Asma Asaad Thamer
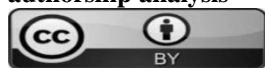
Al-Iraqia University, College of Education for Women

Email:
asmaa.thamer@aliraqia.edu.iq

A Forensic Linguistic Analysis of Authorship Identification of Selected Anonymous Comments on 'Gaza Attack' in Threads Social Media Application

A B S T R A C T

The anonymous comments have become an integral part of the online world as a double-edged sword to either protect identity for avoiding harassment or engage in harmful behaviors without taking responsibility of the consequences. Since social media platforms allow posting of anonymous comments, the present paper deploys forensic authorship identification analysis to evaluate anonymous comments based on authorship identification and authorship characterization. The research generates its data from fifteen anonymous comments in Threads social media platform and qualitatively analyzed using available techniques of the forensic linguistics that include stylistics, sociolinguistics and dialectology to unveil the possible user's profile identity of the Threads users with anonymous identity. The findings of the study show that language always provides a valuable contextual footprint to uncover not only speakers'/writers' intentional and implicational meaning but unveiling those behind the production of the discourse. It also reveals that, in authorship identification methodologies, various forms of lexical choice and semiotic linguistic choice are significant factors for not only identifying but differentiating between a larger set of anonymous authors in social media narratives which would be an important future endeavor.

تحليل لغوي جنائي لهوية المؤلف لعدد مختار من تعليقات مجهولة المؤلف عن "هجوم غزة"
في تطبيق (Threads) وسيلة التواصل الاجتماعي

م. اسماء اسعد ثامر

الجامعة العراقية – كلية التربية للبنات

الملخص:

أصبحت التعليقات مجهولة المصدر جزء لا يتجزأ من عالم الانترنت حيث تعمل كسيف ذو حدين ،أما لحماية الهوية تجنباً للمضايقات والأذى أو من أجل الانخراط في سلوكيات ضارة بدون تحمل مسؤولية العواقب. نظراً لان منصات التواصل الاجتماعي تسمح بنشر تعليقات مجهولة الهوية لذا يستخدم هذا البحث تحليل جنائي لهوية المؤلف لتقييم التعليقات المجهولة بناء على تحديد المؤلف وتوصيف المؤلف. يستمد البحث بياناته من خمسة عشر تعليقا مجهولا في تطبيق Threads للتواصل الاجتماعي تم تحليلها نوعيا باستخدام التقنيات المتاحة من علم اللسانيات (اللغة) الجنائي والتي تشمل الاسلوبية وعلم اللغة الاجتماعي وعلم اللهجات لكشف هوية الملف الشخصي المحتملة لمستخدمي Threads مجهولي الهوية. تظهر نتائج الدراسة ان اللغة توفر بصمة سياقية قيمة دائما ليس فقط للكشف عن المعنى المقصود وغير المباشر(الضمني) للمتحدثين او الكُتاب وانما للكشف عن اولئك الذين يقفون ما وراء انتاج الخطاب. كما يكشف ان في منهجيات تحديد المؤلف تعتبر أشكال اختيار المفردات المتنوعة والاختيار اللغوي السيميائي عوامل مهمة ليس فقط لتحديد وانما التمييز بين مجموعة اكبر من المؤلفين المجهولين في سرديات وسائل التواصل الاجتماعي والتي ستكون مسعى مهم مستقبلا.

الكلمات المفتاحية: علم اللسانيات الجنائية، تعليقات مجهولة ، تحديد المؤلف ، توصيف المؤلف، تحليل التأليف.

1. Introduction

The anonymous profile has become an integral part of the online world, allowing individuals to express themselves without revealing their true identities. As a double-edged sword, anonymous profile has both benefits and disadvantages. Some view it as a way to protect their privacy and avoid harassment, while others consider it a way to engage in harmful behaviors without consequences. Social media platforms have allowed anonymous comments on some topical issues as the result of this, anonymous comments become massive on the internet.

As authorship analysis gained the attention of researchers in recent years to check massive abuse of social media, analyzing social networking applications using authorship analysis will not only highlight contextual bias but providing identification and evidence of either positive or negative bias in the anonymous content.

Authorship analysis deals with 'authorship attribution/ identification', 'authorship profiling/ characterization' and 'authorship verification/ detection'. Under the 'authorship attribution/ identification', the framework identifies not only the author of a given work but also the 'authorship profiling/ characterization'. It equally attempts to formulate the author's demographic features by making inference about age, gender, or education and cultural background level on the basis of the writing style and the 'authorship verification/ detection' to determine whether the supposed author was the one who wrote the examined document (Misini, 2022).

Authorship profiling, which is the key focus of the present study, tackles with determining the characteristics of unknown authors especially their demographic details through the way they use language. Profiling questions are of extreme importance at the investigative phases of an anonymous person of message that exhibits potential anonymity. In this regard, the interrogation of the threads' comments by unmasking the dialectal, sociolinguistic or stylistic features to provide clues about the background of the person behind the unidentified comment on the social media platforms is worth academic adventure.

The essence of this paper therefore is to deploy forensic linguistic model of authorship analysis to evaluate anonymous comment on the Threads application to make inferences about who wrote the comment based on two of the three major focus of the forensic linguistic theory since all discourses are explicitly or subtly encode footprint of its author. These two major approaches are authorship identification and authorship characterization.

2. Brief on Threads Social Media Platform

Threads as social media application which was launched on July 5th 2023, was built by the Instagram team for sharing text updates and joining public conversation through comments. Upon its launch, Threads became the fastest-growing consumer software application in history, gaining over 100 million users in its first five days and surpassing the record previously set by ChatGPT. The new social app Threads, which an offshoot of Instagram is just an X (Twitter) clone (Zhong et. al, 2024). In contrast to prior other platforms, Threads was borne out of an existing parent platform, Instagram, for which all users must already possess an account. This offers a unique opportunity to study platform evolution, to understand how one existing platform can support the "birth" of another.

3. Literature Review

Authorship analysis is the framework that is based its solid foundation on robust linguistic features that include lexical features, syntax, structure features and content-specific features among others. Under authorship analysis, researchers evaluate writing characteristics and linguistic features to make inferences about who wrote it. All these features used in authorship analysis have evolved from basic lexical and syntactical ones to more sophisticated ones, such as personality traits and emotions in written or spoken discourse (Bevendorff et al., 2020). Available literatures on forensic linguistics explain that dealing with authorship issues demands careful examination and extracting linguistic features from written anonymous messages and evaluate their stylistic strategies and patterns to achieve communicative intentions.

3.1 Writing style features

Writing style features that facilitate authorship attribution fall into four categories: lexical, syntactic, structural, and content-specific. *Lexical* features can be either word- or character-based. Word-based lexical features include such characteristics as total number of words, words per sentence, word length distribution, and vocabulary richness. Vocabulary richness measures include the number of words that can be established through several statistical measures (Yule, 1944). On one hand, character-based lexical features include total number of characters, characters per sentence, characters per word, and the usage frequency of individual letters.

Syntax, as mechanism to detect authors, refers to the patterns used to form sentences. This category of features consists of the tools used to structure sentences, such as punctuation and function words *while and upon*. Usage patterns of function words can be effective features for authorship identification. For example, the difference between using the word *thus* or *hence* might seem subtle, but it can constitute a significant stylistic difference (Zhao & Zobel, 2013).

The Structural features, which deal with the text's organization and layout, have proved particularly important in analyzing online/anonymous messages (De Vel, 2001). Researchers traditionally focused on word structures such as greetings and signatures or on the number of paragraphs and average paragraph length. Although these features are important discriminators, they don't capture the additional information contained in online/anonymous messages. For example, fonts, images, and links are not writing style features per se, but they provide important insight into a writer's online style and location where the message is coming from. The use of various font sizes and colors requires a conscientious effort, making it a style marker. Similarly, embedded images and icons or links to different types of Web sites can reflect an author's technical prowess. Evaluating technical characteristics in terms of how images, hyperlinks, and audiovisual media are used isn't novel; researchers have applied it to Web sites for almost a decade. Thus, the research proposes a new subcategory of structural features, called *technical structure,* to encompass font, hyperlink, and embedded image characteristics (Chaski, 2005).

Content-specific features are words that are important within a specific topic domain. An example of content-specific words for a discussion on computers might be *RAM* and *laptop*. The rationale for content-specific words is similar to that of other word usage features but at a finer level of granularity (ibid: 2005).

Generally, Forensic linguistics is a branch of linguistics used to discover not only criminal acts but involving the application of scientific knowledge of language to better understanding of masked language data by revealing some clues that can be useful for exposing those behind the masked or disputed text (Robertson, 2021). Forensic linguistics involves the analysis of spoken and written language for legal and security purposes.

The task of the forensic linguistic is to detect or identify an author from millions of language users on the basis of the linguistic evidences in order to reach opinions about the authorship of the written linguistic evidence in the 'questioned' or 'disputed' text(s) (Coulthard 1994; Grant 2013).

In forensic linguistics, authorship analysis focuses on two broad strands that include authorship profiling and attribution. Authorship profiling deals with examining language to deduce potential demographic information about the author, such as gender, age, education, occupation, and native dialect/language (Bevendorff et al. 2022), while authorship attribution scrutinizes language to identify the text's author (Ainsworth & Juola, 2019). There is interconnectedness between authorship profiling and attribution as the authorship profiling could increase exposing detailed information about a suspected author's background through shared linguistic features. These features when identified could uncover author's demographic information in a consistent manner such as gender, age, occupation and native dialect and

since demographic information and linguistic features remain consistent within a group of authors, this will facilitate the successful execution of the attribution analysis.

In line with the above assertion, McMenamin (2002, p. 47) argues that the linguistic choices authors make reflect their individual linguistic competence and their unique combination of linguistic knowledge, cognitive associations and extra-linguistic influences: their idiolect. In other words, an individual's linguistic choices can be accounted for by their unique social and communicative backgrounds, such as where they have lived, what they have read, the education they have received, the music they have listened to, the jobs they have had, the friends they speak to etc.

McMenamin (2010) provides examples and details of markers that have been important to easily providing the true picture of who behind the write up that authorship attribution requires in the identification of writer and distinguishing from any other writer. Consider the following excerpt:

> Mr. Ramsey. Listen carefully! We are a group of individuals that represent a small foreign faction. We respect your **bussiness** but not the country that it serves. At this time, we have your daughter in our **posession**. She is safe and **un harmed** and if you want her to see 1997, you must follow our instructions to the letter. You will withdraw **S‖118,000.00** from your account. S‖100,000 will be in S‖**100 bills** and the remaining **S‖18,000** in **S‖20** bills. (Extract of JonBenét Ramsey ransom note from McMenamin 2002, p. 182)

In the analysis of the above ransom, McMenamin (2010) identified a series of idiosyncratic spellings, word divisions and ways of writing sums of money – for example, **bussiness**, **posession**; **un harmed**; **S‖118,000.00** and **S‖100** (where S‖ indicates that $ was written as an S with two vertical lines) – which McMenamin felt were collectively likely to be evidential. He then compared the style features of the ransom note with a corpus of 338 typed and handwritten texts from the American Writing Project. He isolated six variables for analysis because they occurred frequently in the comparison corpus and were ones on which Mrs Ramsey, who still remained a suspect, differed from the ransom note.

In this regard, corpus linguistics which is applied in various fields of linguistic enquiry where its quantitative and qualitative results can be used to achieve accurate results (McEnery & Wilson, 2001, pp.76–77). Therefore, corpora can be used to aid forensic authorship attribution in measuring the rarity or frequency of potential style markers identified in stylistic analysis.

4. Method of Data Collection and Analysis

This research is designed to deploy a qualitative method of analysis. From the data that will be collected, the researcher will purposively generate and select only anonymous messages for analysis, even though some of the data may contain names that are pseudonyms of the authors. The rationale behind using purposive sampling is to identify anonymous text for unmasking the authors through linguistics features they used to compose the messages but not their names or pseudonyms. The data of this research are generated from fifteen anonymous comments that are retrieved from Threads social media platform. The researcher makes sure all excerpts that formed the data are posted by Threads users with anonymous

identity so that they can be analyzed qualitatively using available techniques of the forensic linguistics that include stylistics, sociolinguistics and dialectology to unveil the possible user's profile identity of the Threads users with anonymous identity.

5. Theoretical Framework

As the study focuses on authorship analysis that is an indispensable tool in identifying suspects, different approaches are used to identify author(s) of text(s) in question, the paper adopts authorship attribution which is the process in which linguists set out to identify the author(s) of disputed, anonymous or questioned texts (Grant, 2008). This type of linguistic analysis is regarded to be an application of forensic linguistics, especially forensic authorship analysis. Authorship analysis deals with inferring information about the author of a document whose authorship is in dispute by utilizing a technique known as "authorship profiling" to deduce details about an author from their language use.

Under the framework, the study sets to identify the author of anonymous messages based on linguistic tools that include stylistic analysis, text-structure analysis and choice of words that would highlight author attribution from the perspective of forensic linguistic theory. The stylistic analysis does evaluate a person's use of language for easy attribution. The text-structure analysis provides information on how the author approaches the communication goal through the strategic means in the anonymous message. The choice of words is to explain that unique words that easily explains the authors since authorship attribution requires 'the identification of unique words that can mark author uniquely.

6. Data Analysis

It is well established by linguists that language always provide a valuable contextual footprint to uncover not only speakers'/writers' intentional and implicational meaning but unveiling those behind the production of the discourse. This is because every language user has his or her own nature of utilizing certain linguistic features of language and life experience that distinguish him/her from others and these linguistic features and life experience are always consistent (Coulthard, 2017). It is a fact that it is not easy for anyone to avoid deploying his experience in language use. Therefore, idiolect that refers as individual or group register/style of language can be glaring in anonymous language data. For instance, according to the data:

1. We can now go through the replying of this post **when he talked about Karbala and Muharam**

In the above excerpt, the underlined clause structure that '**when he talked about Karbala and Muharam'** suggested that the social media user is a Shi'ite Muslim. Because Karbala and Muharam are sacred place and month for Shi'ite Muslims for the commemoration of the martyr of Imam Hussein (A. S), the author emphasized the reverence and importance the commemorative event of the martyr of the Iman Hussein (A.S.) in Karbala in Muharam of the Hijra calendar to her or his religious ideology. The conclusion is that the author the anonymous text is a Muslim who is Shi'ite by sect ideology.

As the person's writing is influenced by social and regional variables which result in the acquisition of specific vocabulary traits. Forensic linguists can utilize these characteristics to

identify and differentiate a language user's groups, region or country. For example, consider the following anonymous excerpt:

2. Would u like to **hang out with me in Cairo**

The choice of verb '**hang out**' in the excerpt (2) above has amply indicated that the anonymous comment was written by the Thread user and he/she is based in Cairo from Egypt due to this conclusion '**hang out with me in Cairo'.** This indicates that the comment even though was written in anonymous form but it encodes not only writer's age bracket but it encompasses his region which make the research to make inferencing about the nationality of the writer.

Similarly, multi-modal method of analysis can equally assist in identifying author of the anonymous comment especially when the social media user uses emoji that carries identity as in the excerpt (3):

3. Replies through putting emojis of a flag of Lebanon and a broken heart

Such type of anonymous comment clearly indicates the nationality of the social media which is Lebanon and at same time using broken heart emoji signifies that the user is the woman. Since recent research on using emoji on social media platforms attributes frequent use of heart related emoji to women.

As indicated in the example (1), Speakers' or writers' faith and ideology are always reflected in both intentional and implicational meaning that are valuable contextual footprints to unveil who is the language user as in (4):

4. Wwooowwww **alhamdulillah Subhanallah Allahu Akbar Laa illaaha illallaah Muhammad ur Rusoolullah .**

From the above, it is obvious from the propositional content of the reply that the writer here is a Muslim woman by the use of Arabic word by praising Allah and declaration of monotheism is done solely by Muslims and knowing the gender of the commenter is related to the profile photo that she put wearing a veil.

Similarly, language can reveal more about the writers of anonymous text including their feelings and belongingness; especially from mentioning their ethnic nationality or compatriots from where they live or come from and affections, they have for them as you can see in (5) below:

5. My Dear Palestine, some believe we are less than human, without dignity or rights. Yet we endure. Even our language comes from strength hidden far beneath our flesh, requiring all our breath. Sometimes our anger turns to hate. And the world shakes its head, believing we had no choice, or worse, that violence is part of our nature. Yet**, My Beloved**, see how we fill our great cavern of sorrow with even greater love.. love of land, family, tradition? We are worthy of justice, freedom, and yes, peace.

In this excerpt, the social media user utilizes lexical items such as **dear** and **beloved** that carry love, affections and intimacy for his or ethnic nationality or compatriots who are Palestine. This portends that the commentator is from Palestine but no other place even though he or she obscures personal's national identity.

Apart from the individual's linguistic choices that reveal more about the writers of anonymous text, another aspect that evidentially unmask the writers of anonymous text is the

use of ideological designated names/terms that delineate the dichotomy of Us vs Them as mentioned below in (6):

6. Sorry but remove **Morocco** from the zionist team. **We**'re defenitely not standing with a genocidal entity! Only the government. **The people** aren't standing with the genocidal! ⟠MA🫱 🜂 ⟠PS.

The two boldface words are personal reference such as pronoun **we** and **the people** which are used by an anonymous writer that posted a photo containing some flags of countries that support the Zionist Israel activities and another with Palestine indicating that Morocco and their people against the Zionism and since he/she is Morocco citizen due to the usage of **we.**

In addition to Us vs Them, words in narrative that outrightly contextualize and label certain actors with grievous crime to humanity such as **genocide** are overtly revealing not only the where the language user belongs to but reflecting his or her judgement on the debatable events. For instance, the comment posted as another reply by someone indicates that he is from Israel especially by using referring Palestine resistance as genocide as you can see in (7) below:

7. Genocide is what **Iran and its proxies want for Israel and jews worldwide.** Shame on your ignorance!

However, the identity of anonymous writer, sometimes, can be revealed easily specially the place he or she is living in when you look at a word used by in his/her profile as in :

8. Zahraa-**Shaaah :**

2nd Muharram

Imam Hussain A.S's caravan reached Karbala, thousands of Yazid's soldiers were already there, having arrived long before the Imam…….

The usage of onomastic **Shaaah in** (8) indicates that the writer is from Iran since this name is Persian name that is used highly by Iranian to denote royal title position. Though the name is announced as **Zahraa** and this suggests the writer is a woman is from Iran and at the same time embedding the story of Imam Hussein (A.S) in the narrative, also indicated that she is a Shia with a huge love for the family of the Prophet.

As indicated in the literature review discussion, writing style plays an important role in the sphere of unmasking the identity of the writer. When you see some signs that can draw your attention as with the usage of spaces or commas in places that should not be used are revealing more about the writers of anonymous text as in (9) below:

9. we've seen them use kids as human shields **,** test for booby traps **,** tape prisoners **,** use children as bait to kill other children **,** use white phosphorous on civilians **,** target journalists and aid workers **,** execute their own hostages **,** block food going to babies **,** bomb hospitals and tents **,** laugh about shooting kids **,** boast about rape - not looking good.

In the above comment as reply, we can notice that it is not only the excessive use of commas or spaces in between but the identity of a person as being (Palestinian) since he witnesses all these circumstances that his country passed with during Israeli attack.

Similarly, **discourse markers** which serve as a means of communication as mentioned by (Thamer, 2021) can also give a good clue about the excessive way which defines someone feature of writing style and consequently led to uncover his/her identity as in (10) below:

10. And I read these, every morning **and** wish I could tell them that I'm so sorry the world has turned its back **and** I love them.

This reply comes as a response to a post that declares killing 90 people, including many children in their sleep. As it is shown above, the usage of (and) which serves as a means of connection has been used more than required which make it easy to anyone to collect a lot of data from the same person to know his identity whether his age, gender, education or the place that s/he refers to.

From pragmatic dimension, implicit meaning enables the forensic linguist to unveil the identity of the anonymous writer through hints used in the writing as in:

11. But **yall Palestinian** are causing unrest in **the west**, is **the west** Palestinian land too, while **your Arab nation** understand your attack to Israel, so they are watching you pack your mess.

In the comment above that is a reply to the post, it shows that the proposition of the message reflects idea of those who don't care about the impunity action towards Palestinians. The implicit meaning of the structure such as **Yall Palestinian, the west, your Arab nation (sic)** understand your attack to Israel, suggested that the comment emanates from someone who you are not residing within the violent area; but he or she shows the direction of his or her opinion. Similarly, the writer asserts this or her point that he is from the West but not from any of the Middle east countries or Arab countries by mentioning **(your Arab nation)** to indicate that he is not part in.

As the authorship attribution examines language to identify the author of a text, the authorship profiling increases revealing detailed information about an assumed author's background through some common linguistic features. These features when identified in the text could expose author's demographic information in a constant manner such as '**orthographic style**'; such indicator will ease the identification of anonymous writer in the attribution analysis since linguistic features and demographic information remain constant within a group of people as you can see in (12):

12. Ibrahim521

   Valekum salaam mashallah u r looking….

This reply though has been written by a clear name (Ibrahim) but indicates that this person is from Turkey since he **orthographic**ally wrote **Valekum** instead of **Walekum** because in Turkish people **orthography** the used **'v'** for **'w'** for labio-velar glide sound in their writing system and this is part of linguistic feature that is consistent with Turkish people.

Furthermore, the choice of lexical word that carries the content of meaning helps for the identification for anonymous writer who replies to the posted "Five Egyptian planes loaded with rescue equipment and aid for the victims of the hurricane in America! The same regime and ruling clique apologize for their inability to bring a "cup of water" to Gaza." The writer replies with expressions of certainty as in:

13. The only entry of rescue aid and food are from **only Egypt EG do not pretend** that you **do not know** because **you already know.**

Mentioning "**the only entry'** means that he is certain about what he is saying, **do not pretend** as if he is telling the addressee you are liar, I know more than you to assure that he is the citizen of the Egypt so he knows every detail.

Traditionally, the number of paragraphs and average paragraph length are the main concern of word structures which are important features at that time in identifying a piece of writing but with appearance of online anonymous messages and their correlations as images, emojis, fonts they can provide insight about the style marker of the writer and where the message is from. It is better to refer to these current features as technical structures as in:

> 🇮🇱: I want peace with you
> 🇵🇸: I want peace without you
> 🇮🇱: not even negotiation?
> 🇵🇸: never
> 🇮🇱: you aren't really in a position to–
> 🇵🇸: no peace, a fight to the death
> 🇮🇱: fine, have it your way
> 🇵🇸: nooooooo! ceasefire now!
> 🇮🇱: Ok, but you have to ceasefire too.
> 🇵🇸: no

14.   🇮🇱: ...

Here, the message is started with Israel flag which means that the conversation begins by Israeli person and the response came from a Palestinian one which is indicated by Palestinian flag and the conversation goes on between them. Not only using of flags (images/ emojis) is helpful here but through the type of linguistic choices that both elected is the main guide of identification here.

Again the choice of the lexical item plays a significant role in masking the identity of anonymous writer as in:

15. **Allahu raheeem allahu Kareem**

These words are written as a reply for an image posted concerning the starvation in Gaza. They are usually used by Arab Muslims to console someone's bad manners. And if you here the way this consolation said, it can reveal easily the place of the writer.

7. Conclusion

From the authorship analysis, the discussions in the above data analysis has crucially and explicitly established set of linguistic features and techniques that provide not only detailed profile information about anonymous text writers but the linguistic strategy that they utilized to achieve their communication intention especially by expression their opinion without hindrance. The analysis explains how the anonymous authors appeared to be making a concerted effort to state and justify their position by deploying both the linguistic, stylistic, pragmatic and writing techniques from linguistic resources that are available within their geographical proximity.

We can pursue several potential future directions. The current authorship identification methodologies uncover various forms of lexical choice and semiotic linguistic choice and author ideological interests that are significant factors for not only identifying but differentiating between a larger set of anonymous authors in social media narratives which would be an important future endeavor.

References

Ainsworth J. & Juola P. (2019). Who wrote this?: Modern forensic authorship analysis as a model for valid forensic science, *Washington University Law Review*, *96*(5), 1161-1189.

Bevendorff, J. et al. (2022). Overview of PAN 2022: Authorship verification, profiling irony and stereotype spreaders, style change detection, and trigger detection: Extended Abstract. *In Advances in Information Retrieval: 44th European Conference on IR Research, ECIR 2022, Stavanger, Norway*, April 10–14, 2022, Proceedings, Part II (pp. 331–338). Berlin: Springer International Publishing.

Chaski, C. E. (2005). Who's at the keyboard? Authorship attribution in digital evidence investigations. *International Journal of Digital Evidence*, *4*(1), 1-13.

C. Zhong, H.-w. Chang, D. Karamshuk, D. Lee, and N. Sastry, (2024) "Wearing many (social) hats: How different are your different social network personae?" in *AAAI ICWSM*.

Coulthard, M. (1994). On the use of corpora in the analysis of forensic texts. *International Journal of Speech Language and the Law*, *1*(1), 27–43.

Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, *25*(4), 431–47.

Coulthard, M., Johnson, A., & Wright, D. (2017). *An introduction to forensic linguistics: Language in evidence* (2nd ed.). Routledge. https://doi.org/10.4324/9781315630311

De Vel, O. et al., (2001). "Mining e-mail content for author identification forensics," *SIGMOD Record*, vol. *30* (4), pp. 55–64.

Grant, T. (2008). Approaching questions in forensic authorship analysis. In J. Gibbons & M. T. Turell (Eds.), *Dimensions of forensic linguistics* (pp. 215–229). John Benjamins Publishing Company. https://doi.org/10.1075/aals.5.15gra

(2013). Method, consistency and distinctiveness in the analysis of SMS text messages. *Journal of Law and Policy*, *21*(2), 467–94.

McEnery, T. & Wilson, A. (2001). *Corpus linguistics: An introduction. 2nd Edition.*: Edinburgh University Press.

McMenamin, G. R. (2002). *Forensic linguistics: Advances in forensic stylistics*. Boca Raton and New York: CRC Press.

(2010). Forensic stylistics. In *The Routledge handbook of forensic linguistics* (pp. 539–557). Routledge. https://doi.org/10.4324/9780429030581-42

Misini, A. et al (2022). A Survey on authorship analysis tasks and techniques. *Seeu Review* 17 (2):153-167.

Robertson, J. (2013). Understanding how forensic science may contribute to miscarriages of justice. *Australian Journal of Forensic Sciences*, *45*(2), 109–112. https://doi.org/10.1080/00450618.2013.790477

Thamer, A. A. (2021). Discourse markers in selected English editorials: A communicative aspect between text producers and receivers. *International Journal on Humanities and Social Sciences,* (25), 215- 226. https://doi.org/10.33193/IJoHSS.25.2021.318

Yule, G. U. (1944). *The statistical study of literary vocabulary*. Cambridge: Cambridge University Press.

Zhao, Y., & Zobel, J. (2005). Effective and Scalable Authorship Attribution Using Function Words. *Asia Information Retrieval Symposium*.

Websites of data:

1. https://www.threads.net/@binibini_nia/post/C9A0fo0NRe8?xmt=AQGzJLua-LN7PcImZXsZSUuUmjDzoUi34yhsNfWHDn2JnQ
2. https://www.threads.net/@ethancontent/post/DA1RW10o1KI?xmt=AQGzTdb5NOtsMVUel8kDJGsgeROUluPFs0FeA6BUeUU0DA
3. https://www.threads.net/@eye.on.palestine/post/DA3X2ZDsHB_?xmt=AQGzI_YLLS3j2ekp7E1KIUyiQo_-vHYDglEL3Q287hK0sA
4. https://www.threads.net/@dubai_foreigner_community/post/C_MJYO2SnsT?xmt=AQGzSpQJ_2FxB1GHv9ImF0zrdV7s5PnYMSJJOGHFjkHHrw
5. https://www.threads.net/@bridgesforbethlehem/post/C2SZTEgORTY?xmt=AQGzVO7ms4eMuO4Hnmj32qKOpqCqKws68qAglwsYYP1NLQ
6. https://www.threads.net/@akhy_chad_24/post/DA_G37jAYS7?xmt=AQGzM0BeaEfacA9l7ckb6JCzoO0UVX-7miHOhLWyY4HBXQ
7. https://www.threads.net/@jociek1/post/DBBYafezBZv?xmt=AQGzyrI_IB4H4NnzQg_tNYVR4Fp1F3_Dc2_K3TLRGLbmXw
8. https://www.threads.net/@zahra_shahhhh/post/C9KvYIvIWXV?xmt=AQGzMc_7UyQ5i81BGydImazDIfvNH0ddGkjAQac4k0LE0A
9. https://www.threads.net/@thorinseex/post/DCoCTquIzBX?xmt=AQGzS2uhv_qPYuwBhMldE5KRVms4fCejtH-0mJn1qbQkOQ
10. https://www.threads.net/@cynthjoseph/post/DCoiiQ1RdMm?xmt=AQGzU3OwywBhkAx0DpTbsGJwQKnXfSRm2WEqWLOyokEbvw
11. https://www.threads.net/@nini_ajiofficial/post/DCn7UClu9FU?xmt=AQGzxhKsjwBEC0HuAt77VvzbdJwLbHE4w7lp6gKAWLTdrQ
12. https://www.threads.net/@_yasmeen_83/post/DCowv5KB1V3?xmt=AQGzpMqENUJk4aKCt17I91TWFhibkOV6pe1KyDQU7_RGSw
13. https://www.threads.net/@sameh_egypt/post/DBB7mGBtgIh?xmt=AQGzduzcL9BE7Padj4ekAlvBrkxVID-9OBbsbf5aedvbZA
14. https://www.threads.net/@poemcpoe/post/DCqTNOcg2Zy?xmt=AQGzF_rVMvrWstg6JNDxF0K3gTPufrBG_izhXG3QOnvASA
15. https://www.threads.net/@only_adilsayyad_786/post/DCqRv06K1mM?xmt=AQGzgIxAp3-W3ZAZETI-m-F1Zs_0YG0ZSCqJbVmLNVBEXQ