

Lexical and syntactic features in local online commercial posts of buy-and-sell groups in Leyte

Cabornay, John Nicole ✉

Leyte Normal University, Philippines (2002878@lnu.edu.ph)

Received: 31 January 2026

Available Online: 9 March 2026

Revised: 20 February 2026

DOI: 10.5861/ijrse.2026.26809

Accepted: 5 March 2026

ISSN: 2243-7703

Online ISSN: 2243-7711

OPEN ACCESS



Abstract

This study analyzed the lexical and syntactic features present in the online commercial posts of local "buy-and-sell" groups in Leyte, Philippines. Utilizing a descriptive qualitative, corpus-based methodology, the research investigated language use in this digital marketplace. The analysis revealed that the linguistic landscape is characterized through intentional modifications that prioritize efficiency and community-building. Lexically, code-switching between Waray, Filipino, and English is foregrounded which served as a medium to build rapport and a sense of shared identity, while also integrating technical terminology. Moreover, the prevalent use of abbreviations and acronyms like "PM" and "COD" ensures conciseness and speed in communication. Syntactically, the discourse deviates from formal writing through the employment of broken sentences and listed structures. This list-oriented approach enhances scannability which allow buyers to swiftly discern essential information. Furthermore, the corpus is rich in imperative statements that functions as direct calls to action that guide the transaction process. In order to meet the demands of the fast-paced digital marketplace, features such code-switching, abbreviations, and a broken imperative syntax turn the posts into effective, bite-sized, and community-oriented tools that prioritize speed and functionality over traditional grammatical correctness.

Keywords: lexical features, syntactic features, online commercial discourse, code-switching, buy-and-sell groups

Lexical and syntactic features in local online commercial posts of buy-and-sell groups in Leyte

1. Introduction

The constant change of the dissemination of digital technologies and social media platforms have altered human communication wherein it both facilitated the evolution of new linguistic practices and the transformation of existing ones. The Philippines has consistently been ranked among the highest in global social media utilization. In line with this, online platforms have become indispensable for the development of burgeoning commercial activities and social interaction (DataReportal, 2024). This phenomenon is evident in local online "buy-and-sell" groups. These marketplaces facilitate daily interactions, negotiations, and transactions between individuals and small enterprises. Such groups are hosted on platforms like Facebook which are indicative of a linguistic ecology that can be described by a combination of formal sales discourse and informal conversational elements.

However, the impetus for this investigation is directly informed by the fact that the specific language use of these online communities is an under-researched area, despite the widespread adoption and linguistic practices observed, in the province of Leyte, there is still a gap of concentrated linguistic research that investigates the specific lexical and syntactic features employed within these local digital marketplaces. Hence, a knowledge of these aspects can clarify the relationship between language, commerce, and technology in non-formal, vernacular environments (Tuan, 2023). Secondly, the language that these groups employ frequently exhibits a unique lexical choice and syntactic pattern, which is a result of a hybridization of local languages such as Waray, Filipino, and English. This creates an opportunity to investigate the practical and community-specific operationalization of multilingualism through the examination of its influence on sentence construction and word selection (Darus & Ching, 2020; Mahdi, 2020). Furthermore, despite previous research has investigated multilingual practices in Philippine online contexts which gives emphasis on educational or general social interactions (Garcia, 2021; Ong, 2022), this study specifically aimed to broaden these inquiries that focused on the specialized field of online commerce in Leyte which tackled how different languages contribute to the lexical and syntactic characteristics of the discourse. Finally, the identification of the most prevailing linguistic strategies for online commerce can provide insights to local entrepreneurs, digital marketing professionals, and language educators who are interested in functional language use and sociolinguistic variation from an applied linguistics perspective. Moreso, this investigation aimed to enhance our understanding of vernacular online commercial discourse which is a field that remains largely unexplored in regional Philippine contexts through the analysis the language of in the digital context.

The variables that were analyzed by the study are the linguistic features, specifically syntactic patterns and lexical choices. These features are implicitly influenced by contextual variables such as Facebook, the platform that the researcher focused on, the specific genre/register used in online commercial posts, the geographic context which is Leyte, and the time period of data collection which will ensure that the language data is contemporary.

The corpus of literature on online language has conducted a comprehensive examination of general patterns in social media discourse (Darus & Ching, 2020; Hyland, 2021). In the Philippine context, different research has investigated multilingual practices in a diverse array of digital interactions (Garcia, 2021; Ong, 2022; Villamor et al., 2025). Consequently, in formal setting, mass media contexts like print or broadcast are frequently the subject of research (Tuan, 2023; Zhu, 2020). However, in terms of informal online commercial language, it receives only limited attention. To elucidate, there are few studies that conduct a comprehensive linguistic analysis of user-generated content which are focusing on lexical and syntactic features, in local online marketplaces within a regional context like Leyte, even though some research examines consumer language (Kronrod & Pogacar, 2023)

or digital marketing communication (Rammel & Herbig, 2022). Hence, this field remains unexplored by researchers.

Theoretical Framework - This study was anchored on three interconnected theories. Specifically, it integrated Social Information Processing (SIP) Theory (Walther, 1992), Theory of Conversational Code-Switching by Gumperz (1982), and The Principle of Least Effort (Zipf, 2016). Through the intersection of these three domains, the researcher was able to create a lens for understanding how linguistic choices are adapted to the digital marketplace environment.

Social Information Processing Theory (Walther, 1992) - This theory explains how users in online marketplaces adapt their linguistic behavior to build trust and relationships despite the lack of physical interaction. It also accounts for the use of digital-specific lexical features like acronyms and emoticons as substitutes for non-verbal cues.

Conversational Code-Switching (Gumperz, 1982) - In a Sociolinguistic perspective, this clarifies the intentional alternation between Waray, Filipino, and English. Specifically, in this study, code-switching is perceived as a method used by sellers to balance local identity with professional credibility.

The Principle of Least Effort (Zipf, 2016) - This principle was utilized to address the syntactic structures of the corpus. It posits that in high-speed digital environments, language naturally trends toward brevity. Hence, this explains the prevalence of listed structures and imperative fragments in posts. Additionally, this maximizes informational output while minimizing the cost of production and processing time for the reader.

Although, SIP provides the foundation for understanding constraints in the digital environment, it does not fully account for the regional linguistic variation of the Leyteño context. Therefore, there is a need to integrate Conversational Code-Switching to address multilingual identity and the Principle of Least Effort to explain the syntactic brevity observed in the corpus.

Statement of the Problem - This study aimed to analyze the lexical and syntactic features in the online commercial posts of buy-and-sell groups in Leyte. Specifically, it sought to answer the following questions:

1. What are the main lexical and syntactic features used in the online commercial posts of buy-and-sell groups in Leyte?
2. How does the use of these lexical and syntactic features affect the meaning of these online commercial posts?

2. Methodology

Research Design - A qualitative research design, specifically a descriptive qualitative approach was utilized in this study. Specifically, it employed a corpus-based methodology. Furthermore, this design is aligned in the objectives of this research as it offered a comprehensive, contextualized, and detailed comprehension of the lexical and syntactic features that are present in online commercial language. The corpus acted as the authentic linguistic data for in-depth textual analysis.

Selection of the Corpus of the Study - The corpus of this study are publicly available online posts on Facebook. A purposive sampling method was employed to construct it. This was utilized to identify the online buy-and-sell groups and posts for the research questions and to guarantee the representativeness of the target linguistic phenomena (Laghari, 2024). The following criteria were used to select online posts and comments for inclusion in the corpus:

First, posts and comments that originate from public Facebook groups that are explicitly designated as "buy-and-sell," "marketplace," "online bazaar," or similar commercial trading groups, with a primary focus on residents of the Leyte province.

Second, only content from groups that are publicly accessible or that are readily joinable for observation purposes. Third, posts and comments that are evidently intended for or pertinent to commercial transactions within the geographical region of Leyte. Fourth, content in which the primary language of interaction is Waray, Filipino or English, or frequently code-switches between these languages. Lastly, posts and their directly relevant comments that were published within the specified recent time period, specifically from April 1, 2025, to June 30, 2025. Original sales posts and their directly related comments that facilitate a commercial transaction were also noted. Nevertheless, the researcher focused on posts and comments that are primarily textual and contain an adequate amount of linguistic content.

Data Gathering Procedure - The data collection process adhered to established methods for corpus construction from online and user-generated content. These steps were adopted from the research studies of Laghari (2024) and McEnery & Hardie (2021):

The initial step of this research focuses on the formulation of the research questions. This phase is essential for the establishment of the investigation's parameters and for the direction of all subsequent data collection endeavors. Concurrently, the corpus' scope was meticulously defined. This guaranteed that there is an adequate volume of data to conduct a comprehensive linguistic analysis, which includes establishing a target word count of approximately 10,000 words. The nature of the desired linguistic data was characterized, with specific emphasis on the text generated during online commercial interactions. More so, this text is a combination of informal, persuasive, and transactional language. To identify potential data sources, a preliminary and exhaustive search was conducted on the Facebook platform. Publicly accessible and active "buy-and-sell" groups in Leyte, Philippines were the focus of this search. The researcher used keywords such as "Leyte buy and sell," "Leyte B&S," and "Tacloban Marketplace." A preliminary inventory of appropriate groups was compiled during this exploratory phase, from which the final data sources will be selected.

In order to guarantee that the data collection is both representative and relevant to the research questions, a selection procedure was implemented. A set of predefined inclusion criteria was used to select 3 to 5 groups from the initial inventory of identified groups. Factors such as group activity which focused on recent and frequent posts, member count indicated a diverse user base, and relevance of content which the primary function of the group, comprised these criteria. The study ensured a focused and manageable corpus while still capturing a range of local interactions through limiting the number of groups and establishing selection criteria. Upon completion of the selection procedure, the researcher manually accessed the selected public Facebook groups. This was accomplished by utilizing either the official Facebook mobile application or a standard web browser on a desktop computer. The researcher collected data without engaging with or participating in the group's activities, who will only act as an observer within these public spaces. This method was a decision to guarantee a controlled and ethical approach to data acquisition from public social media platforms.

To construct a corpus that is both representative and balanced, a purposive sampling strategy was implemented within the designated groups. The initial posts, which include information such as product descriptions, pricing details, and calls to action, were prioritized in the sampling. Relevant comments that explicitly contribute to the transactional discourse such "how much?", "is this available?", and "pm sent" will also be sampled. Nevertheless, it will be avoided to include lengthy remark chains that diverge from the fundamental commercial interaction. The textual content of the sampled posts and comments were manually copied and extracted. This means that underwent a process of highlighting and duplicating the text from the Facebook interface. Furthermore, this manual method is an ethical decision to circumvent the use of automated data harvesting tools, such as web scraping programs, as recommended by best practices for social media data research (Franzke et al., 2020). The researcher in return minimized the inclusion of extraneous web elements and adheres to ethical guidelines through manual extraction of the data to ensure that only necessary linguistic information is collected.

Next, a thorough initial cleaning procedure was implemented on the extracted raw text. This phase was

essential for the preparation of the data for linguistic analysis, as it eliminated all extraneous, non-linguistic elements. The cleaning process was to identify and remove any other symbols or characters that do not form part of the core linguistic content, as well as identities, timestamps, emoticons, image descriptions, hyperlinks, and advertisements (Laghari, 2024). This made sure that the corpus is exclusively composed of the language employed in online commercial interactions.

The preprocessed text was converted to a standardized format following the initial cleansing. The text that has been cleansed was saved as standard plain text files with a.txt extension. Additionally, to facilitate manual annotation and close reading during the analysis phase, the text was formatted for readability. This may involve the organization of the text by post, with distinct identifiers, to preserve the context of each linguistic sample. This preprocessing phase converts the data into a structured format that is prepared for systematic linguistic investigation.

The unified corpus was created by compiling the cleansed and preprocessed text files. This led to the creation of one or more plain text files, which collectively functioned as the primary data source for the entire study. Each file was named in a way that reflects its origin to facilitate simple identification and administration. Example of this is "Group_A_Posts_1.txt," or "Group_B_Posts_2.txt". The successful construction of the corpus is signified by this compilation step and it will signal the transition from data collection to data analysis.

The corpus was constructed successfully, and the systematic manual linguistic analysis commenced. The methodology for this phase was as specified in the "Data Analysis" section of the research paper. To identify and categorize the lexical and syntactic features specified in the study's framework, the analysis had an examination of the preprocessed text. The research questions are then addressed in this final step in terms of applying the defined methodology to the collected data.

Data Analysis - This study utilized inductive qualitative content analysis to conduct data analysis, with an emphasis on the detailed description and interpretation of lexical and syntactic features in the collected corpus. This process means that the researcher conducted systematic close reading, iterative classification, and pattern identification of linguistic elements to reveal the ways in which they influence communicative intent and meaning. First, the researcher read the corpus on numerous occasions to gain a comprehensive understanding of the data. A prominent lexical and syntactic pattern that emerges from the online commercial discourse will be identified using an open coding approach. Then the analysis was characterized by vocabulary, recurring commercial jargon and semantic fields present in the posts for lexical features. To reiterate, the main objective was to comprehend the persuasive functions and connotative meanings of these word choices in their respective contexts. Next, the analysis described the typical sentence structures, foregrounded clause types, and the use of grammatical patterns such as direct address or imperative mood for syntactic features. The researcher also determined the effect of code-switching on the syntactic structures of these posts.

Ethical Considerations - This research is conducted in accordance with the ethical standards such as the following: The study relied on data obtained from publicly accessible online buy-and-sell groups. The researchers did not make any attempts to access private groups, private messaging, or any content that is not explicitly designated for public viewing. In general, research that aggregates and analyzes publicly posted content in an anonymized manner does not require specific individual consent from content creators, as the data is already in the public domain and freely accessible to anyone with internet access (Franzke et al., 2020; LSE, 2025). The linguistic data that is collected was anonymized. During the data cleaning and formatting phase of corpus construction, any personally identifiable information, including usernames, profile images, unique identifiers, and precise timestamps that could potentially associate content with specific individuals were meticulously removed or obscured. The study was corpus-based, which means that there was no direct interaction with any human. Consequently, the potential dangers associated with participant recruitment, coercion, or direct intervention were eliminated. The objective of this study was to provide empirical insights into the comprehension of contemporary language use in the Philippines. The results have the potential to

provide insights for local entrepreneurs, digital marketers who are interested in optimizing their online communication strategies, and language educators who are interested in sociolinguistics and digital literacy. The final research output provides a comprehensive and transparent account of the entire research process, including the methodology, data sources, corpus construction procedures, and analytical approaches. This dedication to transparency made sure of replicability, permits critical examination by the academic community, and maintains ethical standards of research conduct.

3. Results and discussions

Lexical and Syntactic Structures - The analysis conducted through a systematic process adopted from the research studies of Laghari (2024) and McEnery & Hardie (2021) yielded the following inputs:

In terms of lexical features, the analysis revealed that lexical landscape of Leyte's buy and sell online community is shaped by its multilingual nature. Hence, one of the lexical features foregrounded is code-switching. To elucidate, the languages involved in a single post were Waray, Filipino, and English. The researcher observed that there is a fluid alternation from these languages to provide a clear social and communicative function. An example of this is a post selling a sidecar which had a combination of the aforementioned languages, "Side car only baligya. Bakal iton hiya dere ito pedicab". One implication of this is that it creates a sense of shared identity between individuals and rapport which makes commercial exchange/communication, feel more personal and less transactional. Another post for a computer set utilized a similar approach of advertising: "Affordable ngan good quality naka warranty pa", in this way, the seller can easily connect with the audience and make it more relatable to them. Other posts included:

*Pm me agad sa interesado! paunahay lang! price is very negotiable pm lang pag-usapan natin",
"No repair history! 40k meron pa last price if kunin agad. Can do meet up at Ormoc, Palo,
Guiuan, and Tacloban.*

This finding supports the notion of Ong (2022) that code-switching in digital places is used as a way to build community and establish an authentic voice. Another lexical features present in the posts are abbreviations and acronyms. Examples of this include the internet slangs that users mostly utilize such as "DM" for direct message, "PM" for private message. It also included product terms such as "AC" for air conditioner, "pcs" for pieces and financial terms such as "MOP" for mode of payment and "COD" for cash on delivery. In this way, conciseness and speed are prioritized to enable sellers to convey information quickly and efficiently for mobile users which is aligned to the Social Information Processing Theory by Walther (1992). The post regarding a smartphone shows this: "Iphone 12128gb No issue! Smooth pa, no repair history, 3,500 nala, Pm ha interested".

In lieu of the syntactic structure, it was observed that posts deviated from formal and academic writing. One of syntactic features present in corpus is the employment of broken sentences and a listed structure. The researcher observed that posts often lack a comprehensive subject-verb-object structure, instead it depends on enumerations of essential product characteristics. For instance, a vehicle's specifications are frequently conveyed as a compilation of information: "Toyota Vios, 2018 (50k+++++oddo), Automatic, Complete original papers, Registered, Cold AC...". This aligns well with the Principle of Least Effort by Zipf (1949) as the list-oriented syntax enhances scannability which enables buyers to swiftly discern essential information without the necessity of reading an entire paragraph. It converts the post from a narrative description into a structured data set. Furthermore, it also enhances its utility for the original aim of delivering information succinctly. In addition, the corpus is abundant in imperative statements that serve as direct demands or calls to action for the purchaser. Expressions such as "Pm ha interested" and "Call or txt sa sure buyer" are prevalent. These orders function as a rhetorical way that guides the reader on the subsequent steps of the transaction. This also corresponds with studies in digital marketing literature which further indicates that language on online platforms is frequently optimized to provoke a particular user response (Rammel & Herbig, 2022).

Implications of the Features to Meaning of Posts - To answer the research question 2, the utilization of

lexical and syntactic features in online commercial posts affects meaning through changing it from formal writing to bite-sized, efficient, and community-oriented tools. For instance, code-switching mixes languages like Waray and Filipino to make the tone more approachable for individuals and trustworthy. This allows the sellers to have a conversation with their customers on a personal level and at the same time they are able to use technical terms for their products making it feel more authentic. On the other hand, the use of abbreviation and a broken imperative syntax makes the post quickly readable. Moreover, these features leave from the traditional syntax and instead use direct language for easier comprehension and smooth transaction. Hence, in this way language fits with in the fast-paced nature of digital marketplaces because the main purpose of it is not to advertise using lengthy and flowery words but to inform the buyer immediately.

4. Conclusion

The researcher concludes that the linguistic features identified in the corpus are intentional modifications designed for the digital marketplace. Essentially, code-switching serves two purposes: 1) it develops rapport and a sense of community by speaking the audience's language, and 2) it lets technical or standardized terminology fit in with the local vernacular. Hence, this aligns with the Theory of Conversational Code-Switching by Gumperz (1982) as it shows how language was used intentionally to build trust in an online business setting. Hence, this creates a sense of rapport and community belonging. Conversely, the use of abbreviations and broken imperative grammar is also a practical answer to the ways of the internet. To elucidate, these traits put efficiency, brevity, and clarity ahead of traditional grammatical correctness. These changes the posts from long information into short, action-oriented messages that are easy to scan by the buyers. It also makes the communication and sales process much more efficient. Thus, the researcher also concludes that the language used in these online advertisements is a distinctive dialect that combines words and phrases from different places to fulfill the needs of a regional digital marketplace, and it often embeds speed and functionality for transactions.

Recommendations - Recommendations can be made for stakeholders based on the findings of this study. 1) The strategic value of the linguistic features identified in this corpus should be acknowledged by local entrepreneurs and digital marketers. It is advised that they incorporate abbreviations and code-switching into their online posts to establish a strong connection with their target audience that would also serve as a marketing strategy in other instances. Another one is that through code-switching, they can also cultivate a more relatable and authentic voice that promotes trust and rapport within the community. 2) From the perspective of language educators, the corpus can be a source for sociolinguistic research and a good springboard for discussions on specific lessons. These posts can be implemented in classrooms to show the way language adapts to accommodate new communicative technologies and social contexts. The functional importance of linguistic hybridization and the pragmatic choices made to optimize communication in digital spaces, such as the use of abbreviations and fragmented sentences, can be understood by students through the analysis of these examples. 3) Lastly, this research establishes a foundation for future research that will further investigate the specific social and cultural factors that influence linguistic practices in other online communities in the Philippines specifically in local places such as Leyte.

5. References

- Al-Rawi, A. (2020). The linguistics of online news: Features, functions, and future. *Journal of Applied Linguistics and Language Research*, 7(6), 1-15.
- DataReportal. (2024, February 21). Digital 2024: The Philippines. Retrieved July 27, 2025, from <https://datareportal.com/reports/digital-2024-philippines>
- Darus, S., & Ching, C. F. (2020). Code-switching in Malaysian online forums: A multilingual perspective. *GEMA Online Journal of Language Studies*, 20(4), 160-176. <https://doi.org/10.17576/gema-2020-2004-10>
- Franzke, A., Bechmann, A., Ess, C., & the AoIR Ethics Working Committee. (2020). Internet research: Ethical guidelines 3.0. Association of Internet Researchers. <https://aoir.org/reports/ethics3.pdf>

- Garcia, J. A. (2021). Code-switching in online learning platforms: An analysis of Filipino English. *Asian EFL Journal*, 28(3), 147-165.
- Gumperz, J. J. (1982). *Discourse strategies* (No. 1). Cambridge university press.
- Hyland, K. (2021). *Disciplinary discourses: Social interactions in academic writing* (2nd ed.). University of Michigan Press.
- Kronrod, A., & Pogacar, R. (2023). The emergence and evolution of consumer language research. *Journal of Consumer Research*, 51(1), 42–56. <https://doi.org/10.1093/jcr/ucad003>
- Laghari, R. (2024, October 7). Building a corpus: How to build a corpus? Medium. <https://medium.com/@riazleghari/building-a-corpus-4bd02f69f5c5>
- London School of Economics and Political Science (LSE). (2025, April 1). Using data from the internet and social media in research: Ethics & consent. <https://info.lse.ac.uk/staff/divisions/research-and-innovation/research/Assets/Documents/PDF/ethics-Using-internet-and-Social-media-data.pdf>
- Mahdi, A. K. (2020). Code-switching as a communicative strategy in Iraqi social media discourse. *Journal of Language Teaching and Research*, 11(1), 160-168. <https://doi.org/10.17507/jltr.1101.20>
- McEnery, T., & Hardie, A. (2021). *Corpus linguistics: Method, theory and practice*. Cambridge University Press.
- Ong, K. C. (2022). Exploring multilingual practices in Philippine digital communication: A case study of Facebook interactions. *Journal of Language and Communication*, 9(1), 1-18.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (2020). *A comprehensive grammar of the English language*. Routledge.
- Rammel, D., & Herbig, M. (2022). Language in digital marketing communications: A review and research agenda. *Journal of Business Research*, 142, 582-595. <https://doi.org/10.1016/j.jbusres.2021.12.022>
- Smith, J., & Jones, A. (2021). Social media as a marketplace: An ethnographic study of online buy-and-sell groups. *Journal of Digital Commerce*, 8(2), 112-128.
- Tuan, N. B. (2023). Linguistic features of online commercial advertisements in Vietnam. *Asian Journal of English Language and Literature*, 10(1), 45-60.
- Villamor, J. P., Sumalpong, M. B., & Villamor, V. P. (2025). Exploring the role of code-switching in multilingual classroom dynamics: A comparative study of attitudes and practices among university students in the Philippines. *JPAIR Multidisciplinary Research*, 59(1), 22-40.
- Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. *Communication research*, 19(1), 52-90.
- Zipf, G. K. (2016). *Human behavior and the principle of least effort: An introduction to human ecology*. Ravenio books.