

Modelling the Language of Cyber Scammers: A Forensic Linguistics Analysis of Malaysian Illegal Investment Scheme

Ameiruel Azwan Ab Aziz¹, Sheik Badrul Hisham Jamil Azhar², Ariff Imran Anuar Yatim³, Ahmad Azfar Abdul Hamid⁴, Nor Afifa Nordin⁵, Nur Syafiqah Zabidin⁶ and Adelina Asmawi⁷

¹²³⁴⁵⁶Academy of Language Studies, Universiti Teknologi MARA Melaka, ⁷Faculty of Education, Universiti Malaya

Email: sheik835@uitm.edu.my, ariffimran@uitm.edu.my, ahmadazfar@uitm.edu.my, afifanordin@uitm.edu.my, nursyafiqah@uitm.edu.my, adelina@um.edu.my

To Link this Article: <http://dx.doi.org/10.6007/IJARBSS/v11-i6/10388> DOI:10.6007/IJARBSS/v11-i6/10388

Published Date: 28 June 2021

Abstract

Over the last few years, there are growing concerns over the astronomical increase of reported cases of illegal internet investment schemes in Malaysia. Although the Central Bank of Malaysia (CBM) regulates and supervises financial institutions in order to combat illegal investment scheme through Financial Services Act 2013 (FSA 2013) and Islamic Financial Services Act 2013 (IFSA 2013), they do not protect the consumers should they choose to deal with the illegal financial service providers. The large financial losses have a negative impact fiscally on investors' socioeconomic. Hence, this study is proposed to analyse Malaysian illegal internet investment scammers' language properties to provide cyber defense for the public. Using Zhang and Wildermuth's (2009) content analysis process, this study intends to examine the linguistic strategies used by scammers in their websites to deceive their targets. This study will focus on several websites identified by CBM as the primary data source. Semi-structured interviews will be conducted with personnel from relevant government agencies. For triangulation purposes, individuals who have experienced investing and willing to share their experiences and data will be interviewed. All data will be coded and categorised, and emergent themes will be presented to answer research questions. With great hope, this study's results will help provide comprehensive public education about these scams' linguistics strategies as a strong defense against them. The results may also help the authorities develop better crime prevention methods and contribute to the criminal justice system in presenting more significant linguistic evidence in the court of law.

Keywords: Forensic Linguistic Analysis, Illegal Investment

Introduction

The advent of the digital age brings about numerous changes in all aspects of life. With the widespread use of internet technologies, people from different socioeconomic and geographical divides have seen a rapid shift in the ability to share and receive information freely. Inevitably, people are becoming heavily reliant on internet technologies for personal and business purposes. The latter involves various forms of electronic commerce such as online bill payments, internet transactions, and stocks trading over the internet. However, this convenience also leads to a worrying rise in cybercrimes, including illegal internet investment schemes.

In this day and age, the phenomenon of blind investors being deceived by illegal internet investment schemes has become a trend due to advancements in technology. This illegal investment over the internet phenomenon negatively impacts the national financial order and social credit system by pushing thousands, if not millions, into financial crisis and poverty. The Royal Malaysian Police (RMP) remarked that over RM40 million were lost in the first ten months in 2019, with 9686 cases reported related to this cybercrime (Malaymail, 2019). Surprisingly, the RMP revealed 467 identified illegal internet investment scams with an upsurge in losses amounting to RM52.4 million in the first seven months of 2020 (Malaymail, 2020). It is undeniably a great challenge for the authorities to combat this cybercrime because the network of illegal fund-raising is often done quickly in a brief period. This phenomenon has affected regular financial order and society's well-being, specifically in the social credit system, as it creates financial instability (Jianying, 2018). Furthermore, these schemes' victims have also experienced two-fold impacts; the loss of money and psychological trauma. Many investors have been deceived with a cunning marketing strategy that uses intriguing phrases and promises such as "high profit," "high interest," and "high return". As part of their persuasive strategy, these phrases often mislead investors to fall into these "Get-Rich-Quick" schemes. Besides offering high and unrealistic return rates for a small investment, the schemes also promise that the investment is easy and risk-free.

It is challenging for the public and government authorities to identify these internet investment schemes' legitimacy due to the substantial confusion in its linguistics presentation. The website content's narrative patterns involving vast sums of profit returns and common persuasive strategies that appeal to greed, altruism, trust, and flattery often make the schemes irresistible. Many linguistics studies have documented and discussed the narratives and persuasive strategies in other forms of cyber scams such as love or romance scam (Shaari, Kamaluddin, Paizi & Mohd, 2019), Macau scam, and "Nigerian" e-mail scam (Schaffer, 2012). However, attention on illegal internet investment schemes, especially in the Malaysian context, lacks the literature.

There is a substantial need for studies to analyse and reveal the cyber scammers' language to provide comprehensive public education and defense against this cybercrime. The proposed study intends to utilise forensic linguistic analysis (Tabron, 2016) to achieve its objectives. Although forensic linguistics research is a relatively new discipline in Malaysia, the urgency to carry out the proposed study would be a starting point to reduce the risks of people being deceived in illegal investments. The proposed study intends to achieve the following objectives:

1. To examine the linguistics styles and patterns used in the narrative presentation on illegal investment websites
2. To explore the linguistic strategies of persuasion used to manipulate prospective investors.

3. To describe the development and presentation of linguistics modus operandi model on the illegal internet investment websites in Malaysia

Literature Review

Illegal Internet Investment Scheme

The term illegal internet investment scheme refers to illegal deposit-taking activities over the internet where websites are used as a medium for interaction, communication, and business transactions. The masterminds behind these schemes also offer unlicensed fund management and investment advice to unsuspecting victims. Schaffer (2012) notes that scams of this nature often promise lucrative rewards that can be quickly earned without further details on the associated risks. People who fall victim to illegal internet investment schemes are often swayed by statements of other people deemed 'professionals' and are subsequently vulnerable to half-truths, manipulation, and even absolute deception (Hartwig & Voss, 2017). Examples of such schemes in Malaysia include Dinar Dirham Global, Ethtrade Limited, Sanabil Investment, XIG Limited, and Wadiah Trading.

Over the years, most related studies in this area have focused more on different types of cybercrimes. One example is the online dating romance scam in Malaysia (Shaari et al., 2019), which identified several common steps and persuasive language strategies used by scammers in manipulating the victims. These strategies included both positive (claim common ground, the indication of interest, membership, and similarities) and negative (direct claims, statements, and requests) politeness strategies. In another study on Chinese cyber fraudsters, victims in China were tricked into revealing their bank accounts and depositing money into the fraudsters' bank accounts via 'QQ,' an instant messenger platform (Hua, Abdollahi-Guilani, & Zi, 2017). This study demonstrated that a specific discourse pattern on the way cheats or fraudsters follow to steal money is discernible, and the language they use is formulaic. Besides, an analysis of cyber-related fraud incidents in Malaysia based on seven-year data of reported incidents revealed that this period amounted to 27,671 incidents with an average of 3,953 incidents each year (Hashim et al., 2017). The maximum number of reported incidents from 2011 to 2016 occurred in the first six months each year. This finding can predict and anticipate the likely increase of the said incidents in the coming years. While the above studies have helped shed some light on cybercriminal activities, there have not been enough studies focusing on illegal internet investment schemes in Malaysia. The study of illegal internet investment schemes is still in its infancy in Malaysia, despite the alarming increase in reported cases. This lack of empirical findings proves detrimental to concentrated efforts by relevant parties in providing necessary interventions.

Forensic Linguistic Analysis

This study intends to use forensic linguistics to examine and identify the deceptive language used in illegal internet investment websites and identify and document the linguistics modus operandi used by the scammers to deceive prospective investors of the linguistic styles and patterns in manipulating their targets. Forensic linguistics is the study of language as evidence for the law. It uses linguistics science to examine language as evidence in civil or criminal cases (Tabron, 2016). It is a relatively new field and has not widely been applied to cybersecurity.

The linguistic analysis uncovers several features of language interaction in a limited data set that begin to answer how forensic linguistics could assist in cybersecurity defense. Forensic linguistics could be used to identify the linguistics modus operandi of the crime and, therefore, significantly reduce the risks of being deceived. All subfields of linguistics, namely

phonology, morphology, semantics, pragmatics, sociolinguistics, and psycholinguistics, can be utilized in the forensic analysis to describe how language works from the smallest unit of sound up through words and phrases to sentences and help to understand how specific speech works in interaction across social units such as divisions of gender, ethnicity, communities of practice and other groups (Tabron, 2016). Forensic linguistics is a field that is still growing and evolving, and if viewed as the application of linguistic analysis to forensic texts and contexts, its scope, likely, the need for linguistics in a forensic context will only grow. Technological advances create the potential for increasing perceived and actual anonymity and while computer forensic and other disciplines have considerable contributions to make, so do forensic linguists. Forensic linguistics analysis is significant for identifying the crimes, thus creating better awareness among the public, providing better cybersecurity defense, and presenting empirical linguistics evidence in court cases.

Conceptual Framework of Illegal Internet Investment Scheme Forensic Linguistics Analysis

This study will employ a qualitative content analysis research design as this study intends to examine the general nature, linguistic styles, patterns, and persuasive strategies used in Malaysian illegal internet investment websites. For that purpose, a conceptual framework will be used to guide this study, as shown in Figure 1:

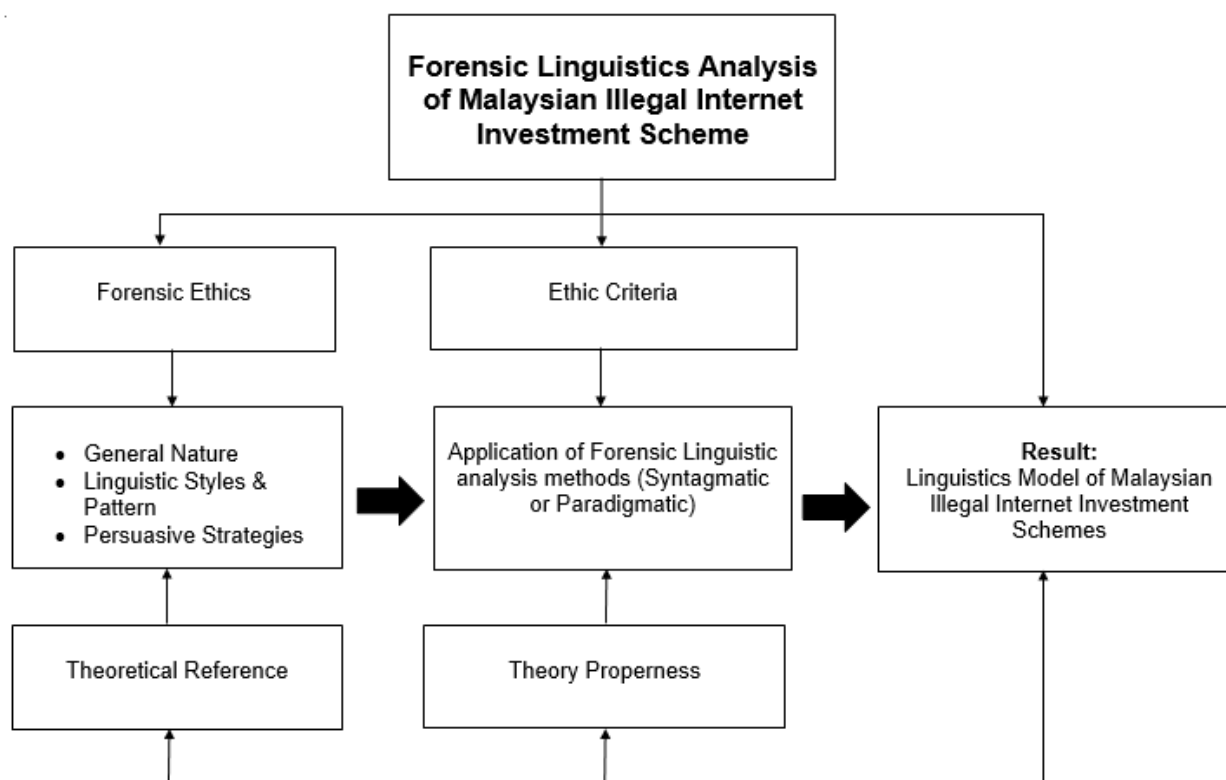


Figure 1: Conceptual framework of forensic linguistics analysis of Malaysian illegal internet investment scheme

The onset of the study is the issue ensuring the trustworthiness of the outcomes of the investigation. Ethics is an essential issue in conducting a study as it has the involvement of humans as participants. Therefore, few factors were taken into considerations to ensure the flow of the study. In the first phase, this study will focus on the forensic ethics concerns by addressing the ethic criteria to ensure a true and accurate representation of data, accurate

data analysis, maintain integrity of the evidence and results, clear and complete documentation, the impartiality of interpretation and reporting and consideration of confidentiality and disclosure. Consent to conduct the study will be obtained from the research ethics committee prior to the actual study.

The second phase of the proposed study involves two research activities. The first activity focuses on examining the language used by the scammers in Malaysian illegal internet investment schemes. Concentration is placed on examining the general language nature, linguistics styles, patterns, and persuasive strategies to recruit potential investors to invest in the schemes. The content of the websites will be analysed from various perspectives, including the sub-fields of textual linguistics elements that can be brought into play for forensic work: phonology, morphology, semantics, and pragmatics to describe how language works from the smallest unit of sound up through words and phrases to sentences. Moreover, the heuristic cues; the non-linguistic elements; will also be examined to analyse their assistance in promoting the schemes. During this process, theoretical reference will be conducted to assist interpretation and discussion of results.

The second activity in this phase will be done by re-analysing the data using forensic linguistics analysis methods. This research activity intends to explore the syntagmatic relationships that involve sequences of linguistics signs that create meaning and the paradigmatic relationships that involve linguistics signs that can substitute each other, where it is commonly used to change the meaning with the substitution. During this process, researchers will focus on theory properness in constructing a new theory or concept of the linguistics features presented in the narration on illegal investment websites.

The final phase of the proposed study is to formulate the linguistics model of Malaysian illegal internet investment schemes used by the scammers to convince prospective investors based on the data analysis. The model will be discussed as the study results that might shed light and offer new knowledge in forensic linguistic analysis.

Methodology

This study will employ a qualitative content analysis research design as this study intends to examine the general nature, linguistic styles, patterns, and persuasive strategies used in Malaysian illegal internet investment websites. Purposive and snowball sampling techniques will be utilised to seek potential documents and participants based on the inclusion and exclusion criteria set based on the study's objectives. This study aims to analyse as many websites as possible, subject to their sensitivity, availability, and consent by relevant authorities. It is estimated that the number of documents to be collected and analysed should be between 40 to 100. Potential Malaysian-based illegal internet investment websites that use either English or Malay language and are identified and listed by the relevant authorities (RMP, CBM, MCMC) will be selected. In addition, this study plans to invite at least three personnel from relevant agencies who have the experiences in investigating illegal internet investment schemes and recruit approximately 15 to 30 participants who have the experiences in investing in such schemes and willing to share their experience and data with the researchers. In respecting the privacy of participants, individual confidentiality is maintained and will be given pseudonyms. Besides, participants should give their informed consent to participate in this study. Each participant will be interviewed in a series of sessions to allow researchers to understand and verify their responses. Field notes will be taken throughout the process to describe non-verbal information.

The numbers of website and participants depend on the study's ability to reach the saturation point. The search for potential documents and participants will end when the data reached saturation level, in which there is sufficient information to replicate the study and no new data emerged. In other words, the gathered data can answer research questions, and the ability to gain additional new information has been accomplished, and further coding was no longer feasible. The inclusion and exclusion criteria are as shown in Table 1 below.

Table 1

Inclusion and exclusion criteria of Malaysian illegal internet investments schemes documents and research participants

| Sample | Inclusion criteria | Exclusion criteria |
|---------------------|--|--|
| Documents | <ul style="list-style-type: none"> • Websites identified by relevant authorities (RMP, CBM, MCMC) • Based in Malaysia • English and Malay language • Focus on investment schemes | <ul style="list-style-type: none"> • Foreign investment schemes • Not listed by authorities • Macau, Nigerian, online purchase, travel/vacation, fake check, advance-fee loan, love/romance scams |
| Participants | <ul style="list-style-type: none"> • Investigating officers/personnel • Individuals/victims with experiences investing in illegal schemes | <ul style="list-style-type: none"> • No direct experience/hearsay |

Data Analysis

Qualitative data analysis is a systematic process to search and arrange data from various sources such as document analysis and interviews to understand the retrieved information. Data collection and analysis in qualitative research should go hand-in-hand and be done simultaneously so that the researchers could focus and shape the study as it proceeds (Fraenkel & Wallen, 2006). Data will be collected, analysed, and arranged into specific patterns, categories, or themes and then will be interpreted using a particular scheme (de-contextualization and re-contextualization). Data analysis will be conducted on every data as soon as the first data collected. This process will be used as a strategy to avoid the abundance of data, which can create difficulties in the analysis process. In the context of this study, data will be primarily collected through document analysis, while data from interviews will be used to support the analysis. The use of various data collection methods enabled possible data comparison and consolidation to be made.

Data will be analysed based on Zhang and Wildemuth's (2009) content analysis process that involves seven steps, which are preparation of data, defining the units of analysis, developing categories and coding scheme, testing coding scheme on a sample of text, coding all texts, assessing coding consistency and drawing a conclusion from the coded data. For triangulation purposes, semi-structured interviews will be conducted to gain insights and discover a perspective concerning the issue. This approach will ensure the trustworthiness of the interpretations of emergent themes. Table 2 is the illustration of content analysis steps.

Table 2. Zhang and Wildemuth's (2009) content analysis steps

| Step | Process | Description |
|------|---|---|
| 1 | Preparation of data | Prepare existing texts, justify the choice of content |
| 2 | Defining the units of analysis | Unitised texts before coding |
| 3 | Developing categories and coding scheme | Develop a coding manual, which usually consists of category names, definitions or rules for assigning codes, and examples |
| 4 | Testing coding scheme on a sample of text | Coding sample text, checking coding consistency, and revising coding rules |
| 5 | Coding all texts | Coding interesting data features systematically across the entire data set, collating data relevant to each code. |
| 6 | Assessing coding consistency | Recheck the consistency of coding |
| 7 | Concluding the coded data | Making sense of the themes or categories identified, and their properties |

The first step of content analysis will be conducted to identify the general steps and strategies presented in the websites before a thorough linguistic analysis is carried out. An interim study will be conducted to identify and then rectify problems before the actual study. This is a crucial phase to provide a preliminary analysis and fulfil the content analysis's first four steps (preparation of data, defining the units of analysis, developing categories and coding scheme, and testing coding scheme on a sample of text). The websites' narratives will be linguistically analysed and triangulated with interview data. To ensure the trustworthiness of the analysis, three content experts will validate the interview protocol. The Atlas.ti software will be used to assist in data management and analysis.

Within any interpretive study, there will be concerns regarding the researchers' intention to interpret the data. This concern might leave room for misinterpretation and introduce a level of uncertainty about the validity of any qualitative study. The research context is a dynamic process where the researchers play a central role; thus, sensitivity to the context and the development of the relationship with participants is vital to its success. It also acknowledges that data interpretation is highly subjective, and interpretations need to be well-evidenced and grounded in raw data. To ensure the trustworthiness of data analysis and interpretation, strategies to establish quality, substantiate the accuracy of data, and verification of methods, meanings, and interpretations of data will be carried out. The data analysis procedures, verification procedures, and procedures to establish quality help the

researcher decide whether the study is credible, dependable, confirmable, and transferable (Creswell, 2008).

Significance of the Study

The significance of this study is four-fold. First, the proposed study focuses on a phenomenon that has not been previously studied within the context of Malaysia. The study results are anticipated to significantly contribute to understanding how language is used to manipulate and deceive blind investors into falling into their scams.

Secondly, this study aims to develop a model of the language used by cyber scammers to provide a comprehensive public education to reduce the risks of financial loss. Through the examination of general language nature, linguistics styles, and persuasive strategies, it is hoped that the study can demonstrate the specific discourse pattern used to scam is discernible and formulaic. Such findings will contribute substantially towards developing a model of linguistic discourse structure, which can raise awareness of such widespread crimes.

Thirdly, the significance of this study may inform relevant authorities to develop and implement better crime prevention methods and improve cybersecurity measures. A monitoring system can be developed as a defense mechanism to combat cybercrime by identifying similar modus operandi in promoting illegal investment schemes.

Finally, the results of this study might be in consideration to be used as a contribution to the criminal justice system in presenting more significant linguistic evidence in the court of law. The vulnerability of victims to such lies and fraud deserves to be proven by a close examination by forensic linguists to preserve their rights and repossession of invested amount.

Conclusion

As the world is scurrying towards a digitalised lifestyle, the need to legitimize forensic linguistics analysis in providing empirical evidence and considerations in legal cases appears to be an urgent necessity. Although the advancement of technology has provided convenience to daily activities, it is inevitable to ignore that it also comes with its opposing sides. The endeavour of this study may become a pioneer for the use of forensic linguistic analysis as forensic science and accepted as concrete arguments in the justice system, or at least as a tool in the examination of language evidence to assist in criminal investigations

References

- Creswell, J. (2008). *Education Research: Planning, Conducts and Evaluating Quantitative and Qualitative Research* (Third Ed). Upper Saddle Revel, NJ: Merrill.
- Fraenkel, J., & Wallen, N. (2006). *How to design and evaluate research in education*. New Jersey: McGraw-Hill.
- Hartwig, M., & Voss, J. A. (2017). *Lie Detection Guide: Theory and Practice for Investment Professionals*. Virginia: CFA Institute.
- Hashim, M. S., Rahman, N. A., Zin, A. N., Mahamad, K. A., & Ahmad, F. A. (2017). Cyber-Related Fraud Incidents in Malaysia. A Seven-Year Analysis of MyCERT Data. *International Journal of Information Security and Cybercrime*, 9 - 16.
- 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.
- Jianying, X. (2018). An analysis model of network illegal fund-raising. *Advances in Intelligent Systems Research*, 48-51.

- Malaymail. (2019). <https://www.malaymail.com>. Retrieved from <https://www.malaymail.com/news/malaysia/2019/11/02/rm400m-lost-to-cyber-crime-in-10-months-say-police/1806272>
- Malaymail. (2020). <https://www.malaymail.com>. Retrieved from <https://www.malaymail.com/news/malaysia/2020/07/17/bukit-aman-be-careful-of-the-ghost-behind-the-computer-screen-scammers-gett/1885291>
- Schaffer, D. (2012). The language of scam spams: linguistic features of " Nigerian fraud" e-mails. *ETC: A Review of General Semantics*, 157-179.
- 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.
- Tabron, J. L. (2016). Linguistic features of phone scams: A qualitative survey. 11th Annual Symposium on Information Assurance (ASIA'16), 52-58.
- Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In B. M. Wildemuth, *Applications of social research methods to questions in information and library science* (pp. 318-328). Santa Barbara, California: Library Unlimited.