INTERNATIONAL JOURNAL OF ACADEMIC RESEARCH IN BUSINESS & SOCIAL SCIENCES

Published Online: 19 January 2023 Vol 13, Issue 1, (2023) E-ISSN: 2222-6990

A Linguistic Analysis of Authorship Attribution in E-Commerce Scams' Promotional Contents and Narratives

Nursyaidatul Kamar Md Shah[1], Ameiruel Azwan Ab Aziz[2] and Aminabibi Saidalvi[3]

[1,2]Academy of Language Studies, Universiti Teknologi MARA, Cawangan Melaka, Kampus Alor Gajah, Melaka, MALAYSIA, [3]Academy of Language Studies, Universiti Teknologi MARA, Cawangan Johor, Kampus Pasir Gudang, Johor, MALAYSIA.

Email: nursyaidatul@uitm.edu.my, ameirul@uitm.edu.my, aminabibi@uitm.edu.my

Abstract

The emergence of e-commerce or online shopping platform has opened-up multiple opportunities for business owners to multiply their earnings. At the same time, this enables consumers to shop conveniently compared to the traditional shopping method, especially during the Covid19 pandemic. In Malaysia, the development opens consumers to the novel threat of fraud and scams, especially on social media and other online shopping platforms such as Shopee, Lazada and Facebook Marketplace. Due to the alarming rise in scams worldwide and the elusive methods offenders use, law enforcement agencies have turned to public education and awareness programs to reduce the number of scam victims. Thus, this study aims to identify authors' attributes in e-commerce scams' promotional contents and narratives to examine the scammers' persuasive strategies to deceive their victims. Specifically, the study intends to look at the linguistic features in authorship attribution, including lexical, syntactic, semantic, structural, and content-specific features of the corpus. Although linguistic analysis has never been used in cybersecurity, learning the language used by scammers could help researchers collect more comprehensive empirical data and educate the public about this common phenomenon.

Keywords: Linguistic Analysis, Authorship Attribution, e-Commerce Scams

Introduction

Linguistics is fundamentally concerned with the nature of language and communication. As a communication tool, language can be utilised in various ways to achieve the basic human need to connect with each other (Akmajian et al., 2017). The internet brings a plethora of benefits and eases to users and online consumers. The emergence of e-commerce or online shopping platform has opened-up multiple opportunities for business owners to multiply their earnings. It is also an opportunity for consumers to shop conveniently compared to traditional shopping. Some famous e-commerce platforms in Malaysia are Shopee, Lazada and Facebook Marketplace (Vasudevan & Arokiasamy, 2021). The opportunities brought upon by the advancement of technologies open possibilities where some people can attempt

malicious acts just by using language to gain personal monetary benefit. People with ill intent or scammers attempt to manipulate potential victims through manipulation, which can be done through language (Iswara & Bisena, 2020).

Table 1.0
The Statistics of e-Commerce Scams in Malaysia from 2018 – May 2022

Year	Cases Reported	Loss (RM)
2022 (January – May)	3833	21.7 million
2021	9569	57.73 million
2020	8851	41 million
2019	3512	28 million
2018	3318	22.39 million
Total	**29083**	**170.82 million**

Source: Bernama (2022) & New Straits Times (2022)

Table 1 illustrates the breakdown of e-commerce cases and the financial loss in the past five years. The number of reported cases and financial losses increased dramatically from 2020 to the current date. The Covid19 outbreak in early 2020 contributed to this disturbing phenomenon as people were in lockdown and spending more time on the internet (Seah et al., 2022). Although the authorities such as Royal Malaysia Police (RMP), Bank Negara Malaysia (CBM) and Malaysian Communication and Multimedia Commission (MCMC) have come up with numerous awareness campaigns against such scams, many still become a victim of e-commerce scams, and this presents a challenge to the authority (David, 2022). Many studies have shown that the best defence from becoming a scam victim is a comprehensive public education to raise awareness from the public themselves (The Sun Daily, 2022; Kadoya et al., 2020; Rahman et al., 2020; Vayansky & Kumar, 2018).

Problem Statement

The main concern that led to this study was the alarming rate of e-commerce scams cases with astronomical financial losses, as reported by several studies around the world (Paintal, 2021; Hanoch & Wood, 2021; Sinha et al., 2020; Zahari et al., 2019). In Malaysia, statistics showed that approximately 5.2 billion (RM) of financial losses were recorded from 2018 – May 2022 due to e-commerce scam-related cases (Shah & Chudasama, 2021). Much research is looking at the technical side of scamming, such as cyber security software and programming (Richardson, 2020; McGowan, 2021). Nonetheless, it is crucial to look at the non-technical side of scamming: the language scammers use to scam their victims. The social engineering part of a scam involves language cues and language features to manipulate someone into believing something they have no interest in, and it can be learned (Pouryousefi & Frooman, 2019; Swain et al., 2017). This study aims to identify authors' attributes in e-commerce scams' promotional contents and narratives. The research questions guiding this research are 1) How do scammers construct their authorship attributions in e-commerce scams' promotional content and narratives? 2) How does authorship attribution concatenate stylistic deception in e-commerce scams' promotional content and narratives? and 3) What are the patterns of communication applied by scammers in e-commerce scams' promotional contents and narratives?

Methodology

Research Design

Authorship analysis examines the characteristics of a piece of work to conclude its authorship (El Bouanani & Kassou, 2014). Authorship attribution is mainly concerned with identifying the real author of a disputed anonymous document (Tabron, 2016). The primary notion underlying statistically or computationally supported authorship attribution is that we may discriminate between texts produced by different authors by assessing various textual features. Previous studies on authorship attribution have proposed taxonomies of features to quantify the writing style (style markers) under different labels and criteria, also known as stylometry. Stylometry is computational linguistics that studies the quantitative assessment of linguistic features in natural language texts. It is closely related to the terms of the author's style and idiolect that imply a system of language features used by the author.

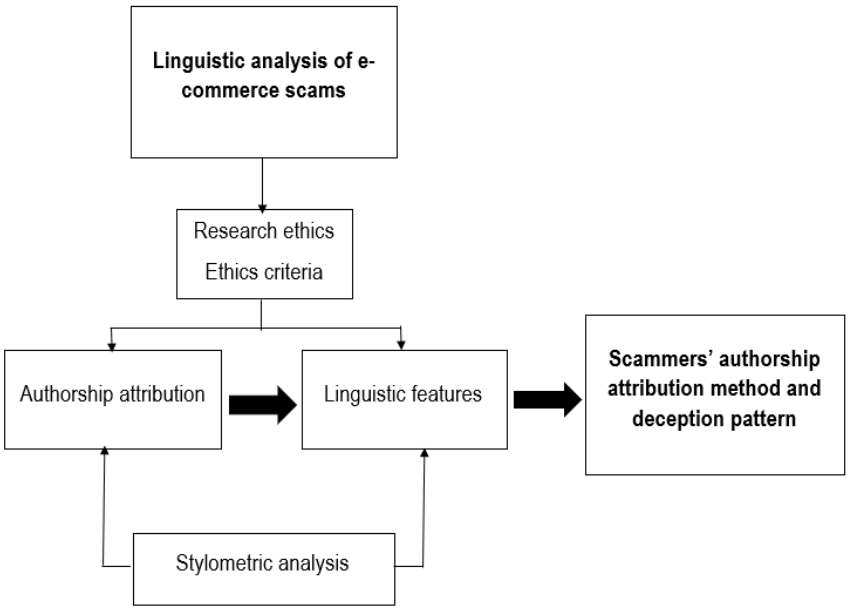

Figure 1. Conceptual framework of the study

This study aims to provide an understanding of and recognise scamming works through authorship linguistic analysis. A conceptual framework will be used to guide this study, as shown in Figure 1. The starting point of the discussion is the language scammers use in their promotional content and narratives. First, the proposed study will apply for approval from the research ethics committee. This is to ensure that this study meets recognised ethical standards, which include respect for the dignity, rights, safety, and well-being of participating subjects. Once approved, the study will move to the next phase, analysing authorship in e-commerce scams based on text analysis.

In this phase, the analysis will deal with authorship to infer the author's characteristics from the documents produced by that author since the scammers use manipulative words to deceive their victims. Authorship attribution is the process by which a linguist attempts to identify the author of an anonymous or unknown document based on linguistic clues left by the author. The dataset from websites, social media, and text messages is analysed using Linguistic Inquiry and Word Count (LIWC) text analysis software based on lexical features (number of words, word n-grammes, average word length, frequency of same words) and character-based features (total number of characters, ratio of capital letters, frequency of special characters), syntactic features

(POS tags, frequency of punctuation marks, ratio of singular to plural nouns, proper nouns, pronouns), content-specific features (topic-specific features, sentiment words), and unique words used in the text. In addition, errors and idiosyncrasies of the data are also evaluated during the analysis. Detecting a text's authorship by analysing an author's writing style is called stylometry. Finally, a number of patterns can be derived from the result that can be identified as the method of authorship attribution and deception patterns of the scammers.

Sampling

This study will use purposive and snowball sampling techniques to seek potential documents and samples based on the inclusion and exclusion criteria set based on the study's objectives. The documents will consist of scammers' promotional websites, social media, text messages, and customers' testimonies and reviews. This study aims to analyse as many documents as possible, subject to relevant authorities' sensitivity, availability, and consent. It is estimated that the number of documents to be collected and analysed should be within the range of 40 to 100. The dataset in the proposed study will be sourced from Malaysian-based illegal online investment, and e-commerce scams' promotional materials and narratives identified and listed by the relevant authorities (MCMC, RMP, CBM) will be selected. Priority will be given to data written in English; however, the Malay language will also be considered. Atlas.ti and LIWC software will assist in data management and analysis.

Significance of Proposed Study

Therefore, the linguistic analysis of authorship attribution in e-commerce scam's promotional contents and narratives may reveal how the scammers construct their authorship attribution and how the linguistic features concatenate stylistic deception in e-commerce scams' promotional contents and narratives. It is believed that the linguistic strategies used are a major contributing factor in influencing prospective victims' behaviour in their decision. This study will first explore the authorship attribution in e-commerce scams' promotional contents and narratives using stylometric analysis with LIWC software. Next, it will investigate the linguistic features in e-commerce scams' promotional contents and narratives through content analysis with Atlas.ti software. Finally, interviews with the respective parties will be conducted to triangulate and unearth the explanation to provide a clearer understanding and a conclusive narration of this phenomenon.

Implications for Research

The linguistic analysis uncovers several features of language interaction in a limited data set and could potentially assist cybersecurity defence. It could also be used to identify the linguistics modus operandi of the crime and significantly reduce the risks of being deceived. All subfields of linguistics can be utilised in the linguistic analysis to describe how language works from the smallest unit of sound up through words and phrases to sentences and help to understand how language functions in discourses across social units such as divisions of gender, ethnicity, communities of practice and other groups (Tabron, 2016). While linguistic analysis has not previously been applied in cybersecurity, gaining a scholarly understanding of the language of the scammers could provide more comprehensive empirical data and public education against this prevalent phenomenon.

Literature References

The rapid growth of Internet technologies has brought numerous changes in various spheres of life. Since the advent of the digital age, people from different socioeconomic and geographic backgrounds have become increasingly connected as the internet opens more accessible ways to stay in touch, meet new people, and network. In addition, the internet's extensive influence on society also means that people have become highly dependent on it for personal and business purposes. In particular, the internet enables various forms of e-commerce to be conducted with relative ease.

However, this ease also contributes to a concerning increase in cybercrime. According to Singh et al. (2021), there were nine different types of cybercrime attacks in Malaysia between 2008 and 2020, including "Cyber Harassment," "Fraud and Forgery," "Malicious Code," "Denial of Service (DOS)," "Intrusion," "Content-Related," "Intrusion Attempt," "Spam," and "Vulnerabilities Report." Fraud and forgery, in particular, are on the rise at an alarming rate and include fraudulent transactions through e-commerce sites, unauthorised transactions, illicit investments, and Nigerian online scams. The internet has opened the floodgates to cyberscams, defined as any fraud that uses mass communication technology to defraud people of their money, even if Whitty (2020) observed that scams or frauds have occurred for a very long time before the invention of the internet.

The majority of earlier research has tended to concentrate on various forms of cybercrime over the years. A case in point is the online dating romance fraud that took place in Malaysia (Shaari et al., 2019; Kamaruddin et al., 2020) which exposed a number of common techniques and deceptive linguistic tactics that scammers employ to manipulate victims. These techniques include positive and negative politeness techniques, such as the assertion of commonality and the hints of affiliation, interest, and similarity. In a different study of Chinese cybercriminals, victims were duped into providing their bank account information and making financial transfers into the accounts of the fraudsters using the instant messaging service "QQ" (Hua et al., 2017). This study has demonstrated a distinct discourse pattern that may be recognised in how criminals of money use formulaic language. According to Tan et al. (2020), online con artists continued to target the vulnerable throughout the Covid 19 pandemic by seducing victims into shady business deals and dubious investment schemes. Talib and Rusly (2015) explained that victims are cheated out of their money when a seller makes them promises for products or services that either do not exist, are not meant to be supplied, or are misrepresented. The Department of Statistics Malaysia (DSM) crime statistics show that cybercrime complaints skyrocketed by 99.5% from 10,426 in 2019 to 20,805 in 2020, illustrating the disturbing trend in cyber-related scams. This discovery allows us to foresee and estimate the expected growth of the same incidents during the following decades.

The ability to utilise language for good or bad makes it a vital instrument. Speaking a language gives speakers a variety of instruments to fulfil their manipulative objectives. They can use language to manipulate others in subtle, frequently unfair, or self-serving ways. In most scams, words and phrases that appeared to be warnings, flattery, or promises were used to influence, deceive, and persuade victims (Media Prima Television Network, 2021; Shaari et al., 2019). Scam victims are frequently persuaded by the scammers' convincing promises, making them susceptible to half-truths, manipulation, and even blatant lies (Hartwig & Voss, 2017). Although the previous research has offered some insight into trends in cybercrime, there is a lack of studies that concentrate on cybercrime in Malaysia, particularly in terms of the language tactics that scammers employ. This lack of scientific proof negatively impacts

the targeted efforts of the appropriate authorities to implement the necessary actions. Therefore, it is crucial to analyse the language used in scam situations to inform the public, stop them from falling for the scam, and reduce financial losses.

Conclusion

While linguistic analysis has not previously been applied in cybersecurity, gaining a scholarly understanding of the language of the scammers could provide more comprehensive empirical data and public education against this prevalent phenomenon. Therefore, this study aims to explore the authorship attribution of scammers and linguistic clues in the advertising materials and narratives in e-commerce scams. This study presents a detailed investigation of sub-fields of pure linguistics, including morphology, syntax, and semantics. The study may reveal the validity of the scammers' technique for better financial scam detection and public awareness. Additionally, it is believed that the findings of this study would offer pertinent empirical linguistic evidence for improved public education, increased cyber security awareness, and strengthened cyber security in line with the fourth strategic pillar of the Malaysia Cyber Security Strategy 2020–2024. Hopefully, this study will contribute to comprehensive public education regarding the scammers' linguistic techniques as an effective defence against them to avoid more monetary losses and declining quality of life.

Acknowledgement

The authors would like to acknowledge the support from the Ministry of Higher Education (MoHE) Malaysia through a research grant award (Project code: FRGS/1/2022/SSI09/UiTM/02/10). Our gratitude is also extended to all parties directly or indirectly involved in completing the research project.

References

Akmajian, A., Farmer, A. K., Bickmore, L., Demers, R. A., & Harnish, R. M. (2017). *Linguistics: An introduction to language and communication*. MIT Press.

David, A. (2022). *RM5.2b in losses through online scams since 2020*. New Straits Times. Retrieved September 10, 2022, from https://www.nst.com.my/news/crime-courts/2022/08/819331/rm52b-losses-through-online-scams-2020

Hanoch, Y., & Wood, S. (2021). The scams among us: Who falls prey and why. *Current Directions in Psychological Science*, 30(3), 260-266.

Hartwig, M., & Voss, J. A. (2017). *Lie Detection Guide: Theory and Practice for Investment Professionals*. Virginia: CFA Institute.

Hua, T. K., Abdollahi-Guilani, M., & Zi, C. C. (2017). Linguistic Deception of Chinese Cyber Fraudsters. 3L: The Southeast Asian Journal of English Language Studies, 108 – 122.

Iswara, A. A., & Bisena, K. A. (2020). Manipulation And Persuasion Through Language Features In Fake News. *RETORIKA: Jurnal Ilmu Bahasa*, *6*(1), 26-32.

Kadoya, Y., Khan, M. S. R., & Yamane, T. (2020). The rising phenomenon of financial scams: evidence from Japan. *Journal of Financial Crime*. Vol. 27, No. 2, pp. 387-396

Kamaruddin, S., Wan Rosli, W. R., Abd Rani, A. R., Md Zaki, N. Z. A., & Omar, M. F. (2020). When love is jeopardised: Governing online love scams in Malaysia. *International Journal of Advanced Science and Technology*, 29(6), 391-397.

McGowan, E. (2021). *How to identify the language tech support scammers use to scam*. Avast Blog. Retrieved September 10, 2022, from https://blog.avast.com/tech-support-scammer-language-avast

Media Prima Television Network [TV3MALAYSIA Official]. (2021, November 11). *Laporan Kes Jenayah Komersial (Jenayah Dalam Talian)* | MHI (11 November 2021) [Video]. YouTube. https://www.youtube.com/watch?v=Yjagq1WnTEM

Paintal, S. (2021). E-commerce and Online Security. *International Journal of Management (IJM)*, 12(1).

Pouryousefi, S., & Frooman, J. (2019). The consumer scam: an agency-theoretic approach. *Journal of Business Ethics*, 154(1), 1-12.

Rahman, A. A., Azmi, R., & Yusof, R. M. (2020). Get-Rich-Quick scheme: Malaysian current legal development. *Journal of Financial Crime*.

Richardson, J. (2020). Is there a silver bullet to stop cybercrime? *Computer Fraud & Security*, 2020(5), 6-8.

Seah, C. S., Loh, Y. X., Wong, Y. S., Jalaludin, F. W., & Loh, L. H. (2022, April). The Influence of COVID-19 Pandemic on Malaysian E-Commerce Landscape: The case of Shopee and Lazada. *In Proceedings of the 6th International Conference on E-Commerce, E-Business and E-Government* (pp. 215-221).

Shaari, A. H., Kamaluddin, M. R., Paizi, W. F., & Mohd, M. (2019). Online dating romance scam in Malaysia: An analysis of online conversations between scammers and victims. *GEMA Online® Journal of Language Studies*, 19(1), 97-115.

Shah, A., & Chudasama, D. (2021). Investigating Various Approaches and Ways to Detect Cybercrime. *Journal of Network Security*, 9(2), 12-20.

Singh, M. M., Frank, R., & Zainon, W. M. N. W. (2021). Cyber-criminology defense in pervasive environment: A study of cybercrimes in Malaysia. *Bulletin of Electrical Engineering and Informatics*, *10*(3), 1658-1668.

Sinha, P., Sharma, U., Kumar, D., & Rana, A. (2020, June). A conceptual framework for mitigating the risk in e-commerce websites. *In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 217-221). IEEE.

Swain, S., Mishra, G., & Sindhu, C. (2017, April). Recent approaches on authorship attribution techniques—An overview. In *2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)* (Vol. 1, pp. 557-566). IEEE.

Tabron, J. L. (2016). Linguistic features of phone scams: A qualitative survey. *11th Annual Symposium on Information Assurance (ASIA'16)*, 52-58.

Talib, Y, Y. & Rusly. F, H. (2015). Falling Prey for Social Media Shopping Frauds: The Victims' Perspective. *Proceeding of the International Conference of E-Commerce (ICoEC)*, Kuching, Sarawak, 2015.

Tan, S. L., Vergara, R. G., Khan, N., & Khan, S. (2020). Cybersecurity and privacy impact on older persons amid covid-19: A socio-legal study in Malaysia. *Asian Journal of Research in Education and Social Sciences*, *2*(2), 72-76.

The Sun Daily. (2022). *Empowering the public against online harm.* https://www.thesundaily.my/spotlight/empowering- the-public-against-online-harm-FM9592887

Vasudevan, P., & Arokiasamy, L. (2021). Online Shopping Among Young Generation in Malaysia. *Electronic Journal of Business and Management*, *6*, 31–38.

Vayansky, I., & Kumar, S. (2018). Phishing–challenges and solutions. *Computer Fraud & Security,* 2018(1), 15-20.

Whitty, M. T. (2020). Is there a scam for everyone? Psychologically profiling cyber scam victims. *European Journal on Criminal Policy and Research*, 26(3), 399-409.

Zahari, A. I., Bilu, R., & Said, J. (2019). The Role of Familiarity, Trust and Awareness Towards Online Fraud. *Journal of Research and Opinion*, 6(9), 2470-2480.