

Identifying Code Mixing on Instagram: The Case of Chinese-Indonesians

Rafi Muhammad Lubis¹, Hafwen Anargya², Fabian Putra Nazreensyah³, Daniel Fernando Siahaan⁴, Rahmadsyah Rangkuti⁵

Universitas Sumatera Utara

rangkuti@usu.ac.id

ABSTRACT

This study looks into code mixing in Chinese-Indonesian users' Instagram descriptions and comments. The research employs a qualitative descriptive approach to identify and categorize the types of code mixing found in online discourse, using Muysken's (2000) typology of code mixing as the analytical framework. Data were carefully gathered from 100 Chinese-Indonesian Instagram caption pages and 40 Instagram posts consist of public Instagram accounts of Chinese-Indonesian people, with an emphasis on comments and captions that included multilingual content. Three forms of code mixing are identified by the analysis: congruent lexicalization, alternation, and insertion. The most common type was insertion, with 58 instances, followed by alternation with 30 instances and congruent lexicalization with 12. The results demonstrate the ongoing role of social media in maintaining linguistic diversity by indicating that Instagram functions as a platform for the active performance and maintenance of multilingual practices.

Keywords: Code mixing, Chinese-Indonesians, Instagram, Multilingualism

Received:
12-09-2025

Revised:
20-10-2025

Accepted:
29-10-2025

INTRODUCTION

Social media has been a potent medium for multilingual users to express their identities and communicate in recent years. Local languages coexist with Bahasa Indonesia and other languages like English, Mandarin, and Hokkien in Indonesia, where multilingualism is both a social reality and a cultural standard (Nugraha et al., 2024). Particularly among minority groups like the Chinese-Indonesian community, platforms like Instagram have developed into linguistic spaces where various languages interact dynamically. Their usage of many languages on the internet reveals intricate social, cultural, and identity debates in addition to communication demands.

The role of code-mixing in online written discourse, especially among Chinese-Indonesian Instagram users, has not been well studied despite Indonesia's rich multilingual context. Instagram, as a platform that has been integrated into daily life for over a decade and able to facilitates mutual connections among users, serves as a significant medium for this analysis. The majority of research

conducted in Indonesia on code-mixing has been on spoken situations (Safitri, 2021; Astri & Fian, 2020) or students' overall opinions of multilingualism (Nugraha et al., 2024; Abdulloh & Usman, 2021). Nonetheless, recent studies have also highlighted the increasing importance of code-switching and code-mixing as linguistic and identity performance strategies in popular culture and social media (Nazri & Kassim, 2023; Wulandari & SusyLOWATI, 2021; Sari et al., 2020; Nurchaerani et al., 2022). As a result, little is known about the structures and patterns of code-mixing in digital environments, especially among minority language populations. There is a knowledge deficit about how internet communication helps minority speakers preserve their linguistic identity and variety as a result of this neglect.

In order to solve this problem, the current study uses Pieter Muysken's (2000) typology of code-mixing to characterize and categorize the structural patterns of code-mixing employed by Chinese-Indonesian Instagram users. In order to differentiate between various mixing techniques in bilingual or multilingual interactions, Muysken outlines three primary structural processes of code-mixing: insertion, alternation, and congruent lexicalization. Congruent lexicalization occurs when two languages share a grammatical structure and lexical items can appear interchangeably; alternation occurs when speakers switch between two languages throughout clauses or sentences; and insertion entails embedding elements from one language into the grammatical structure of another. The way Chinese-Indonesian users mix Mandarin, Hokkien, English, and Indonesian in their Instagram posts and comments may be systematically categorized by using this framework (Sari, Arifin, & Harida, 2020).

The discipline of sociolinguistics, which studies the interaction between language and society, has a strong foundation in the idea of code-mixing. Code-mixing, according to Harya (2018), is a normal occurrence among bilingual and multilingual speakers who incorporate linguistic aspects from many languages into a single discussion. Within a particular group, code-mixing is a means of expressing closeness, establishing identification, and conveying meaning. Code-mixing, according to Abdulloh and Usman (2021), is a sign of linguistic innovation and adaptability, particularly for younger generations navigating numerous language systems in social and academic situations. Similarly, research by Slamet et al. (2020), Wulandari and SusyLOWATI (2021), and Nurchaerani et al. (2022) shows that sociolinguistic elements including social identity, solidarity, and intergenerational communication styles affect code-mixing.

The sociolinguistic and identity-related purposes of code-mixing have been highlighted in a number of studies. According to Akhtar et al. (2020), code-switching and code-mixing are tools for negotiating one's identity and sense of cultural belonging, especially in heterogeneous communities. Similarly, Guo (2023) emphasizes that code-mixing among Chinese college students in China is a reflection of language adaptability in transnational contexts, identity negotiation, and modernity. According to Nordin (2023), social media users actively utilize code-mixing and code-switching in online settings in order to express themselves, establish connections with a variety of audiences, and preserve linguistic hybridity. Furthermore, Meliani et al. (2021) show that insertion-type code-mixing is prevalent on social media sites like Twitter, where users purposefully mix languages to convey modernity and bilingualism. Similar to this, code-switching in popular culture and community contexts serves as a tool for identity negotiation, language adaptation, and cultural preservation, according to Nazri and Kassim (2023), Airemionkhale (2023), and Nurchaerani et al. (2022). According to Muysken's (2000) theory, these investigations support the idea that code-mixing is systematically influenced by environment, cultural background, and language proficiency rather than occurring at random.

Understanding social media code-mixing from a sociolinguistic standpoint sheds light on how linguistic variety is preserved in the digital era. According to earlier studies on YouTube comments by Safitri (2021) and Astri and Fian (2020), multilingual users use several languages to highlight concepts, exhibit comedy, and express who they are. However, the structural aspects of code-mixing were not the main focus of these investigations, which instead examined its communication function. When it comes to explaining how mixed utterances are created online, this creates a theoretical and methodological vacuum. This study helps close that gap by concentrating on the Chinese-Indonesian community, a minority population with a sophisticated bilingual repertory.

METHOD

This study uses a qualitative descriptive research design in the field of sociolinguistics. The aim of the study is to find and categorize the different forms of code mixing in Instagram conversation created by Chinese-Indonesian users. The main source of data for this paper is the Chinese-Indonesian users' Instagram comments and captions. By using Instagram as the data source, this study is able to document real-world examples of code mixing in a digital space where people are free and easy to express themselves.

Using a purposeful sampling technique, the focus is on Instagram accounts, both individual users and community pages, that represent Chinese-Indonesian identity. The statistics were restricted to posts made in the year 2025 in order to preserve consistency. The dataset only included comments and posts that showed signs of multilingualism, which in this study refers to the use of two or more languages within a single post and comment, accompanied by an appropriate understanding and contextual use of each language. Screenshots were used to collect the data, which were then transcribed for analysis. Usernames and other identifying information were anonymised before processing in order to protect privacy. Then, Muysken's typology of code mixing was used to evaluate the data. Following classification, relevant instances were chosen and explained to show how code mixing appears in Chinese-Indonesian Instagram users' conversations.

We collected 100 Instagram captions and comments from 40 posts made by Chinese-Indonesian users. These posts were specifically chosen from public accounts, such as individual users and community pages, that explicitly represented Chinese-Indonesian identity. The posts were selected because they covered a variety of subjects, such as everyday expressions, informal online interactions, and cultural celebrations (such as the Lunar New Year and Mid-Autumn Festival), after that the data that already been collected are arranged in analysis table then using Muysken's (2000) typology, the comments and captions were analyzed to determine the types and frequency of code mixing, Code mixing is categorized as either congruent lexicalization, alternation, or insertion, and the last step, after the data has been analyzed, is selecting the data that best represents each type found among all the data gathered for the article.

FINDINGS

There is unmistakable proof of code mixing in the descriptions and comments posted by Chinese-Indonesian Instagram users. Insertion, alternation, and congruent lexicalization were the three

categories that were determined using Muysken’s (2000) typology. Every kind demonstrates a distinct method by which speakers blend languages in their written messages on the internet.

TABLE 1. Distribution of Code Mixing Types in the Dataset (N = 100)

Types of Code Mixing	Frequency	Percentage	Description
Insertion	58	58%	Incorporating single words or brief phrases—typically from English or Hokkien—into sentences written in Indonesian.
Alternation	30	30%	Switching at the end of a sentence or clause between two or more languages.
Congruent Lexicalization	12	12%	Combining vocabulary from several languages while maintaining a common grammatical structure.
Total	100	100%	-

Insertion was the most common form of code mixing in the dataset, according to an analysis of all the data collected, as about 58 instances of insertion are observed from the data. This pattern involved inserting words, phrases, or speech particles from English, Mandarin, or Hokkien into sentences that were primarily in Indonesian. The caption ‘Happy Mid Autumn Festival semua, jangan lupa makan mooncake ah!’, for example, shows how to insert a Hokkien language marker and an English greeting into an Indonesian sentence. These insertions bring out informality, add cultural nuance, and enhance the words themselves. The ease with which speakers incorporate elements of other languages into Indonesian structures is demonstrated by the frequent usage of insertion.

The second type that appears is alternation, where in the data, about 30 instances are observed, which happens when users switch between two or more languages at the end of a sentence or clause. “Besok kita makan bareng ya. Wa ai chia pui!” is one example. The sentence has a Hokkien second clause and an Indonesian first clause. This switch implies intentional language use to emphasize cultural expressions, show humor, or indicate intimacy with in-group members. Although less frequent than insertion, alternation shows how Chinese-Indonesian users can use multiple languages with ease.

Although more rare, congruent lexicalization is still important, as about 12 instances are observed in the data. This happens when English or Hokkien lexical items are freely put into a frame provided by Indonesian grammar. ‘Kemarin aku lagi chill di rumah sambil dengerin lagu Hokkien’, for instance, shows how English, Hokkien, and Indonesian can all be seamlessly

blended within a common grammatical structure. This demonstrates a high degree of bilingual or multilingual competence by demonstrating that users view various linguistic resources as compatible.

According to the results, congruent lexicalization occurs the least frequently in the dataset, while insertion and alternation are the most common forms of code mixing. The widespread use of insertion supports Muysken's (2000) theory that the most popular tactic in multilingual contexts is insertion. This pattern shows how simple and organic it is to incorporate individual lexical items into an Indonesian sentence structure.

The simplicity and naturalness of insertion in informal online communication are the reasons for its high frequency in the dataset. Without changing Indonesian's primary grammatical structure, insertion enables users to incorporate individual words or language particles from another language. This makes it a useful tactic for conveying emphasis, emotion, or cultural nuances in quick social media exchanges. Alternation is less common because it requires more grammatical control and is frequently used on purpose to emphasize contrast, humor, or group identity rather than as the default mode of communication. Meanwhile, congruent lexicalization is less common since it requires high levels of bilingualism and structural compatibility between languages, requiring a more intricate integration of two linguistic systems within common grammatical frames.

TABLE 2. Sample Data

No	Data	Language(s)	Types of Code Mixing	Explanation
1.	Happy Mid Autumn Festival semua, jangan lupa makan mooncake ah!	Indonesian + English + Hokkien	Insertion	English greetings and the Hokkien particle 'ah' are used to add cultural nuance to an Indonesian sentence.
2.	Besok kita makan bareng ya. Wa ai chia pui!	Indonesian + Hokkien	Alternation	Display group identity by switching between Hokkien in the second clause and Indonesian in the first.
3.	Kemarin aku lagi chill di rumah sambil dengerin lagu Hokkien.	Indonesian + English + Hokkien	Congruent Lexicalization	The English word "chill" and the Hokkien word "lagu Hokkien" are both supported by the Indonesian grammatical frame.

4.	Jangan lupa ikut meeting nanti sore lah!	Indonesian + Hokkien	Insertion	'Lah' is a Hokkien discourse particle that is used for informality and emphasis.
5.	Hari ini capek banget. Tomorrow I will just stay at home.	Indonesian + English	Alternation	The sentence switches between Indonesian and English, emphasizing contrast and emphasis.
6.	Lagi craving dim sum, tapi too lazy keluar rumah.	Indonesian + English	Insertion	The English adjective "too lazy" is used to convey a modern tone and casualness in an Indonesian sentence.
7.	Wa boh jio you tadi, tapi semua sibuk kerja.	Hokkien + Indonesian	Alternation	The user demonstrates ease of switching between codes within a single utterance by alternating between Hokkien ("Wa boh jio you" – I didn't invite you) and Indonesian.
8.	Jangan lupa pakai cheongsam waktu Imlek ya girls!	Indonesian + Mandarin + English	Congruent Lexicalization	The English plural "girls" and the Mandarin loan "cheongsam" are integrated into Indonesian grammar, demonstrating fluid multilingual usage.
9.	Makan bareng keluarga besar, so happy lah!	Indonesian + English + Hokkien	Insertion	English phrase "so happy" and Hokkien particle "lah" are inserted into an Indonesian sentence to convey warmth and excitement.

10.	Aduh cuaca panas banget, I cannot tahan!	Indonesian + English + Hokkien	Alternation	Alternation occurs between Indonesian and English-Hokkien phrases.
-----	--	--------------------------------	-------------	--

DISCUSSION

The extension of multilingual practices into written online contexts is demonstrated by the existence of code mixing among Chinese-Indonesian users on Instagram. First, it shows that multilingualism in daily life is not limited to verbal exchanges but rather occurs in online settings where casual conversations occur. Second, even when the conversation is primarily in Indonesian, speakers maintain cultural identity markers by frequently using Hokkien words and particles like "ah" and "lah." Third, the language switch emphasizes how code mixing can serve as a sign of unity and in-group affiliation, especially in online communities where a common cultural background is important. Lastly, the regular usage of traditional languages in online conversation points to the potential of social media for language preservation, enabling younger generations to continue speaking Mandarin and Hokkien in addition to English and Indonesian.

The study identified three categories of code mixing in Chinese-Indonesian Instagram discourse: congruent lexicalization, alternation, and insertion. The most common method was insertion, which suggests that it naturally integrates linguistic components into Indonesian structures. Although less frequent, alternation demonstrated the use of several languages to highlight significance, promote unity, and preserve cultural ties. And then, despite being the least common, congruent lexicalization showed how flexible speakers are at using several linguistic resources at once. All things considered, the results demonstrate that Instagram is a significant platform for multilingual expression and helps maintain the vitality and visibility of minority languages in Indonesian society.

CONCLUSION

This study used Muysken's (2000) typology as the analytical framework for analyzing code mixing in Instagram comments and descriptions written by Chinese-Indonesian users. All three forms of code mixing insertion, alternation, and congruent lexicalization were found to exist, with insertion being the most common type. The findings show that Instagram serves as a significant forum for the active expression of multilingual practices in casual written discourse as well as their maintenance.

The way Chinese-Indonesians incorporate various linguistic resources into daily communication is demonstrated by the usage of Hokkien particles like ah and lah in addition to English and Mandarin insertions. Even though they are less common, alternation and congruent lexicalization emphasize how flexible language choice is in this community. These results highlight the ongoing vitality of minority languages like Hokkien, which are still actively used in digital contexts despite having a small formal presence. Because the dataset was limited to a comparatively small number of Instagram posts and comments from Chinese-Indonesian users, the scope of this study is limited. Additionally, the focus was descriptive, identifying different kinds of code mixing without delving deeper into pragmatic functions or presenting a

comprehensive quantitative analysis. Furthermore, the data came exclusively from Instagram, excluding other social media sites where distinct linguistic patterns might be found. Future studies could overcome these constraints by using mixed-methods approaches that integrate quantitative frequency analysis and qualitative description, as well as by enlarging the dataset to encompass more users and platforms. The relationship between code mixing and identity, language attitudes, or intergenerational language transmission within the Chinese-Indonesian community may also be the subject of future research.

The study concludes by demonstrating the important role social media sites like Instagram play in maintaining and promoting multilingualism in Indonesia. The Chinese-Indonesian community views code mixing as a reflection of their diverse linguistic repertoire as well as a communication tactic. As a result, this study advances our knowledge of code mixing in online settings and emphasizes how important digital discourse is to maintaining linguistic diversity.

REFERENCES

- Abdulloh, A., & Usman, D. R. (2021). Students' Perception towards Code-Switching and Code-Mixing in Sociolinguistic: A Case at an English Education Major Awab. *EDULINK Journal*, 3(1), 6.
- Akhtar, S., Baig, F. Z., Aslam, M. Z., Khan, T., Tayyaba, S., & Iqbal, Z. (2020). Code-Switching and Identity: A Sociolinguistic Study of Hanif's Novel Our Lady of Alice Bhatti. *International Journal of English Linguistics*, 10(1), 364. <https://doi.org/10.5539/ijel.v10n1p364>
- Astri, Z., & Fian, A. (2020). The Sociolinguistics Study on the Use of Code Mixing In Gita Savitri Devi's Youtube Channel Video. *Seltics*, 3(2), 83–92. <https://doi.org/10.46918/seltics.v0i0.728>
- Guo, Y. (2023). The Code-Mixing of Chinese and English Among Chinese College Students: A Qualitative Study. In *Proceedings of the 2022 4th International Conference on Literature, Art and Human Development (ICLAHD 2022)*. Atlantis Press SARL. <https://doi.org/10.2991/978-2-494069-97-8>
- Harya, T. D. (2018). Sociolinguistic (Code: Code Switching and Code Mixing). *LENTERA: Jurnal Ilmiah Kependidikan*, 11(1), 87–98. <http://jurnal.stkipgribl.ac.id/index.php/lentera>
- Meliani, P. E., Ratminingsih, N. M., & Mahendrayana, G. (2021). CODE MIXING AS LANGUAGE STYLE IN COMMUNICATION THROUGH SOCIAL MEDIA TWITTER. *Yavana Bhasha: Journal of English Language Education*, 4(2), 91–101.
- Muysken, P. (2000). Bilingual Speech A Typology of Code-Mixing. In *Cambridge University Press*. <https://doi.org/10.7551/mitpress/15113.003.0009>

- Nazri, S. N. A., & Kassim, A. (2023). Issues and Functions of Code-switching in Studies on Popular Culture: A Systematic Literature Review. *International Journal of Language Education and Applied Linguistics*, 13(2), 7–18. <https://doi.org/10.15282/ijleal.v13i2.9585>
- Nordin, N. R. M. (2023). Code-switching and code-mixing among users of social media. *Jurnal Javanologi*, 6(2), 1267. <https://doi.org/10.20961/javanologi.v6i2.75114>
- Nugraha, I., Jaelani, A., & Sukma, B. P. (2024). How Indonesian Students View Code-Mixing in Daily Conversations. *Journal of English Teaching and Linguistics Studies (JET Li)*, 6(2), 88–94. <https://doi.org/10.55215/jetli.v6i2.10407>
- Nurchaerani, M., Evto, R. H., Angelica, V., & Agriska. (2022). CODE-MIXING WITHIN THE USE OF KHEK LANGUAGE ON CHINESE-INDONESIAN ETHNIC IN TAMBORA. *English Journal Literacy UTama*, 6(2), 544–550.
- Omoningho, A. E. (2024). An Autoethnographic Exploration of Code-switching as a Tool for Identity Negotiation: Reflections of a Nigerian Postgraduate Student. In *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4942932>
- Safitri, M. (2022). A Sociolinguistics Analysis of Code-Swicing and Code-Mixing Used by Netizen Comments in Puella Podcast. In *UMS Library*. [https://eprints.ums.ac.id/96760/9/Naskah Publikasi %28Mirah Safitri A320170107%29_Wisuda.pdf](https://eprints.ums.ac.id/96760/9/Naskah_Publikasi_Mirah_Safitri_A320170107%29_Wisuda.pdf)
- Sari, M., Arifin, A., & Harida, R. (2021). CODE-SWITCHING AND CODE-MIXING USED BY GUEST STAR IN HOTMAN PARIS SHOW Maya. *Journal of English Language Learning (JELL)*, 5(2), 105–112.
- Slamet, J., Ninik, & Sulistyaningsih. (2020). A Sociolinguistics of Code Mixing and Code Switching Between English and Indonesian Language at STKIP PGRI Sidoarjo. *JournEEL (Journal of English Education and Literature)*, 2(1), 65–77. <https://doi.org/10.51836/journeel.v2i1.84>