International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Lexical Features and Linguistic Environments of Twitter Slang

Louie Cristian Estrella Simbajon

Department of Languages and Literature, College of Arts and Sciences, Central Mindanao University, Philippines

Abstract

This study analyzed tweets from Filipinos containing slang words. Specifically, it categorized the recurrent kinds of slang words used in tweets, identified their lexical features, and determined other linguistic environments where these slang words appear. The concepts of internet slang and slang classification were used to identify the recurrent slang categories, while the Linguistic Features of Short Texts and Shortening Strategies were used to analyze lexical features. The Corpus of Contemporary American English (COCA) was also used to identify the linguistic environments of the selected slang words. Using qualitative content analysis, the study found nine recurrent kinds of slang in tweets, with letter homophones, keyboard-generated icons and smileys, and onomatopoeic spellings as the most prominent categories. The findings further showed the extensive use of shortening strategies and short-text linguistic features, including clipping and contraction, vowel omission, non-standard spelling, abbreviations, emoticons, and transmissions in characters. The identified slang words appeared in eight COCA genres, suggesting that Twitter slang shares features with other linguistic environments while also reflecting localized and platform-based lexical innovation.

Keywords: Lexical Features, Linguistic Environments, Slang, Tweets, Corpus Linguistics

1. Introduction

The digital age has intensified the role of information technology in communication and has expanded the spaces in which new words, altered spellings, and informal lexical items circulate. The growth of internet use has contributed to the rapid appearance and circulation of new lexical forms, while social media has enabled users to create, distribute, and normalize language innovations across communities (Glance, 2015; Kemp, 2019).

In the Philippine context, social media has become a productive site for linguistic creativity. Twitter, as a microblogging platform, has encouraged short, fast, and socially responsive communication. With millions of tweets being produced daily in the Philippines, the platform provides a rich environment for examining how users create and circulate slang (Statista Research Department, 2019). Even when tweets contain non-standard spellings, abbreviations, or informal expressions, the communicative priority of social media remains pragmatic clarity rather than strict conformity to formal written norms (Hoffmann & Bublitz, 2017).

Slang is significant because it performs both linguistic and social functions. It may index identity, humor, solidarity, playfulness, secrecy, and group membership. It may also reduce excessive seriousness in interaction and allow speakers or writers to display creativity (Moore, 2012; Partridge, 1970). In social

media discourse, these functions are intensified because users write for audiences that are immediate, interactive, and often community-specific.

The study of Twitter slang is therefore useful in understanding how language users modify form, meaning, and usage in digital discourse. Language has the capacity for change because its signs are productive and context-sensitive; through language, users do not merely record reality but also create social meanings (Ahmed, 2013; Halliday & Matthiessen, 2014). Word meaning also depends on linguistic environment because a lexical item can shift in use depending on its textual and social surroundings (Ahmed, 2013; Halliday & Matthiessen, 2014).

Previous studies have examined slang, word formation, sociolinguistic awareness, and language use in digital and social-media contexts (Gomaa, 2015; Grieve et al., 2018; Squires, 2010; Tagliamonte & Denis, 2008; Zappavigna, 2011). However, there remains a need to examine the lexical features of slang used in Philippine Twitter discourse and to identify the broader linguistic environments where these forms appear. This study addresses that need by analyzing Filipino tweets containing slang words and by examining their relationship to lexical innovation and corpus-based linguistic environments.

This study aimed to analyze tweets containing slang words. Specifically, it sought to categorize the recurrent kinds of slang words widely used in tweets, identify the lexical features of the slang words, and determine other linguistic environments where the slang words appear.

2. Related Research Work

2.1 Slang and Its Function

Slang is commonly understood as informal and context-bound language used by particular groups or communities. It emerges from social, historical, and cultural conditions and is often associated with novelty, playfulness, and creative expression. It may involve the renaming of objects, actions, feelings, or social experiences in ways that are meaningful to a particular speech community (Fasola, 2014; Moore, 2012).

In the Philippine setting, slang is used not only within one social group but also among students, young people, social media users, and other communities. It may be used for greetings, opinions, emotions, characterizations, questions, and other interactional purposes (Pascual, 2016). These functions show that slang is not simply an informal substitute for formal vocabulary. Instead, it operates as a social resource that allows users to position themselves within communities and communicative situations.

2.2 Internet Slang and Word-Formation Categories

Internet slang covers a wide range of lexical forms shaped by digital communication. Some forms come from short messaging practices, where words are replaced by phonetic equivalents, numbers, symbols, shortened forms, or expressive spellings. Common categories include letter homophones, punctuation and capitalization for emphasis, onomatopoeic spellings, keyboard-generated icons and smileys, leetspeak, flaming, clipping, compounding, derivation, and blends (Barseghyan, 2013).

Letter homophones include forms such as abbreviated spellings, acronyms, and initialisms. Punctuation and capitalization may be used to signal emphasis, emotion, or intensity. Onomatopoeic spellings, such as repeated laughter forms, represent sound-based writing conventions. Keyboard-generated icons and smileys represent affective or emotional meanings through characters, while clipping and compounding reduce or combine lexical items to meet communicative needs in digital spaces (Barseghyan, 2013).

2.3 Lexical Features and Linguistic Environment

Lexical features refer to observable forms through which words are shaped, shortened, modified, or repu-

rposed. In short-text communication, users often employ abbreviations, transmissions in characters, contractions, non-standard spelling, vowel omission, and other economy-driven strategies (Dąbrowska, 2011; Ling & Baron, 2007). These features are useful for analyzing how users maximize meaning within limited or fast-moving digital contexts.

Linguistic environment refers to the contexts in which words appear and gain meaning. Corpus-based analysis is useful for identifying these environments because it shows how words occur across genres and registers. The Corpus of Contemporary American English is a large, register-balanced corpus that provides information about word frequency and use across spoken, fiction, magazine, newspaper, academic, television and movie, blog, and general web genres (Davies, 2008–; Davies & Gardner, 2010; Yusu, 2014).

Table 1: COCA Database Used in Identifying Linguistic Environments

Genre	Number of Texts	Number of Words
Academic	26,137	120,988,361
Fiction	25,992	119,505,305
Magazine	86,292	127,352,030
Newspaper	90,243	122,958,016
Spoken	44,803	127,396,932
Television and Movies	23,975	129,293,467
Web (Blog)	98,748	125,496,216
Web (General)	88,989	129,899,427

The corpus information in Table 1 shows the range of genres used in this study for identifying linguistic environments. Because COCA is organized across multiple genres, it allows the study to observe whether slang forms found in Philippine tweets also appear in broader English-language contexts (Davies, 2008–).

3. Research Methods

3.1 Research Design

The study used a qualitative research design and employed qualitative content analysis. The approach was appropriate because the study did not only identify slang words in tweets but also interpreted their forms, lexical features, and linguistic environments. The study was also corpus-based because COCA was used to examine the contextual occurrence of selected slang forms across different genres.

3.2 Sampling Procedure and Sampling Period

The required sample size was computed using a sample size calculator. With a 5 percent confidence interval, a 95 percent confidence level, and an estimated population of 10.2 million tweets per day in the Philippines, the computed sample size was 384 tweets (Statista Research Department, 2019). The actual selection of tweets was purposive because the study focused on public tweets that contained slang words relevant to the research objectives.

The sampling period covered 16 consecutive nights. Twenty-four tweets were gathered each night between 6:00 p.m. and 10:00 p.m., which were treated as active evening hours for social media use. This procedure produced the required corpus of 384 tweets. The fixed nightly quota helped distribute the collection across the sampling period rather than concentrating the data on a single date or short time block.

3.3 Keyword/Search Procedure and Inclusion/Exclusion Criteria

The search procedure was category-led and objective-based. The researcher used the Twitter mobile application to observe and search for public Filipino tweets containing recognizable slang forms that could be examined under the study's categories, such as letter homophones, onomatopoeic spellings, keyboard-generated icons and smileys, punctuation and capitalization, clipping, compounding, and other slang-related forms. The search and selection process prioritized textual items that showed clear word modification, informal lexical use, or platform-based spelling practices.

Tweets were included when they met all of the following criteria: they were publicly available at the time of collection; they contained at least one identifiable slang word, shortened form, modified spelling, character-based expression, or other relevant slang feature; they contained enough textual context for classification; and they could be documented through a screenshot showing the tweet text, username, date, and time of posting. Tweets were excluded when they did not contain analyzable slang, were duplicates or repeated captures of the same item, were private, deleted, or inaccessible at the time of review, consisted only of images or non-textual material, or lacked sufficient context for coding.

3.4 Data Gathering and Analysis Procedure

Tweets were documented through screenshots using the Twitter mobile application. Each screenshot preserved the tweet text and basic contextual information, including the username, date, and time of posting. For publication reporting, the manuscript presents aggregated findings and does not reproduce usernames or account-level identifiers.

After gathering the corpus, slang words were categorized using frameworks on slang and internet slang (Barseghyan, 2013; Moore, 2012; Partridge, 1970; Pascual, 2016). Frequency counts were used to identify recurrent kinds of slang. The lexical features of the recurrent slang words were then analyzed using the Linguistic Features of Short Texts and Shortening Strategies (Dąbrowska, 2011; Ling & Baron, 2007). Finally, the selected slang words were entered into COCA to identify the linguistic environments in which they appeared (Davies, 2008–). The analysis focused on genre distribution and contextual occurrence rather than on individual user identity.

3.5 Ethical Considerations

Only publicly available tweets were analyzed. The presentation of findings avoided usernames, handles, and account-level identifiers to keep the discussion focused on discourse features rather than individual users. The study therefore reports patterns, frequencies, categories, and linguistic environments without exposing individual Twitter users.

4. Results and Discussion

4.1 Recurrent Kinds of Slang in Tweets

The analysis identified 783 total occurrences of slang in 384 tweets. Nine recurrent kinds of slang were found: cant and swear words, clipping, compounding, keyboard-generated icons and smileys, letter homophones, medical slang, onomatopoeic spellings, punctuation, capitalization and other symbols, and society slang.

Letter homophones were the most recurrent kind of slang, accounting for 51.60 percent of the total occurrence. Keyboard-generated icons and smileys followed with 17.11 percent, while onomatopoeic spellings accounted for 12.00 percent. These findings show that Twitter users frequently relied on abbreviated, character-based, and sound-based forms to communicate meanings quickly and creatively (Table 2).

Table 2. Dominant recurrent kinds of slang in the dataset

Recurrent kind of slang	Frequency	Percentage	Interpretive note
Letter homophones	404	51.60%	Words are compressed through sound-letter substitution while remaining recognizable.
Keyboard-generated icons and smileys	134	17.11%	Character-based forms carry affect, stance, or emphasis.
Onomatopoeic spellings	94	12.00%	Sound-based spellings reproduce expressive or speech-like effects.
Other recurrent kinds combined	151	19.29%	Includes cant/swear words, clipping, compounding, medical slang, symbols, and society slang.
Total	783	100.00%	Total recorded slang occurrences in the sampled tweets.

The prominence of letter homophones suggests that users often preferred forms that compress words while preserving recognizability. This pattern supports the idea that internet language is commonly associated with acronyms, abbreviations, respellings, and other non-standard written features (Squires, 2010). The result also aligns with findings that digital and social-media writing uses abbreviations, emotion markers, and audience-oriented forms (Tagliamonte & Denis, 2008; Zappavigna, 2011).

4.2 Lexical Features of Twitter Slang

The study found that shortening strategies were extensively used in the sampled tweets. Among these strategies, clipping and contraction were the most frequent at 32.19 percent, followed by vowel omission at 26.03 percent and non-standard spelling at 20.55 percent. These features show that Twitter slang frequently modifies word form through reduction and spelling variation.

The analysis also identified linguistic features of short texts. Abbreviations accounted for 51.47 percent, emoticons for 18.63 percent, and transmissions in characters for 8.58 percent. These features indicate that Twitter users relied on compact linguistic signals to express content, stance, and emotion (Table 3). Such usage reflects the broader function of internet slang as a tool for rapid and socially marked communication (Barseghyan, 2013; Dąbrowska, 2011; Ling & Baron, 2007).

Table 3. Lexical features and communicative functions of Twitter slang

Lexical feature	Recorded share	Communicative function
Clipping and contraction	32.19%	Reduces longer expressions while preserving recoverable meaning.
Vowel omission	26.03%	Compresses spelling and supports faster typed interaction.
Non-standard spelling	20.55%	Indexes informality, pronunciation, humor, or group style.
Abbreviation	51.47%	Condenses repeated expressions into compact written forms.
Emoticons	18.63%	Adds affective and interpersonal meaning to short posts.

Lexical feature	Recorded share	Communicative function
Transmissions in characters	8.58%	Encodes emphasis or stance through symbols and character sequences.

The findings suggest that lexical innovation in Twitter discourse is not random. Instead, it follows recognizable patterns of economy, sound representation, affective signaling, and visual emphasis. These patterns are consistent with broader evidence that social media supports lexical innovation in English-language digital discourse (Grieve et al., 2018).

4.3 Linguistic Environments of Twitter Slang

Using COCA, the study found that the identified slang words appeared across eight linguistic environments: web blog, web general, magazine, television and movie, fiction, spoken, newspaper, and academic genres. The recorded appearances were web blog with 147, web general with 136, magazine with 134, television and movie with 133, fiction with 132, spoken with 129, newspaper with 125, and academic with 122 (Table 4).

Table 4. COCA linguistic environments of selected Twitter slang forms

COCA environment	Recorded appearance
Web blog	147
Web general	136
Magazine	134
Television and movie	133
Fiction	132
Spoken	129
Newspaper	125
Academic	122

The distribution suggests that Twitter slang forms are not confined to social media. Some forms also appear in broader linguistic environments, particularly web-based and media-oriented genres. This finding indicates that slang can circulate across multiple contexts and may become recognizable beyond the platform where it is actively used.

At the same time, the academic genre showed the lowest recorded appearance among the eight genres. This result is expected because academic writing follows more formal conventions and tends to limit the use of informal or platform-specific expressions. Keyboard-generated icons and smileys also showed zero appearance in COCA because they are character-based rather than word-based forms, and COCA primarily stores and retrieves lexical items.

Some Filipino expressions and recently modified English slang forms did not appear in COCA. This absence suggests that local slang forms may remain limited to particular communities, platforms, or national contexts. However, the presence of other slang forms in COCA also shows that some Twitter slang shares features with wider English-language usage. The finding supports the view that youth and popular language can influence media, print, and other literacy practices (Djenar, 2015).

5. Conclusion

The study concludes that Twitter is a productive environment for slang and lexical innovation. The sampled tweets showed extensive use of slang, particularly letter homophones, keyboard-generated icons and smileys, and onomatopoeic spellings. These forms demonstrate how users reshape words and symbols to make communication shorter, expressive, and socially recognizable.

The lexical features identified in the study show that Twitter slang is shaped by patterns of shortening, spelling modification, abbreviation, emotive marking, and character transmission. These features suggest that informal digital writing is systematic rather than careless. Users modify words according to communicative needs, platform practices, and shared community knowledge.

The study further concludes that slang found in Philippine tweets can appear in other linguistic environments, including web, magazine, television and movie, fiction, spoken, newspaper, and academic genres. However, the uneven distribution across genres also shows that acceptance and use vary according to context. Overall, the findings affirm that slang contributes to language adaptation, digital creativity, and the changing relationship between informal discourse and broader linguistic environments.

Because the corpus was limited to selected public tweets from Filipino Twitter discourse, the findings should not be generalized to all Filipino Twitter users. Future studies may compare larger datasets across platforms, age groups, regions, and Philippine languages to examine whether similar lexical patterns appear in other digital communities.

7. Acknowledgement

The author acknowledges Ms. Jasmin T. Balabagno for her guidance, patience, suggestions, and support during the development of the study. The author also expresses gratitude to Ms. Danna Karyl Jane C. Talde and Dr. Jeneifer C. Nueva for their comments and recommendations. Appreciation is likewise extended to the Department of Languages and Literature of Central Mindanao University, the author's family, classmates, friends, and all individuals who offered support throughout the completion of the research.

8. References

1. Ahmed, A. R. I. (2013). Lexis development: The importance of contextual clues. European Scientific Journal, 9(35), 339–348. https://eujournal.org/index.php/esj/article/view/2233
2. Barseghyan, L. (2013). On some aspects of internet slang. Graduate School of Foreign Languages, 14, 19–31. https://artsakhlib.am/wp-content/uploads/2019/12/Liana-Barseghyan-On-Some-Aspects-of-Internet-Slang.pdf
3. Dąbrowska, M. (2011). Language economy in short text messages. ResearchGate. https://www.researchgate.net/publication/297735378_Language_economy_in_short_text_messages
4. Davies, M. (2008–). The Corpus of Contemporary American English (COCA). English-Corpora.org. Retrieved May 9, 2026, from https://www.english-corpora.org/coca/
5. Davies, M., & Gardner, D. (2010). A frequency dictionary of contemporary American English: Word sketches, collocates, and thematic lists. Routledge. https://www.routledge.com/A-Frequency-Dictionary-of-Contemporary-American-English-Word-Sketches-Collocates-and-Thematic-Lists/Davies-Gardner/p/book/9780415490641

6. Djenar, D. N. (2015). Youth language in Indonesia and Malaysia: From slang to literacy practices. NUSA: Linguistic Studies of Languages in and Around Indonesia, 58, 1–8. http://hdl.handle.net/10108/84125
7. Fasola, I. (2014). Slang and its history. In Proceedings of the 53rd International Scientific Conference of Daugavpils University. Daugavpils University. https://dukonference.lv/files/proceedings_of_conf/53konf/valodnieciba_literaturzinatne/Fasola.pdf
8. Glance, D. (2015). How technology is changing language and the way we think about the world. The Conversation. https://theconversation.com/how-technology-is-changing-language-and-the-way-we-think-about-the-world-35856
9. Gomaa, Y. A. (2015). Saudi youth slang innovations: A sociolinguistic approach. International Journal of Linguistics and Communication, 3(2), 98–112. https://doi.org/10.15640/ijlc.v3n2a11
10. Grieve, J., Nini, A., & Guo, D. (2018). Mapping lexical innovation on American social media. Journal of English Linguistics, 46(4), 293–319. https://doi.org/10.1177/0075424218793191
11. Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). Halliday's introduction to functional grammar (4th ed.). Routledge. https://www.routledge.com/Hallidays-Introduction-to-Functional-Grammar/Halliday-Matthiessen/p/book/9780415826280
12. Hoffmann, C. R., & Bublitz, W. (Eds.). (2017). Pragmatics of social media. De Gruyter Mouton. https://doi.org/10.1515/9783110431070
13. Kemp, S. (2019, January 31). Digital 2019: Global digital overview. DataReportal. https://datareportal.com/reports/digital-2019-global-digital-overview
14. Ling, R., & Baron, N. S. (2007). Text messaging and IM: Linguistic comparison of American college data. Journal of Language and Social Psychology, 26(3), 291–298. https://doi.org/10.1177/0261927X06303480
15. Moore, R. L. (2012). On swearwords and slang. American Speech, 87(2), 170–189. https://doi.org/10.1215/00031283-1668191
16. Partridge, E. (1970). Slang: To-day and yesterday (4th ed.). Routledge & Kegan Paul. https://openlibrary.org/books/OL5474174M/Slang_to-day_and_yesterday
17. Pascual, G. F. S. (2016). Swardspeak or gay lingo in the Philippine context: A morphological analysis. International Journal of Advanced Research in Management and Social Sciences, 5(12), 32–36. https://garph.co.uk/IJARMSS/Dec2016/4.pdf
18. Squires, L. (2010). Enregistering internet language. Language in Society, 39(4), 457–492. https://doi.org/10.1017/S0047404510000412
19. Statista Research Department. (2019). Twitter: Number of monthly active users 2010–2019. Statista. Retrieved May 9, 2026, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
20. Tagliamonte, S. A., & Denis, D. (2008). Linguistic ruin? LOL! Instant messaging and teen language. American Speech, 83(1), 3–34. https://doi.org/10.1215/00031283-2008-001
21. Yusu, X. (2014). On the application of Corpus of Contemporary American English in vocabulary instruction. International Education Studies, 7(8), 68–73. https://doi.org/10.5539/ies.v7n8p68
22. Zappavigna, M. (2011). Ambient affiliation: A linguistic perspective on Twitter. New Media & Society, 13(5), 788–806. https://doi.org/10.1177/1461444810385097