

# Visualising Deception: A Linguistic and Word Cloud Analysis of Malaysian Scam Discourse

Nursyaidatul Kamar Md Shah\*, Ameiruel Azwan Ab Aziz, Ariff Imran Anuar Yatim, Amirah Mohd Juned

Academy of Language Studies, Universiti Teknologi MARA, Cawangan Melaka, MALAYSIA

\*Corresponding Author

DOI: <https://dx.doi.org/10.47772/IJRISS.2025.913COM0039>

Received: 01 September 2025; Accepted: 06 September 2025; Published: 06 October 2025

## ABSTRACT

This study examines the linguistic strategies employed by scammers in the context of the Malaysian *Emas Lelong* scam by using word cloud visualisation to reveal lexical patterns in deceptive communication. Drawing from a corpus of scammer-authored messages, six word cloud tools were used to surface frequent lexical items, highlighting the consistent use of relational address terms, transactional verbs, institutional references, and directive expressions. These patterns were analysed through the lenses of Speech Act Theory, Politeness Theory, Interpersonal Deception Theory, and Forensic Linguistics to uncover how scammers construct trust, simulate authority, and guide victim behaviour. The findings demonstrate that even surface-level frequency analysis can offer meaningful insights into scammer intent, especially when interpreted through linguistic frameworks. While word clouds lack syntactic and contextual depth, they remain valuable tools for exploratory linguistic profiling and public awareness. This study contributes to understanding how language functions as a vehicle for deception and suggests further integration of visual analysis with deeper linguistic and computational approaches.

**Keywords:** forensic linguistics, linguistic deception, scam discourse, Malaysian online scams, word cloud analysis

## INTRODUCTION

Online scams have become a widespread and evolving threat in digital spaces, leveraging technical vulnerabilities and linguistic manipulation to deceive victims. In Malaysia, scam cases involving fake investment schemes, fraudulent online purchases, and imitation brand promotions have surged recently (Kosmo, 2024; Harian Metro, 2024). One particularly notorious example is the *Emas Lelong* scam, in which perpetrators pose as gold traders offering promotional “lelong” (auction) prices through messaging platforms such as WhatsApp, Telegram, and Facebook. Victims are drawn into conversations that appear professional and trustworthy, only to realise later that the promised products never arrive. The success of such scams depends heavily on language, particularly the careful construction of trust, urgency, authority, and social rapport.

Language is a strategic tool in scam communication. Scammers construct persuasive narratives using relational markers, transactional verbs, and legitimising vocabulary (Shah et al., 2023). Previous studies have documented various linguistic strategies in fraudulent communication, including speech acts, politeness markers, authority impersonation, and emotional appeals (Chiluwa, 2015; Male et al., 2021; Ibrahim, 2022). In Malaysian contexts, these linguistic cues are often embedded in informal, mixed-language communication, combining colloquial Malay with English and internet-specific expressions (Aziz et al., 2023; Shah et al., 2024). However, while researchers have analysed scammer discourse through thematic or conversational frameworks, there is limited research on how surface-level linguistic patterns can be quickly and visually explored using accessible tools such as word clouds.

Word clouds are visual representations of the most frequent terms in a text, offering a rapid overview of salient language features. They are widely used in educational, journalistic, and research contexts due to their intuitive

design and low barrier to entry (McNaught & Lam, 2010; Viégas et al., 2009). In deception research, word clouds can help reveal lexical dominance and discourse focus without requiring deep computational analysis. However, most studies employing word clouds do so for illustrative purposes, without critically examining the linguistic implications of the emerging patterns. Moreover, in contexts involving informal, multilingual, or nonstandardised language such as scammer messages, word cloud outputs can vary significantly depending on how the tool handles tokenisation, stopword filtering, and case sensitivity. (Yu & Wang, 2019; Lam & Muniandy, 2021).

This study addresses this gap by exploring how lexical patterns in scammer communication are surfaced through multiple word cloud tools. Rather than comparing the tools for technical performance, the focus here is on understanding the linguistic features that consistently appear in scam messages, and how these features relate to persuasive and deceptive strategies. The data consist exclusively of text authored by scammers involved in verified *Emas Lelong* scam cases in Malaysia. Messages from victims are excluded to ensure a focused linguistic analysis of the scammers' rhetorical strategies. By generating word clouds using six different platforms, this study identifies recurring themes, key lexical choices, and discourse patterns that define scammer language in this context.

Word clouds offer a quick, surface-level language analysis by visually representing lexical frequency. This makes them particularly appealing for exploratory research or public awareness, where rapid identification of dominant themes or persuasive patterns is needed. However, despite their accessibility, word clouds are limited in depth and are susceptible to misinterpretation if used without methodological caution (McNaught & Lam, 2010; Viégas et al., 2009). In scam communication, where subtle linguistic cues and persuasive strategies are critical, even a surface analysis must be scrutinised for accuracy and representational fidelity.

This study examines the linguistic patterns in scammer-authored messages, focusing on relational, transactional, and persuasive markers consistently highlighted across multiple word cloud tools. It also aims to explore how visual representations of word frequency can help identify surface-level deception strategies used in Malaysian online scam discourse. This research highlights the connection between visual text analysis and forensic linguistics by examining how different tools process real-world scammer language. It provides theoretical insights and practical implications, particularly for researchers, educators, and digital investigators seeking to use word cloud tools to analyse deceptive content.

## LITERATURE REVIEW

Scam discourse refers to fraudulent communication practices that manipulate victims by creating perceptions of trust, urgency, or authority. In digital environments, particularly within the Malaysian context, scammers often exploit cultural cues, relational language, and affective expressions to deceive. The *Emas Lelong* scam exemplifies this strategy, as perpetrators present themselves as legitimate gold traders offering discounted jewellery through online channels. Victims are lured into persuasive conversations that mimic trustworthy transactions, only later discovering that the offers were fraudulent.

Existing research has documented how scammers craft messages using carefully selected lexical and rhetorical patterns. For instance, Aziz et al. (2023) found that Malaysian online investment scams rely on imperative structures, exaggerated politeness, and recurring financial terminology to construct believability. Similarly, Juned et al. (2024) highlighted how scammers use emotional appeals, cognitive bias triggers, and credibility-enhancing phrases to influence victims' decision-making processes. These studies underscore the significance of language in shaping the scam narrative and controlling the flow of communication.

In studies focusing on online romance scams, researchers have observed the use of deeply emotional language and politeness markers to develop intimacy and trust over time. Shaari et al. (2019) revealed how scammers strategically use relationship-building phrases and deferential politeness to maintain engagement and extract financial or personal information. These findings reflect the broader patterns of digital deception, where the success of a scam hinges on linguistic manipulation rather than just technological means. The literature affirms that Malaysian scammers consistently employ affective, directive, and persuasive language to maintain control, simulate legitimacy, and ultimately extract financial gains.

This study draws on multiple linguistic and communicative theories to understand the underlying communicative functions of scammer discourse. These frameworks provide insight into how language operates as a vehicle for manipulation, particularly in digital scam contexts. Speech Act Theory (Austin, 1962; Searle, 1969) proposes that language conveys information and performs actions. Scammers engage in various illocutionary acts such as assertives (stating that payment has been made), directives (instructing victims to transfer money), and commissives (making promises of refunds). These acts are designed to construct a legitimate and procedural reality, thereby facilitating deception. Politeness Theory (Brown & Levinson, 1987) helps explain how scammers reduce suspicion and build rapport. By employing both positive politeness strategies (using friendly address terms like “sis” or “cik”) and negative politeness strategies (hedging, apologising for delays), scammers manage impressions and create a cooperative tone. Such strategies are powerful in Malaysian discourse, where politeness norms are culturally significant and often signal sincerity. Interpersonal Deception Theory (Buller & Burgoon, 1996) provides a framework for understanding how deceivers manage their behaviour to appear credible. According to this theory, deception is an interactive process where the scammer adjusts their discourse in response to cues from the target. This includes avoiding direct denial, maintaining consistency, and using ambiguous or vague statements to manage suspicion. These features are frequently observed in long-form scam messages, particularly emotionally driven or high-stakes exchanges. Forensic Linguistics (Coulthard & Johnson, 2010) frames deception as a subject of linguistic evidence. In scam discourse, forensic linguists analyse patterns such as repeated lexical choices, intentional vagueness, and textual structures that serve fraudulent intent. Studies by Aziz et al. (2021) have shown that scammers use predictable patterns such as references to “resit,” “akaun,” and “polis” to simulate legitimacy. This theoretical lens situates scam messages within the context of broader criminal language use and legal accountability. Together, these theories allow for a deeper interpretation of the linguistic patterns identified through word cloud visualisation, moving beyond frequency to functional and strategic analysis.

Word clouds are graphical representations of word frequency within a body of text. While they do not capture syntactic or contextual detail, they offer an accessible way to identify dominant lexical items and emerging themes in discourse. In linguistic research, they are widely used as preliminary tools for exploring textual data, particularly in educational and qualitative studies. In the field of education, Calle-Alonso et al. (2018) used word clouds to visualise student reflections and monitor the focus of collaborative learning. Viégas et al. (2009) demonstrated how word clouds can reveal discourse patterns through repetition and thematic clustering in digital journalism and public communication. Within linguistic studies, McNaught and Lam (2010) noted that word clouds help researchers quickly surface high-frequency terms that may indicate stylistic trends or rhetorical focus.

Although not designed for in-depth discourse analysis, word clouds offer practical advantages in exploratory research. They allow researchers to visually map lexical salience across large datasets and identify starting points for more nuanced theoretical analysis. This is particularly useful in scam discourse, where repeated use of specific terms such as directives, financial verbs, and institutional references signals the presence of patterned rhetorical strategies. However, word clouds are limited in several ways. They do not account for multi-word expressions, grammatical functions, or pragmatic meaning. Additionally, they may misrepresent frequency when informal spelling, abbreviations, or multilingual text is involved. Despite these constraints, word clouds remain valuable for forensic linguistic research as tools for initial detection of lexical markers and framing hypotheses for deeper analysis.

Although research on scam discourse has grown, several gaps remain. First, few studies combine visual lexical analysis with linguistic theory to explore deception. Word clouds are often used illustratively in reports and awareness campaigns, but are seldom integrated with theoretical frameworks such as Speech Act Theory or Politeness Theory. Second, there is limited work on authentic scammer-authored datasets in Malaysia, particularly those composed in informal Malay-English mixtures commonly used in honest scam communication. Third, existing studies tend to focus on victim narratives or police reports, with less attention given to the actual language of the scammers. This study addresses these gaps by applying linguistic theory to a scammer-only dataset and analysing it through word cloud visualisation. Doing so contributes to forensic linguistic scholarship and practical efforts in scam detection and public education.

## METHODOLOGY

### Research Design

This study adopts a qualitative descriptive research design to uncover the lexical and rhetorical patterns scammers use in digital fraud communication. Rather than evaluating the technical functionality of word cloud tools, the analysis focuses on the linguistic content surfaced through visual frequency representation. Word clouds identify dominant lexical items contributing to deceptive messaging, particularly concerning personal address terms, financial references, directive verbs, and legitimacy framing.

The purpose of using word clouds is to gain a surface-level view of which words appear most frequently in scammer messages and how these words reflect broader communicative strategies. This approach allows for a quick mapping of lexical salience that may point to underlying persuasive intentions. The outputs from six different word cloud tools are examined collectively, not to compare their efficiency, but to determine which linguistic elements remain consistently visible regardless of visual layout or processing differences. This reinforces the central focus of the study, which is to dissect scam language rather than assess the tools themselves.

The research is exploratory in nature and grounded in forensic linguistic inquiry, where the goal is to make visible the structure and style of deceptive discourse in a specific scam context. It aims to support further research on scam detection and public awareness through linguistically informed visualisation methods.

### Dataset and Sampling

The dataset for this study was collected from publicly available social media channels involved in the widely reported *Emas Lelong* scam in Malaysia, such as WhatsApp, Facebook Messenger and Telegram. This scam involves the fake sale of discounted gold jewellery through online advertisements and direct messages. Victims are persuaded to make payment in advance for items that are never delivered. The selected conversations were identified based on multiple local media reports and victim testimonies that linked these sources to fraudulent activity (Kosmo, 2024; Harian Metro, 2024).

Only text authored by scammers was included in the dataset. All messages sent by victims were deliberately excluded to maintain analytical focus on the rhetorical and lexical strategies of the perpetrators. This ensures that the patterns observed in the word clouds reflect the scammers' language choices without influencing recipient responses. The compiled corpus contains approximately 7300 words, drawn from various promotional and transactional exchanges between the scammers and their targets.

The sampling was purposive and guided by three criteria: (1) the conversation was verified to be part of an actual scam case, (2) the message thread included sustained promotional content, and (3) the data were collected from cases occurring within the last two years. No translations or language corrections were applied to the messages. Informal spellings, mixed Malay and English usage, emotive expressions, and digital markers such as emojis and abbreviations were retained to preserve the natural linguistic features of the discourse. This dataset provides an authentic representation of how scammers engage with victims using casual, persuasive, and structured messaging. It also reflects real-world language use within the Malaysian digital context, making it a suitable foundation for analysing surface-level linguistic indicators of deception through visual frequency tools.

### Data Preprocessing

All scammer messages were extracted and compiled into a single text corpus to prepare the dataset for analysis. Basic cleaning procedures were applied, including removing duplicated phrases, empty responses, and irrelevant symbols like repeated emojis. No translation or standardisation was conducted, as the study sought to preserve the linguistic authenticity and stylistic variation found in the scam messages. The dataset retained nonstandard spellings, informal discourse markers, and code-switched lexical items to ensure that the visualisations reflected the true nature of scammer communication. This was particularly important for maintaining repetition, politeness



strategies, and formulaic requests, forming part of the scammers' rhetorical tactics. Where required by the tool, the cleaned text was converted into plain text format or pasted directly into the online input interface.

This study employed six established word cloud tools for comparative analysis: Atlas.ti24, WordClouds.com, WordArt.com, Voyant Tools, JasonDavies.com, and TagCrowd.com. These tools were selected based on their demonstrated use in linguistic, qualitative, or educational research and diverse text visualisation approaches as summarised in Table 1.

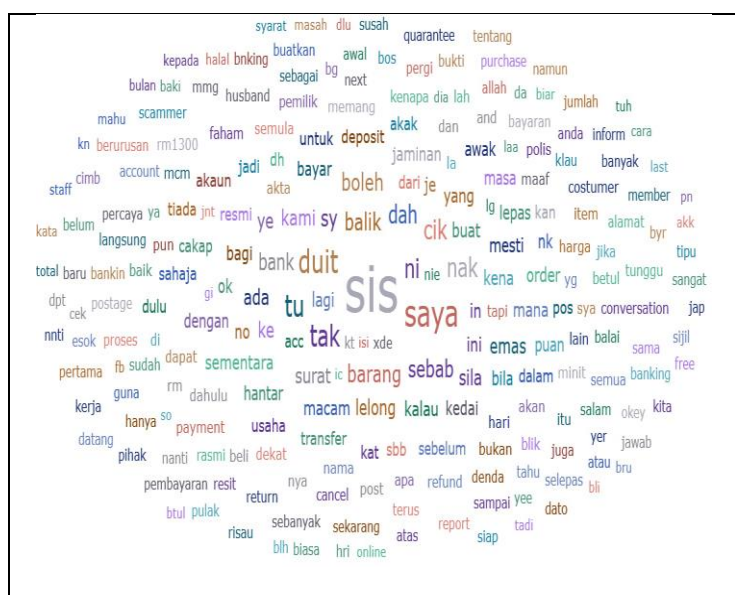
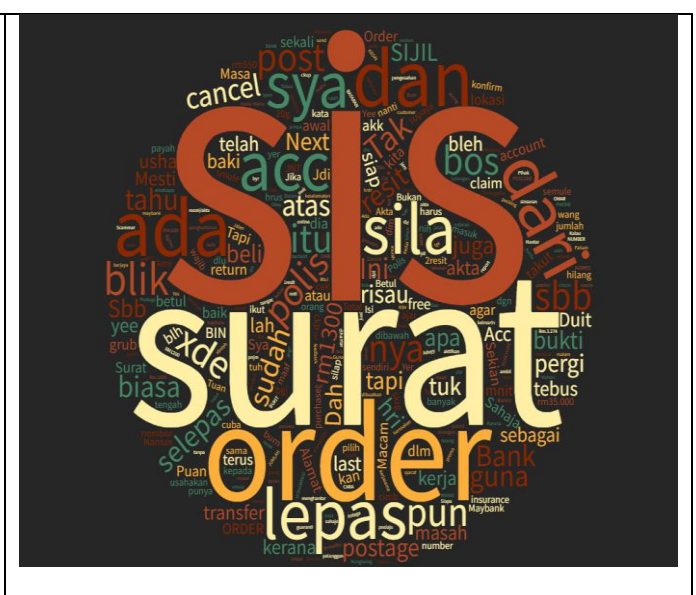
Table 1: Word Cloud Tools


Tool Name	Developer / Source	Research Use / Citation
Atlas.ti24	ATLAS.ti Scientific Software	Used in qualitative and linguistic analysis (Yusof & Ghazali, 2020)
WordClouds.com	Zygomatic	Employed in educational reflection studies (Ahmad et al., 2022)
WordArt.com	WordArt.com	Used to enhance vocabulary and engagement in ESL (Kassim et al., 2021)
Voyant Tools	Sinclair & Rockwell (2016)	Common in digital humanities and discourse analysis (Nicolaidis et al., 2018)
JasonDavies.com	Jason Davies	Cited in visualisation pedagogy and comparison layouts (Lohmann et al., 2012)
TagCrowd.com	Daniel Steinbock	Early word cloud tool used in qualitative themes (McNaught & Lam, 2010)

Each of the six tools was provided with the same scammer-only dataset drawn from *Emas Lelong* scam cases. Inputs were submitted in raw or plain text format, depending on the tool. The visual outputs served as the foundation for evaluating how linguistic features, particularly persuasive terms, transactional verbs, and personal address markers, were emphasised differently across the various platforms. The findings from each tool were synthesised to highlight methodological implications.

## RESULTS AND FINDINGS

The analysis of six word clouds generated from the scammer-only dataset revealed consistent linguistic patterns that reflect scammers' communicative strategies. Across all tools, high-frequency terms illustrated a repeated reliance on personal address, transactional vocabulary, institutional references, and persuasive markers. These patterns serve as lexical evidence of fraudulent behaviour and reveal the communicative techniques used to build credibility, pressure victims, and construct a facade of legitimacy.

	
Atlas.ti24	Word clouds.com

	
<p>Word Art.com</p>	<p>Voyant Tools</p>
	
<p>Jason Davies.com</p>	<p>Tag Crowd.com</p>

One of the most dominant features was the repeated use of relational address terms, particularly the word “sis”, which appeared prominently in all six visualisations. Its frequent use suggests an intentional strategy to foster familiarity, trust, and casual rapport. Alongside “sis,” other forms of polite or familiar address, such as “cik,” “puan,” and “kami”, were also common, reflecting a calculated effort to personalise the interaction and lower suspicion. These choices contribute to the social engineering aspect of the scam, allowing the scammer to appear approachable and respectful.

Another salient lexical cluster involved transactional and financial terms such as “duit,” “bayar,” “deposit,” “akaun,” “transfer,” and “refund.” The recurrence of these items underscores the primary purpose of the scammer’s discourse: to persuade the victim to send money or provide banking details. The variation in these terms also reflects the different stages of the scam process, from initial payment requests to promises of refunds or returns. This pattern mirrors typical scam narratives where urgency is introduced through requests for payment, followed by reassurances that the process is legitimate and reversible.

The dataset also featured institutional and formal vocabulary, including terms like “surat,” “resit,” “balai,” “polis,” “akta,” and “sijil.” These words were often embedded within persuasive messages intended to project legal or procedural authority. The prominence of “surat” and “resit” suggests that scammers deliberately used bureaucratic or administrative language to increase credibility and suggest formality. Similarly, references to enforcement agencies or regulations such as “polis” and “akta” may be employed to introduce consequences or pressure for compliance.



The word clouds also revealed a pattern of instructional and directive language. Words such as “*sila*,” “*balik*,” “*buat*,” “*hantar*,” “*isi*,” and “*tunggu*” were frequently used, indicating the step-by-step nature of the scammer’s approach. This form of language positions the scammer as a figure of guidance or authority, instructing the victim through each phase of the fraudulent process. The frequent use of imperatives softens the demand by embedding them within polite expressions such as “*sila buat pembayaran*”, masking coercion as procedural instruction.

Emotional appeals and justifications also surfaced through words such as “*sebab*,” “*kena*,” “*risau*,” “*jaminan*,” and “*balik*.” These terms indicate persuasive narratives that exploit the victim’s emotions, often by providing reasons for delays, invoking urgency, or offering guarantees. The presence of these lexical items reflects an empathetic yet manipulative tone, characteristic of scams that rely on perceived sincerity and shared understanding.

Moreover, several function words and discourse particles in informal Malay, such as “*ni*,” “*tu*,” “*je*,” “*tak*,” and “*ye*”, appeared frequently. These items, often excluded in standard academic analysis, are integral to the conversational tone of scam discourse. Their consistent visibility across tools reflects the naturalistic style of the scammers’ messages and reinforces the need for language analysis tools that accommodate informal or mixed-register input.

The linguistic patterns observed in the visualised data suggest a coherent communicative strategy that blends relational warmth, transactional clarity, formal authority, and persuasive reasoning. The scammers’ lexical choices are not random but reflect deliberate rhetorical choices designed to build trust, guide actions, and reinforce legitimacy. These patterns offer valuable insights for linguistic profiling and can inform the development of automated detection tools or training materials for public awareness.

## DISCUSSION

The linguistic patterns visualised through the word clouds reveal consistent strategies scammers use to manipulate, persuade, and control the interaction with victims. Rather than being random or casual, these lexical choices reflect deliberate attempts to construct social familiarity, simulate professionalism, and reinforce legitimacy. This study builds on forensic linguistic perspectives, where language is central to committing and concealing fraudulent acts (Coulthard & Johnson, 2010). By drawing from Speech Act Theory (Searle, 1969), Politeness Theory (Brown & Levinson, 1987), and Interpersonal Deception Theory (Buller & Burgoon, 1996), the findings underscore how surface-level lexical frequency can reflect deeper communicative intentions.

Relational address terms such as “*sis*,” “*cik*,” and “*puan*” appeared prominently in all visualisations. Their frequent use is a social anchoring device to build rapport and reduce suspicion. According to Politeness Theory, such terms operate as positive politeness strategies that reduce social distance and increase the likelihood of compliance. In the Malaysian context, this linguistic familiarity aligns with cultural expectations of friendliness and respect, making the scammer appear trustworthy and relatable (Male et al., 2021; Ibrahim, 2022). Though seemingly casual, these terms form part of the scammer’s overall social engineering tactics.

The dominance of transactional lexis, including “*duit*,” “*bayar*,” “*deposit*,” and “*akaun*,” reflects the central purpose of scam discourse, which is financial extraction. These words function as the backbone of the fraudulent narrative, embedding monetary references in routine and legitimate ways (Aziz et al., 2023). From a speech act perspective, these terms are often delivered through directives, whereby the scammer instructs the victim to make payments, share account details, or confirm transfers. Using polite modifiers such as “*sila*” and “*harap*” softens these imperatives, masking coercion in the form of procedural guidance. This blending of directive acts with politeness strategies reflects what Searle (1969) describes as indirect speech acts, where language is shaped to appear cooperative while pursuing self-serving goals.

The frequent appearance of institutional or legal vocabulary such as “*surat*,” “*resit*,” “*polis*,” and “*akta*” serves to legitimise the interaction. These lexical items act as assertives, in which the scammer presents statements intended to simulate formal authority or bureaucratic procedures. Chilwa (2015) highlighted that impersonation and reference to official institutions are standard features of scam narratives designed to evoke trust or fear. The

strategic placement of such vocabulary reinforces the perception that the scammer operates within a legitimate framework, thereby neutralising suspicion and increasing compliance.

In addition, the presence of commissive and expressive speech acts further illustrates the scammers' rhetorical flexibility. Words like "*sebab*," "*jaminan*," and "*balik*" suggest promises, reassurances, and explanations intended to maintain victim engagement and deflect doubt. According to Interpersonal Deception Theory, such expressions are essential to sustaining the illusion of sincerity and controlling the victim's emotional state (Buller & Burgoon, 1996). These acts function not only to delay disengagement but to extend the narrative and deepen the victim's emotional and financial commitment.

Function words and informal Malay particles such as "*ni*," "*tu*," "*tak*," and "*ye*" also appeared frequently, reflecting a conversational tone that is consistent with natural chat-based interactions. Though often overlooked in formal linguistic analysis, these elements help shape the scammer's persona as approachable and spontaneous, lowering the victim's guard. Their visibility across tools also reinforces the value of word cloud analysis in contexts involving informal or mixed-register communication.

The consistent lexical patterns observed across tools confirm that scammers strategically employ language to build trust, instruct behaviour, and simulate legitimacy. These surface-level findings reveal a more complex network of communicative functions beyond word frequency when interpreted through theoretical lenses. This underscores the importance of what is said and how it is said through a combination of speech acts, politeness strategies, and emotional appeals embedded in discourse. Although word clouds are limited in their analytical depth, they remain valuable for surfacing patterns that prompt further linguistic investigation or public awareness efforts.

## LIMITATIONS

This study is limited by the relatively small dataset, consisting of approximately 7,300 words of scammer-authored messages taken from verified Emas Lelong scam cases in Malaysia. While this provided authentic linguistic evidence of deceptive strategies, the size of the dataset constrains the extent to which the findings can be generalised to other scam types or wider populations of deceptive communication. Another limitation lies in the exclusion of victim responses. By focusing exclusively on scammer-authored texts, the analysis highlights how scammers design and deliver persuasive strategies, but does not capture how these strategies shift in response to victims' replies. Since scammers frequently adapt their tactics to manage suspicion or resistance, the absence of victim discourse restricts the study's ability to examine deception as an interactive process. These limitations should be considered when interpreting the results, and they point to the need for larger and more balanced datasets that incorporate both scammer and victim discourse in future studies.

## CONCLUSION

This study has demonstrated how Malaysian Emas Lelong fraud scammers employ language for deception and persuasion. The analysis surfaced prominent lexical items that reveal consistent communicative strategies by applying word cloud visualisation to a corpus of scammer-authored messages. Interpreted through Speech Act Theory, Politeness Theory, and Interpersonal Deception Theory, these patterns reveal how language functions as action and manipulation. Directives, commissives, and assertives are deployed to compel victims to act, reassure them of legitimacy, and present fraud as routine transactions. Politeness strategies further reduce suspicion by framing demands as respectful requests, while deception is sustained interactively by offering reassurances and explanations when needed. These insights confirm the value of linguistic theory in exposing the mechanics of fraudulent communication, showing that deception is achieved not only through what is said but also through how it is said.

Methodologically, this study demonstrates that word cloud tools, while often regarded as simplistic, can provide meaningful entry points for forensic and applied linguistics. Their ability to highlight surface-level frequency patterns allows researchers and practitioners to quickly identify recurring markers of persuasion and deception. When combined with theoretical frameworks, word clouds can move beyond illustration and serve as useful exploratory devices in scam detection, digital literacy campaigns, and even preliminary forensic investigations.



This highlights the importance of critical interpretation, since visualisation alone is not neutral but must be contextualised through linguistic analysis.

Beyond its academic contributions, the study also carries practical implications. For educators, word clouds offer a simple yet effective way to raise awareness of deceptive strategies, making them valuable tools for teaching digital literacy. The identified patterns can inform scam detection frameworks for investigators and analysts by providing lexical markers of persuasion and fraud. For the wider public, this research underscores how everyday language can be manipulated to commit a crime, encouraging greater vigilance in online interactions.

Looking ahead, there is significant potential to build on these findings. Future research could incorporate larger and more varied datasets that cover different scam types, platforms, and cultural contexts to assess whether the identified strategies are universal or locally specific. Integrating victim discourse would provide insight into the interactive dynamics of deception, capturing how scammers adapt to resistance or suspicion. Advances in computational linguistics also offer opportunities to combine word cloud outputs with machine learning models, enabling automated detection of scam language at scale. Finally, comparative research across regions could deepen understanding of how cultural and linguistic norms shape the language of fraud.

In conclusion, this study has shown that scam discourse is neither random nor simplistic. It is a carefully orchestrated communication that draws on linguistic resources to create trust, urgency, and legitimacy. By situating word cloud analysis within theoretical frameworks of speech acts, politeness, and deception, the research provides both scholarly and practical contributions to the study of fraudulent language. It emphasises the enduring importance of linguistic awareness in combating digital crime and invites continued interdisciplinary engagement at the intersection of language, technology, and security.

## ACKNOWLEDGMENT

This research was supported by the *Skim Geran Dalaman* TEJA 2024, Universiti Teknologi MARA (UiTM), Cawangan Melaka (Grant Code: GDT2024/1-3).

## REFERENCES

1. Aziz, A., & Azhar, S. (2021). Modelling the language of cyber scammers: A forensic linguistic approach. *Asian Journal of Forensic Sciences*, 4(1), 12–25.
2. Aziz, A., Juned, S., & Rahman, A. (2023). Linguistic cues of deception in Malaysian online fraud. *International Journal of Applied Linguistics and English Literature*, 12(2), 56–67.
3. Austin, J. L. (1962). *How to do things with words*. Oxford University Press.
4. Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge University Press.
5. Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, 6(3), 203–242.
6. Calle-Alonso, F., Calle-Martínez, C., & Botón-Fernández, V. (2018). Word clouds as a learning analytic tool for cooperative work in online higher education. *International Journal of Educational Technology in Higher Education*, 15(38), 1–18.
7. Chiluwa, I. (2015). Discourse strategies of cyber fraudsters in Nigeria. *Journal of African Media Studies*, 7(2), 191–206.
8. Coulthard, M., & Johnson, A. (2010). *An introduction to forensic linguistics: Language in evidence* (2nd ed.). Routledge.
9. Harian Metro. (2024). Penipuan emas lelong semakin berleluasa. Harian Metro. <https://www.hmetro.com.my/>
10. Ibrahim, F. (2022). Manipulative language in Malaysian online scams: A discourse-pragmatic perspective. *GEMA Online® Journal of Language Studies*, 22(3), 45–61.
11. Juned, A. M., Aziz, A. A., Sharif, N. A. M., Shah, N. K. M., Yatim, A. I. A., & Fakhruddin, W. F. W. W. (2024). The language of lies: An analysis of deceptive linguistic cues on Malaysian investors' decision making. *International Journal of Research and Innovation in Social Science*, 8(10), 668–673. <https://doi.org/10.47772/IJRISS.2024.8100056>

12. Kosmo. (2024). Sindiket emas lelong tipu ratusan mangsa. Kosmo Online. <https://www.kosmo.com.my/>
13. Lam, K. S., & Muniandy, B. (2021). Using corpus linguistics to detect patterns in phishing emails. *Pertanika Journal of Social Sciences & Humanities*, 29(1), 455–469.
14. Male, M., Yusof, N., & Rahim, H. A. (2021). Discursive legitimization strategies in online scam messages. *GEMA Online® Journal of Language Studies*, 21(4), 39–56.
15. McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *The Qualitative Report*, 15(3), 630–643.
16. Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.
17. Shaari, A. H., Kamaluddin, M. R., & Zakaria, N. Z. (2019). Online romance scam in Malaysia: An analysis of victimisation and deception. *Journal of Language and Communication*, 6(1), 89–102.
18. Shah, N. K. M., Ab Aziz, A. A., Juned, A. M., Yatim, A. I. A., & Fakhruddin, W. F. W. W. (2024). Scheming in Syntax: Analysing Scammer-Victim Conversations in Malaysian E-Commerce Scams. *3L, Language, Linguistics, Literature*, 30(4), 47-59.
19. Shah, N. K. M., Aziz, A. A. A., & Saidalvi, A. (2023). A Linguistic Analysis of Authorship Attribution in E-Commerce Scams' Promotional Contents and Narratives. *International Journal of Academic Research in Business and Social Sciences*, 13(1), 996 – 1004.
20. Viégas, F. B., Wattenberg, M., & Feinberg, J. (2009). Participatory visualisation with Wordle. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1137–1144.
21. Yu, H., & Wang, S. (2019). A comparative study on text visualisation tools: Wordle, TagCrowd, and WordArt. *Journal of Visual Languages and Computing*, 50, 74–86.