samples of the written speech of the person to be identified must be provided for the purposes of the authorship examination. Attributional methodology is often used, combining qualitative and quantitative



approaches to text attribution. Qualitative analysis is a study of an author's individual style as a holistic construct, which represents the linguistic competencies of the author of a text and the level of proficiency in them (Yunusova, 2021). With this type of analysis it is possible to create a linguistic model of an author's individual style. Quantitative analysis is the extraction of statistical data from a text array. This data helps to make the attribution examination more complete, comprehensive and objective (Khomenko, 2019).

It is worth noting that in today's realities there is an increase in the number of crimes involving digital technologies. These online offences are facilitated due to the anonymity and reach allowed on social media platforms. Proper identification of individuals who post threatening, defamatory or false messages online is of paramount importance to investigators as it can help protect those who are attacked. The authorship establishment is based on the correct grouping of texts created by one author by highlighting textual features characteristic of that author. These characteristics are usually related to grammatical structure and are deeply rooted in each person's individual authorial style. Establishing authorship in social media is challenging as there is no "text imprint" or clear pattern of language use. Examples of text that are often analysed by forensic linguists include blackmail letters, confessions, wills, suicide letters and plagiarism, online texts such as online chat logs of sexually explicit content between middle-aged men and underage girls, etc. Author profiling is used when researching social bots, the most common of which are Twitter bots. Social bots are considered a threat, given their commercial, political and ideological influence. The bots designed to look like human accounts can mostly be identified by the information in their profiles, such as username, profile picture and posting time. However, the task of diagnosing bots based solely on textual data is a much more complex one, requiring the methods of author profiling. This usually involves classification tasks that are based on semantic and syntactic characteristics.

The subject of linguistic examination is also the determination of the text's semantic content and the identification of criminally relevant information in the text (Koutchadé et al., 2018). The linguistic investigations most in demand today are those in cases of verbal imagery, protection of honour, dignity and business reputation as well as in cases of extremist crime (Order of the Central Department of the Security Service of Ukraine No 215, 2020). Extremist crime cases may include the presence or absence in the text of a publication or document of a threat to commit a particular crime (murder, terrorist act), calls to overthrow the constitutional order, to seize state power, to commit deliberate acts with the aim of changing the territory or state border, statements aimed at inciting national, racial or religious hatred and enmity, expressions that can be classified as propaganda for war, fascism, etc.

The materials are provided to the expert linguist for examination in the form of printed texts, oral speech materials on digital media (CDs, flash cards, computer hard drives). If the text under investigation is printed in a periodical as well as in a book, a complete copy of the publication or a good quality copy thereof shall be submitted for examination. In the case of a text document from the Internet, a complete digital copy of the page(s) containing the text for examination, including illustrations, or a printout of the page(s) in question, indicating the moment when

the seizure was made, should be submitted for examination (Sokolova-Vysochyna, 2018). In addition to verbal data, the texts under examination may include non-verbal components (graphics, symbols, drawings, photographs, etc.), which also serve as objects of linguistic examination due to the fact that they contain additional important information. The expert also has to be provided with a protocol of inspection and listening to the audio recordings of the conversations under investigation with their typed text, drawn up in accordance with the procedural requirements.

The texts of oral speech (speeches at rallies, conferences, interviews) can be provided for linguistic investigation. The amount of language material should be approximately 5-10 minutes of the test person's speech. When it comes to linguistic examination of spoken language, speaker identification is the most common task of the relevant investigation. This technology involves comparing language models to assess the speaker's identifiability. Comparisons are made by a comprehensive examination of the phonetic, linguistic and acoustic components of speech and can be supplemented by automated procedures using speaker recognition technology. A comprehensive examination of an oral audio recording is carried out jointly by an expert linguist and an expert engineer in the field of video and sound recording. The examination of oral statements involves analysing problem parts of the recordings to determine what was said. This can be an argument, for example, through the noisy environment, overlapping language, accents, dialects. Improvements can be made to help with this type of work, which include the use of digital filters and dynamic processors in recordings to reduce background noise and improve speech intelligibility. Profiling technology involves analysing the recordings of language unknowns to obtain information on the regional and social origin of the speaker. In this case, background noise can also provide information on where and when the recording was made. Forensic linguists can analyse interrogations recorded by the police, which can be used to decipher the facts about a person's guilty plea – whether or not it was consciously admitted. It may also be established that the person "only underwent interrogation" or only "understood the conversation" during the investigation or interrogation. Recorded interrogations are presented as important evidence in court. Thus, the technique of dialogic interview analysis is used by forensic linguists who can prove someone's guilt or innocence. This analysis can also identify potential inconsistencies in the interview process, making the recordings inadmissible evidence in court. The defence can thus demonstrate that the recorded speech does not necessarily indicate the guilt of the defendant.

Increasingly, forensic linguistic examination uses a combination of software, experience and statistical approaches in its analysis. Computer scientists have developed technologies to automate linguistic and phonetic analysis. These approaches do not require an expert to implement them, but do require expert interpretation (Forensic Language Analysis, 2015). Computational technology extracts biometric information (based on the physiology of an individual's vocal tract) from speech samples. These samples can be compared with others to perform an automatic comparison (sometimes known as recognition) or to check whether the same person is speaking in several samples (verification).

Technological advances in recent decades have opened up new possibilities for forensic linguistic analysis: new forms of online interaction have required more advanced forms of computer-mediated discourse analysis as well as synchronous and direct forms of communication, such as those provided on online platforms, which have enabled users to communicate with almost anyone, anywhere, at any time, from any mobile device using online communication (Sousa-Silva, 2019). In this case, corpus linguistics technology is applied to the analysis of meaning (written and spoken): software processes hundreds of documents such as extremist texts on the Internet and identifies key words, phrases and themes that can be used for intelligence gathering and investigative purposes.

Knowing the author of the information can help better determine its reliability, which is critical in light of the increasing amount of online misinformation spread by so-called trolls, bots and other online agitators. A great deal of work on implementing machine learning algorithms by computational and statistical methods to determine authorship of textual works based on writing style: word choice, use of punctuation marks, idiosyncratic grammatical errors, and in the latest digital texts – the use of emoticons – is carried out by, for example, PAN, PolEval – active international communities in which computational approaches to author identification are developed and evaluated (Ordoñez et al., 2020).

Based on the aforesaid, it can be concluded that at the present stage of development of Ukraine the activity of forensic experts in conducting linguistic examination of speech is regulated at the legislative level. The relevant expert actions are carried out with the help of special knowledge and the use of methods of forensic science and forensic examination. In procedural terms, the conclusions of an expert investigation are considered reliable evidence, if the expert has conducted a full examination of the object within the scope of his or her task, and has provided sound and objective written conclusions, the evidential value of which directly or indirectly establishes the facts, circumstances relevant to the investigation of the crime. The linguistic investigations most in demand today are those in cases of verbal imagery, protection of honour, dignity and business reputation as well as in cases of extremist crime. Increasingly, forensic linguistic examination uses a combination of software, experience and statistical approaches in its analysis.

## **Discussion**

Forensic linguistic analysis can be of value in almost any case where oral or written speech utterances may be considered as evidence (Leonard et al., 2017). Recently, researchers have emphasised the important role of linguists involved in forensic linguistics (Houtman & Suryati, 2018), paying special attention to the interdisciplinary nature of the investigations and highlighting the importance of pragmatics as the linguistic basis of professional activity. Moreover, practitioners emphasise the importance of collecting and providing accurate data in terms of applied characteristics (Meluzzi et al., 2020). Conclusions based on the performed linguistic analysis of documents acquire the status of reliable evidence through further research of disputed meanings, the application of authorship analysis techniques in response to new needs (e.g. cybercrime investigations) and attempts

to develop new theories, such as synthesis of author speech (Grant & MacLeod, 2018).

An important aspect of forensic linguistics is the identification of the author of a particular written text. Certain problems should be highlighted in this aspect. Thus, documents available for forensic analysis are sometimes very short extracts. These can be ransom notes or letters containing blackmail materials. Consequently, these documents can hardly be used as reliable identification of the author (Ahmed, 2021). Another problem is that it remains to be identified and determined which linguistic features can be reliable indicators for determining authorship. The credibility of linguistic characteristics is also a big issue in the field of forensic linguistics. There is an acute need to identify reliable attribution methods for the application in short texts. However, several important issues remain open regarding "author's indication", in which the most important issue is the required length of the text. Various investigations and reports have shown fruitful results with short criminal texts that are less than 1000 words in length, but forensic linguists have not yet been able to determine the minimum text length that can be recognised as reliable to establish the author (Najam-Us-Sahar, 2020).

One of the controversies in the application of forensic linguistics is the divergence identified among forensic linguists. This divergence relates to the accuracy of the speaker's identification. There may be certain "vocal distortions" in the speaker's voice sample provided for verification, and it can be very difficult for specialists to deal with this. The probability of identification in such a sample is lower than in conventional voice samples. This problem could be solved if the experts suggest to the court to oblige the suspect to give a voice sample that repeats certain phrases in a natural conversational voice (Ahmed & Arcelus-Ulibarrena, 2021). Voice samples obtained through such specific recommendations are usually very suitable for comparison purposes, but there is a universal standard for the number of words required for identification. It can range from 10 to 20 words for various situations (Machado et al., 2019). Using spectrographic voice identification technology, it can be stated that the variability within the technical means that records and reproduces oral speech differs from the variability between different technical means. This means that differences in the same expression of different speakers are noticeable, but scientific theory and data do not consider this assumption to be confirming. There are certain points of view based on the actual number of errors associated with professional judgements about them. Moreover, "fragmented experimental results" also affect these perceptions of the actual error rate. It should be noted that in this case "objective data" is not given much significance, one could say that it is not representative of the results when it comes to forensic examinations (University of York, 2021).

In terms of the international development of forensic linguistics, the following challenges arise before attaining appropriate status: the integrated study of forensic linguistics / language and law through different judicial systems and geographical boundaries; the development of repeatable analysis techniques for use in expert conclusions to ensure the internal and external validity of the investigation; extensive detailing of codes of good practice and conduct; cooperation of the International Association of Forensic Linguistics (IAFL) with

other forensic science associations and societies (Ahmed & Arcelus-Ulibarrena, 2021). It will also be important for linguists in the era of international courts to understand the practice of international law discourse and become familiar with the customs of other countries' legal systems. It is likely that in the future an increasing number of those seeking to enter the field of forensic linguistics will have additional qualifications in areas such as law and will gain a better understanding of scientific techniques, methods and presentations (Ariana et al., 2014).

Forensic linguistics can be applied simultaneously in civil, administrative and criminal cases to support legal conclusions; hence it is necessary to focus on each aspect of the latest technologies that are changing traditional methods of investigation and interrogation in pre-trial investigations. When oral or written speech utterances can be regarded as evidence, every linguistic investigation can be relevant (Syam, 2018). It should also be noted that there is an urgent need to develop methodological approaches to the investigation of physical products containing other destructive information due to such growing threats of virtual space as: trolling, bullying, promotion of suicidal behaviour (including "death groups" in social networks), propaganda of underground culture, the cult of violence and cruelty (including prison culture), destructive propaganda using "fan-fiction" content, etc. (Nikishin, 2018).

There is now an acute need to develop a methodology for the forensic authorship examination of printed texts, taking into account the improvement of already existing authorship examination techniques developed with handwritten texts (Goloborodko, 2020). As a result of the fact that typed texts have features of composition (automatic correction of errors, intervention by another individual (performer) etc.), scholars insist that it is necessary to investigate and identify a set of features that are stable for the author of a typed text, including graphic ones (use of emoticons, other markings etc.). Assessing the validity and reliability of corpus procedures is very challenging, as data on the linguistic and phonetic levels of the population are limited, making it difficult to establish the prevalence of features. This means that conclusions cannot be expressed statistically or with the same certainty as in other fields of forensic science. In linguistics, experts use their knowledge and experience as well as computational methods in their analysis; in phonetics, experts use their software to process the relevant sounds and speech. However, since variability within the speaker makes each linguistic expression unique, the expert cannot draw definite conclusions about, for instance, the extremist orientation of texts. Computational procedures can be used in some circumstances and give numerical conclusions that also express a degree of certainty (Forensic Language Analysis, 2015).

Computational methods in linguistics have developed considerably in recent decades. The increase in computing power of the computer, together with the increasing attention of computer scientists to the processing of natural language, has enabled more in-depth research into computational and computer-linguistic analysis. Sophisticated computing systems and models have been developed that allow large volumes of linguistic data to be analysed with little human intervention, at a pace and degree of efficiency with which linguists can scarcely compete (Sousa-Silva, 2019). However, according to the scholar, this does not

mean that linguists cannot be – or cannot act as – computational linguists, rather the opposite; even if linguists do not have the advanced programming skills of computer scientists, they have the knowledge necessary to assess the cost of computing resources under certain circumstances and choose the most appropriate computing tools for solving a particular linguistic problem. This is particularly important in forensic contexts where linguists, in addition to reporting their analysis, have to justify their findings scientifically and provide transparency for court purposes.

The development of machine learning techniques and eventually artificial intelligence (AI) raises new questions for forensic linguists. In addition to PAN contests over the years, PolEval has organised a task aimed at identifying harmful tweets in general and identifying the type of harm (cyberbullying or hate speech). If, on the one hand, AI in particular will become increasingly competent at producing human-like texts, then on the other hand (computational) forensic linguists will face the need to develop, test and refine their methods and techniques to solve the even greater problem of forensic linguistic examination arising from the increasing complexity of computer systems (Sousa-Silva, 2019). Since AI systems operate like black boxes, the results of their analysis cannot be explained – and certainly not to the extent and with the level of transparency required by the courts; however, they can play a key role in the context of an investigation (Ionova, 2017). If machines are able to generate human-like text, forensic linguists will have to be able to distinguish between human-generated texts and machine-generated texts. Moreover, as the scholar notes, forensic linguists may need help on cases of machine-generated text to establish whether this text has some resemblance to the text production of whoever is controlling the system, or vice versa, or whether the text was created by a machine to resemble someone else's text.

In the case of evaluating electronic means of proof, it would be more efficient to prepare for the process an electronic evidence bases on electronic media and in printed form, provided that the relevant data allows for such actions (Kalamaiko, 2016). As a consequence, this situation will enable the court to establish the affiliation of the evidence, and other persons involved in the process will also have an opportunity to examine the case materials. The legal scholar identifies three options for preserving electronic evidence and emphasises the need to avoid losing important information for the case electronically and to avoid errors in procedural implementation: out-of-court, contractual and judicial, and makes the following suggestions: to increase the activities of the notarial system to certify the use of electronic means of proof by amending the legal framework, which already has relevant blanket rules; to introduce sanctions for spoiling evidence in the form of monetary and court decisions on the existence of a specific fact in favour of the other party into the legislation (Kalamaiko, 2016).

In order to solve the problem of assessing the admissibility and credibility of the results of the expert examination, the investigator, prosecutor, defence counsel and judge can use the following forensic recommendations: to find out whether the expert has scientific, technical or other special knowledge necessary to give a conclusion on the objects submitted for examination; to check whether the expert has been warned of the relevant criminal responsibility; to check whether there

are circumstances precluding the possibility of expert examination or whether the expert is subject to recusal; to establish whether the expert understood the task correctly and did not go beyond his or her competence; to find out whether the expert informed the prosecution, defence or court about the impossibility to carry out the examination due to lack of necessary knowledge or insufficient objects; to find out whether the expert requested to provide additional materials and samples for examination; to check whether the expert has received permission from the initiator of the examination to fully or partially destroy the object of expert examination or to change its properties; to check whether the expert has used modern science and technology, means and tools, methods and techniques in carrying out the examination; to find out whether he or she personally conducted the examination of objects; to check whether the expert used special knowledge when forming the conclusion; logically justified inferences based on the results of the examination; to check whether the expert indicated in the conclusion the information revealed during the examination, concerning which he or she received no questions, but which were important for the investigation of the crime (Shcherbakovskiy, 2013).

As a result of the discussion, it can be argued that the conclusions based on the conducted linguistic analysis of documents acquire the status of reliable evidence through further research into disputed meanings, the application of authorship analysis techniques in response to new needs and an attempt to develop new theories. It will be important for linguists in the era of international courts to understand the practice of international law discourse and become familiar with the customs of other countries' legal systems. It should also be noted that there is an urgent need to develop methodological approaches to investigate physical products containing other destructive information, due to the growing threats of the virtual space. The investigator, prosecutor, defence counsel and judge can use the forensic recommendations suggested by the scientists to solve the problem of procedural verification of the results of the expert examination and the objectivity of the conclusions.

## **Conclusions**

Science is becoming increasingly important in relation to the law, and forensic linguistics is one of the fields where investigations lead to advances that are increasingly being used to uncover crimes. Forensic linguistics under many circumstances provides powerful methods and techniques to substantiate criminal charges or verdicts. Forensic linguistic expertise can also be used in civil cases to support or appeal legal conclusions. At the present stage of Ukraine's development, the activities of forensic experts in conducting linguistic examination of speech are regulated at the legislative level. The relevant expert investigations are carried out with the help of special knowledge and the use of methods of forensic science and forensic examination.

In procedural terms, the conclusions of experts are considered reliable evidence, if the expert has conducted a full examination of the object within the scope of his or her task, and has provided sound and objective written conclusions, the evidential value of which directly or indirectly establishes the facts, circumstances relevant to the investigation of the crime. Moreover, the identification of

individuals who post threatening, defamatory or false messages online is of paramount importance to investigators as it can help protect those who are attacked.

When performing linguistic examination of written language in Ukraine, authorship and semantic-textual studies are carried out. Linguistic examination of spoken language involves identification and diagnostic studies of a person's oral speech and semantic studies of spoken language. In accordance with the tasks set, the linguistic expert uses standard techniques and algorithms for the examination of documents. The linguistic investigations most in demand today are those in cases of verbal imagery, protection of honour, dignity and business reputation as well as in cases of extremist crime.

At the present stage in the implementation of linguistic examination, the following technologies are mainly used: speaker comparison, examination of controversial statements, profiling technology, technology for dialogic analysis of interviews, technology for the automation of linguistic and phonetic analysis, corpus linguistics technology. Increasingly, forensic linguistic examination uses a combination of software, experience and statistical approaches in its analysis. Furthermore, the study has shown that there is currently an acute need to identify reliable attribution methods for use in short criminal texts.

Particular attention should be paid to the international development of forensic linguistics, the practice of international law discourse and familiarity with the customs of other countries' legal systems. It should be noted that there is an urgent need to develop methodological approaches to investigate physical products containing destructive information, due to the growing threats of the virtual space. Therefore, there is now an acute need to develop a methodology for the forensic authorship examination of printed texts, taking into account the improvement of already existing authorship examination techniques developed with handwritten texts.

According to scholars, in the case of evaluating electronic means of proof, it would be more efficient to prepare for the process an electronic evidence base on electronic media and in printed form, provided that the relevant data allows for such actions. Legal scholars suggest the measures of providing electronic means of proof to prevent the loss of case-relevant information in electronic form, as well as for the purpose of its proper procedural design. The investigator, prosecutor, defence counsel and judge can use the forensic recommendations suggested by the scientists to solve the problem of procedural verification of the results of the expert examination and the objectivity of the conclusions.

## References

- Ahmed, H. R. (2021). The Role of Forensic Linguistics in Crime Investigation: Uses in Legal Proceedings. *ANGLISTICUM. Journal of the Association-Institute for English Language and American Studies*, 10(2), 23-31.
- Ahmed, H., & Arcelus-Ulibarrena, J. M. (2021). The role of forensic linguistics in understanding legal documents in court proceedings: Controversies and future



- perspectives. *Electronic Research Journal of Social Sciences and Humanities*, 3(II).
- Ariani, M. G., Sajedi, F., & Sajedi, M. (2014). Forensic linguistics: A brief overview of the key elements. *Procedia-Social and Behavioral Sciences*, 158, 222-225. <https://doi.org/10.1016/j.sbspro.2014.12.078>
- Bataille, M., Crainic, K., Leterreux, M., Durigon, M., & de Mazancourt, P. (1999). Multiplex amplification of mitochondrial DNA for human and species identification in forensic evaluation. *Forensic science international*, 99(3), 165-170. [https://doi.org/10.1016/S0379-0738\(98\)00185-6](https://doi.org/10.1016/S0379-0738(98)00185-6)
- Biedermann, A., Taroni, F., Bozza, S., & Mazzella, W. D. (2011). Implementing statistical learning methods through Bayesian networks (Part 2): Bayesian evaluations for results of black toner analyses in forensic document examination. *Forensic science international*, 204(1-3), 58-66. <https://doi.org/10.1016/j.forsciint.2010.05.001>
- Bogoslavskaya, M. O. (2020). Current issues of modern linguistic examination. *Theory and Practice of Forensic Science and Criminology: Materials of the III All-Ukrainian Scientific-Practical Conference*.
- Chang, C. W., Wu, C. R., & Chen, H. C. (2008). Using expert technology to select unstable slicing machine to control wafer slicing quality via fuzzy AHP. *Expert Systems with Applications*, 34(3), 2210-2220. <https://doi.org/10.1016/j.eswa.2007.02.042>
- Coulthard, M., Johnson, A., & Wright, D. (2016). *An introduction to forensic linguistics: Language in evidence*. Routledge.
- Gabidullina, A., Sokolova, A., Kolesnichenko, E., Zharikova, M., & Shlapakov, O. (2021). Metonymy in scientific linguistic discourse. *Linguistics and Culture Review*, 5(S4), 71-83. <https://doi.org/10.21744/lingcure.v5nS4.1556>
- Goloborodko, K. (2020). Linguistic examination of the text: Legal and linguistic analysis. *Ukrainian World in Scientific Paradigms: Collection of Scientific Works of H.S. Skovoroda Kharkiv National Pedagogical University*. Kharkiv: HIFT.
- Gow Jr, D. W., & Im, A. M. (2004). A cross-linguistic examination of assimilation context effects. *Journal of Memory and Language*, 51(2), 279-296. <https://doi.org/10.1016/j.jml.2004.05.004>
- Grant, T., & MacLeod, N. (2018). Resources and constraints in linguistic identity performance—a theory of authorship. *Language and Law/Linguagem e Direito*, 5(1), 80-96.
- Houtman, H., & Suryati, S. (2018). The History of Forensic Linguistics as an Assisting Tool in the Analysis of Legal Terms. *Sriwijaya Law Review*, 2(2), 215-232.
- Ionova, S. V. (2017). Aspects of the study of the written text as an object of linguistic expertise. *VolSU Bulletin. Series 2, Linguistics*, 16(2), 28-38.
- Kalamaiko, A. (2016). *Electronic means of proof in civil proceedings*.
- Khomenko, A. Yu. (2019). Linguistic attributional study of short written texts: qualitative and quantitative methods. *Political Linguistics*, 2(74).
- Korma, V. D. (2017). *Scientific foundations of criminology*. Moscow: Prospekt.
- Koutchadé, I. S., Datondji, A. C., & Salami, A. (2018). Exploring textual meaning in Chukwuemeka Ike's *Sunset at Dawn*: a systemic functional approach. *International Journal of Linguistics, Literature and Culture*, 5(1), 1-11. <https://doi.org/10.21744/ijllc.v5n1.475>

- Leonard, R. A., Ford, J. E. R., & Christensen, T. K. (2017). Forensic linguistics: Applying the science of linguistics to issues of the law. *Hofstra Law Review*, 45(3), Article number 11.
- Machado, T. J., Vieira Filho, J., & de Oliveira, M. A. (2019). Forensic speaker verification using ordinary least squares. *Sensors*, 19(20), 4385.
- Madea, B., & Preuß, J. (2009). Medical malpractice as reflected by the forensic evaluation of 4450 autopsies. *Forensic science international*, 190(1-3), 58-66. <https://doi.org/10.1016/j.forsciint.2009.05.013>
- Meluzzi, C., Cenceschi, S., & Trivillini, A. (2020). Data in forensic phonetics from theory to practice. *TEANGA, the Journal of the Irish Association for Applied Linguistics*, 27, 65-78.
- Najam-Us-Sahar. (2020). *A critical forensic linguistic analysis of media discourse on court decisions in Pakistan*: thesis for the degree of master of philosophy. Islamabad: International Islamic University Islamabad.
- Nikishin, V. D. (2018). The place of forensic examination of materials of a religious nature of extremist-terrorist orientation in the classification of forensic examinations. *Journal of the OE Kutafin University*, 7/2018, 187-198.
- Ordoñez, J., Soto, R. R., & Chen, B. Y. (2020). Will longformers PAN out for authorship verification. *Working Notes of CLEF*.
- Peniro, R. ., & Cyntas, J. . (2019). Applied linguistics theory and application. *Linguistics and Culture Review*, 3(1), 1-13. <https://doi.org/10.21744/lingcure.v3n1.7>
- Pfefferli, P. W. (1983). Application of microspectrophotometry in document examination. *Forensic science international*, 23(2-3), 129-136. [https://doi.org/10.1016/0379-0738\(83\)90142-1](https://doi.org/10.1016/0379-0738(83)90142-1)
- Seyari, A., & Bagheri, M. S. (2019). Perspectives of Forensic Linguistics Research in Iran. *Journal of Language Horizons*, 3(1), 187-205.
- Shcherbakovsky, M. G. (2015). *Conducting and using forensic examinations in criminal proceedings*. Kharkiv: V dele.
- Shcherbakovskyi, M. H. (2013). *Conditions for the formation of a reliable expert report. Theory and practice of forensic science and criminalistics*. Kharkiv: Pravo.
- Sokolova-Vysochyna, Ya. A. (2018). Appointment of linguistic examination of speech in the investigation of crimes against the foundations of national security of Ukraine. In: O. G. Ruvyn (Ed.), *Current issues of standardization of forensic justice in Ukraine. Development prospects: Materials of the international scientific-practical conference, dedicated 105th anniversary of forensic examination in Ukraine and 95th anniversary of the birth of Academician M. Ya. Segai*. Kyiv: Kyiv Scientific Research Institute of Forensic Expertise of the Ministry of Justice of Ukraine.
- Sousa-Silva, R. (2019). Computational forensic linguistics: an overview of computational applications in forensic contexts. *Language and Law/Linguagem e Direito*, 5(2), 118-143.
- Suryasa, W. (2019). Historical Religion Dynamics: Phenomenon in Bali Island. *Journal of Advanced Research in Dynamical and Control Systems*, 11(6), 1679-1685.
- Syam, S. K. M. A. (2018). Aspects of forensic linguistics in policing. *Language in India*, 18, 100-111.

- Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, 92(1-2), 231-270. <https://doi.org/10.1016/j.cognition.2003.10.008>
- Ullman, M. T., & Pierpont, E. I. (2005). Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex*, 41(3), 399-433. [https://doi.org/10.1016/S0010-9452\(08\)70276-4](https://doi.org/10.1016/S0010-9452(08)70276-4)
- Umiyati, M. (2020). A Literature Review of Forensic Linguistics. *IJFL (International Journal of Forensic Linguistic)*, 1(1), 23-29.
- Waxman, S. R., Senghas, A., & Benveniste, S. (1997). A cross-linguistic examination of the noun-category bias: Its existence and specificity in French- and Spanish-speaking preschool-aged children. *Cognitive Psychology*, 32(3), 183-218. <https://doi.org/10.1006/cogp.1997.0650>
- Wen, W. (2010). An intelligent traffic management expert system with RFID technology. *Expert Systems with Applications*, 37(4), 3024-3035. <https://doi.org/10.1016/j.eswa.2009.09.030>
- Yaroshchuk, I. A., Zhukova, N. A., & Dolzhenko, N. I. (2020). *Linguistic examination*. Belgorod: BelGU Publishing House, BelSU Research Institute.
- Yin, W., & Zubiaga, A. (2021). Towards generalisable hate speech detection: a review on obstacles and solutions. *PeerJ Computer Science*, 7, e598.
- Yunusova, S. M. (2021). Paraphrases related to the language of advertising. *International Journal of Linguistics, Literature and Culture*, 7(4), 236-240. <https://doi.org/10.21744/ijllc.v7n4.1766>