International Summit on Science Technology and Humanity
ISETH 2024
ISSN: 2615-1588 (online)

Analysis of Code-Mixing Found in YouTube Content

Rifda Ghaitsa Tsuraya[1], Sigit Haryanto[2,*]

[1] Faculty of Teacher Training and Education, Universitas Muhammadiyah Surakarta, Surakarta, Indonesia

[*] Corresponding author: sh288@ums.ac.id

Abstract

Purpose: This study examines code-mixing in YouTube content created by Indonesian YouTubers Jesselyn Lauwreen, Gita Savitri Devi, and Sandy Kristian Waluyo, focusing on the types and motivations behind their language use.
Methodology: Adopting a qualitative descriptive approach, the research uses content analysis to collect and categorize instances of code-mixing. Data collection began with selecting and transcribing relevant videos to capture spoken content, which was then organized into tables for analysis. Code-mixing was categorized into three types: insertion, alternation, and congruent lexicalization.
Results: The reasons for code-mixing were analyzed, including discussing specific topics, quoting, showing emphasis, and expressing group identity.
Applications/Originality/Value: The findings reveal that insertion is the most frequent type of code-mixing across all three YouTubers, followed by alternation and congruent lexicalization. The primary reason for code-mixing was talking about a particular topic, with expressing group identity also being significant.

Introduction Section

People often modify their speech according to the person they are speaking to, considering things like intonation, language proficiency, cultural differences, and other factors. They use codes in the process, which helps them understand to communicate their ideas. When speakers communicate with people whose languages are different from their own, code switching and code mixing frequently takes occurred. Code-mixing refers to the incorporation of one language into the framework of another language, occurring in specific contexts (Lubis & Hendrawan, 2022). Code-switching is a linguistic phenomenon that frequently takes place within bilingual and multilingual speech communities (Nurhayati, 2021). These phenomena are common among bilingual or multilingual individuals. In Indonesia, regional languages serve as the mother tongue, while Indonesian is the national language, making many Indonesians bilingual. Additionally, basic English is widely understood, as it is taught from elementary school through university, and the global role of English as an international language is further reinforced by the widespread availability of information in English, especially in the digital age. The phenomenon of mixing Indonesian and English in one sentence has also become a trend among young people. As found on YouTube content,

"I was walking around terus ada stall yang ini aku lihat menarik ya udah cobain untuk pertama kalinya jadi aku iseng-iseng coba dan ternyata enak banget." Jesselyn Lauwreen YouTube channel (June 22nd, 2024)

"Buat kalian-kalian yang udah lama nge-follow gua you know that fact about me."Gita Savitri Devi YouTube channel (July 27th, 2024)

"Jadi sekarang jam lima lebih empat belas ya, tanggal sebelas Agustus dan hari ini aku ada flight ke Singapore jam delapan tiga lima pagi, habis itu bakal ke NUS, buat moving in ke dorm" Sandy Kristian Waluyo (September 19th, 2024)

The English terms used by Jesselyn and Sandy, such as 'stall,' 'flight,' and 'dorm' are placed in sentences that are mostly in Indonesian, indicating insertion code mixing. This may be because the English terms are more specific or familiar. However, Gita Savitri switch from Indonesian to English with "you know that fact about me." This shows alternational code mixing, where the speaker alternates between languages. This type of mixing often happens when a speaker is comfortable with both languages and switches for emphasis or ease of communication.

Given the varying instances and types of code-mixing displayed across different YouTube channels, several studies on code-switching and code-mixing have yielded diverse findings, with researcher distinguishing studies based on their data sources, some of which were drawn from YouTube channel videos. Astri and Fian's (2020) study on Gita Savitri Devi's YouTube channel revealed fifty-two instances of code-mixing, identifying intra-sentential code-mixing as the most

prevalent type, while pronunciation changes were the least frequent. Similarly, Simanungkalit and Putra (2021), analyzing Daniel Mananta's YouTube content, identified 23 instances of intra-sentential information, 6 intra-lexical cases, and 1 instance involving articulation changes, highlighting factors like topic focus and engaging atmospheres as influential. Dewi et al. (2022) examined Sacha Stevenson's channel, where they identified three primary types of code-mixing—insertion, substitution, and congruent lexicalization—and noted factors such as interlocutor language background, language dominance, language attitude, psycholinguistic motivations, ease of understanding, participant roles, social context, topic, and emotional influences. Nurhayati (2021) analyzed the YouTube talk show CURHAT BANG Denny Sumargo featuring Cinta Laura, noting intra-sentential code-mixing and insertion as the most frequent types and identifying nine purposes for code-mixing, where clarification was the most common.

Previous studies employing audiovisual content as data sources have yielded a range of findings. Herman et al. (2022) conducted research using television commercials, identifying two forms of code-mixing in advertisements: compatible lexicalization and insertion. This study also identified four key motivations for code-mixing, including expressing group identity, discussing specific topics, using interjections, and providing conversational clarification. Another study by Lubis and Hendrawan (2022) analyzed "Breakout", a music program on Net TV, as a data source. They found that presenters predominantly used two types of code-mixing—intra-sentence and extra-sentence—along with five types of code-switching: situational, metaphorical, tag code, inter-sentential, and intra-sentential. This study, drawing on Dell Hymes' framework, highlighted that presenters' speech is influenced by participant roles, the setting and scene, goals, keys, and interaction norms, in addition to language background, topics, and vocabulary limitations in Indonesian that sometimes lead to ambiguity. Further, Adnyani et al. (2023) explored the "Cek Toko Sebelah" movie, identifying three types of code-mixing: congruent lexicalization, alternation, and insertion. In this case, code-mixing served two main purposes: fulfilling communication needs and conveying social status. Achmad et al. (2023) used the "Ali & Ratu-Ratu Queens" movie as their data source, concluding that tag code-switching and insertion code-mixing were the most prominent forms observed in the dialogue. Following this, Aslim et al. (2024) conducted a study involving TikTok beauty influencers, revealing 954 instances of code-mixing in video transcripts. Based on Hoffman's categorization, three types of code-mixing were identified: intra-sentential, intra-lexical, and pronunciation variations, with beauty influencers' primary reason for code-mixing attributed to habitual language use. Lastly, Putra et al. (2024) provided insight using data from the "99 Cahaya di Langit Eropa" movie. Their findings indicated that inter-sentential code-switching was the most frequently occurring type, appearing in 37 utterances, with the main purpose of code-switching and code-mixing being to emphasize points in the conversation, observed in 13 utterances.

This study introduces a new exploration of code-mixing on YouTube by analyzing content from three different YouTube channels, especially those featuring well-known Indonesian YouTubers such Sandy Kristian Waluyo, Gita Savitri Devi, and Jesselyn Lauwreen. Although Gita's YouTube channel has been previously studied by Astri & Fian (2020), their research focused on different videos, whereas this study uses content specifically uploaded on July 27, 2024. Notably, there has been no prior research examining videos from Sandy and Jesselyn's YouTube channels. Hence, this research aims to clarify what types of code-mixing used in YouTube content and why code-mixing is used.

Literature Review

Code-Mixing

According to Musyken (2006), code-mixing is a manner of speaking that incorporates a significant amount of morpho-syntactic and/or lexical elements from at least two distinct languages. Holmes (2022) defines code-swiching as the same as code-mixing, when people switch codes in a domain or social situation. Lubis and Hendrawan (2022) state that code-mixing is found mainly in informal interactions. In such contexts, speakers tend to switch between languages in a single utterance or conversation. This study will examine code-mixing based on its type and reasons. Hoffman (1991) identifies three types of code-mixing: intra-sentential, intra-lexical, and mixing that involves a change in pronunciation. There are several types of code-mixing according to Nurhayati (2021); insertion (words or phrases), alternation (clauses), and congruent lexicalization (dialect). Nurhayati (2021) uses the theory from Musyken (2000). The theory from Musyken (2000) is applied in order to discover the types of code-mixing.

Insertion

Insertion is the proses of adding lexical pieces or complete constituents from one language into the structure of another. The underlying languages or matrix structural characterisics place restrictions on this process. For example:
1. Aku lagi **meeting** sama tim sekarang.
2. **Weekend** ini aku mau liburan ke Bali

Alternation

Alternation refers to the limitation of mixing, which depends on the compatibility or equivalence of the languages at the point where the mix occurs, typically at the clause level. For example:

1. Aku udah bilang dari tadi, **but you never listen to me**.
2. Iya, aku selalu coba makan sehat. Tapi malemnya when **I'm hungry**, suka ngemil.

Congruent Lexicalization

Congruent lexicalization involves incorporating elements from different lexical sources into a common grammatical framework. It also reflects the impact of dialect on language usage. For example:

1. Aku tadi udah **nge-check** the schedule, tapi kayaknya salah.
2. Gue bakal join meeting nanti, udah **confirm** ke mereka.

People mixed from their native languages to English for a variety of reasons. Here are the seven reasons for bilinguals or multilingual to switch or mix their languages (Harya, 2018), those are:

Talking about a particular topic

Speakers may switch languages when discussing specific topics that are more easily or naturally expressed in one language. This often happens when a term or concept does not have an exact equivalent in another language or when the nuances to be conveyed are better captured in that language. For example:
"Aku suka traveling, especially kalau **talking about Europe**."

Quoting somebody else

Switching occurs when a speaker quotes someone, often in the original language. This allows the speaker to maintain the authenticity of the words spoken by the quoted person and convey the exact meaning or emotion. For example:
"Dia bilang, **'I will always support you,**' dan itu bikin aku semangat."

Being emphatic about something

Language mixing is used to emphasize a point or emotion. When speakers want to emphasise the importance of a statement, they sometimes use phrases in different languages to give it greater impact. For example:
"Aku udah bilang dari kemarin, **this is really important!**"

Interjection

Mixing languages is common when using interjections to express sudden feelings or reactions. Interjections often go untranslated as they are emotional expressions spoken spontaneously. For example:
"**Oh my God**, kamu serius?"

Repetition used for clarification

Speakers may repeat a phrase in another language to ensure understanding. This is important in situations where there is potential for confusion or where the listener may not be familiar with certain terms. For example:
"Kita harus pergi sekarang, **we have to go now**."

Intention of clarifying the speech content for interlocutor

Language mixing helps clarify information for the listener by expressing it in both languages. This is useful in contexts where the listener may be more comfortable with one language than the other, or when the concept being expressed is easier to understand in a particular language. For example:
"Jadi meeting-nya besok jam 10, **so the meeting is tomorrow at 10**."

Expressing group identity

Code mixing can be used to express belonging to a particular social group that shares bilingual or multilingual abilities. For example:
"Kita emang anak-anak **creative team**, jadi biasa kerja sampai malem."

Research Methods

This study uses a qualitative descriptive approach to analyze the types of code-mixing found in YouTube content, with a particular focus on videos from Indonesian YouTubers Sandy Kristian Waluyo, Gita Savitri Devi, and Jesselyn Lauwreen. A qualitative approach was chosen because it allows for a comprehensive understanding of the linguistic patterns and contexts in which code-mixing occurs. Unlike conventional quantitative procedures, a qualitative approach was used for data collection, analysis, interpretation, and report writing (Creswell 2009). in line with what Miles and Huberman (1994) said, to describe the major phases of data analysis: data reduction, data display, and conclusion drawing and verification. The primary data for this study were transcripts of selected YouTube videos, including "HOW SANUK STARTED, WHY SANUK? WHY THAI BOAT NOODLES?" by Jesselyn Lauwreen, "32 🎂 " by Gita Savitri Devi, and "nus dorm move in vlog | utown residence" by Sandy Kristian Waluyo. These videos were selected based on their rich use of code mixing between Indonesian and English, making them ideal for this analysis. In this study, the researcher used content analysis to collect data. According to Kothari (2004), content analysis consists of analyzing the content of documentary materials such as books or newspapers that can be spoken or printed in full and analyzing communication. The data collection process began with the selection of relevant videos, followed by transcription of each video to capture the exact spoken content, including all instances of code mixing. The transcriptions were then arranged in a table format, where each sentence was analyzed according to the type of code mixing using Musyken's theory (2000), which includes Insertion, Alternation, and Congruent Lexicalization. In addition to identifying the types, the reasons for code mixing were analyzed based on Hoffman's theory (1991) which outlines several motivations for code mixing, such as talking about a particular topic, quoting others, emphasizing something, interjections, repetition used for clarification, the intention to clarify the content of the utterance for the interlocutor, and expressing group identity. After the types and reasons were classified, the frequency of each type and reason was calculated. The percentage for each category was determined using the following formula:

$$Percentage = \left(\frac{Frequency\ of\ specific\ type\ or\ reason}{Total\ occurances\ of types\ or\ reasons}\right) \times 100$$

Subsequently, the data were organized into a frequency table that displays the types and reasons for code-mixing in each video, facilitating a more straightforward comparison and analysis of the results. The researcher then composes a discussion of the findings and concludes with a summary.

Research Results and Discussion

Types of Code-Mixing

The findings of this study show the data of types of code-mixing with the details below:

Table 1. Frequency of Types of Code Mixing in Jesselyn Lauwreen YouTube Channel.

No	Type of Code Mixing	Frequency	Precentage
1	Insertion	15	46.88%
2	Alternation	7	21.88%
3	Congruent Lexicalization	10	31.25%
	Total	**32**	**100%**

Table 2. Frequency of Types of Code Mixing in Gita Savitri Devi YouTube Channel

No	Type of Code Mixing	Frequency	Precentage
1	Insertion	88	65.67%
2	Alternation	36	26.87%
3	Congruent Lexicalization	10	7.46%
	Total	**134**	**100%**

Table 3. Frequency of Types of Code Mixing in Sandy Kristian Waluyo YouTube Channel

No	Type of Code Mixing	Frequency	Precentage
1	Insertion	38	60.32%
2	Alternation	16	25.40%
3	Congruent Lexicalization	9	14.29%
	Total	**63**	**100%**

According to Musyken (2011), code-mixing is frequently observed and, in some communities, occurs at a very high rate. In the context of this study, the research subjects—Indonesian YouTubers studying abroad—belong to a similar community, which may explain the high frequency of code-mixing found in their content. The data reveals that one-word code-mixing at the sentence level is the most prevalent across all three research subjects, followed by code-mixing at the phrase and clause levels. Code-mixing in the form of congruent lexicalization appears less frequently but still shows significant usage. This pattern, with insertion being the dominant type, followed by alternation and congruent lexicalization, is consistent across all subjects. This trend may be attributed to the shared experience of living and studying abroad, where speaking English—despite being a foreign language for the subjects—becomes a necessity. This aligns with Holmes' (2013) assertion that switching languages can act as a marker of identity, reflecting the speakers' social and cultural positioning in a multilingual context.

These findings echo the work of Herman et al. (2022), who also identified insertion as the most frequent type of code-mixing. However, they contrast with the findings of Simanungkalit and Putra (2021), who observed that intra-sentential code-mixing, occurring within phrases and clauses, was the most common. This divergence may be due to differences in the context of code-mixingwhile the current study focuses on YouTubers' content, which may favor simpler and more frequent forms of insertion, Simanungkalit and Putra's research may have examined contexts where more complex structures are employed. This variation underscores the need for further investigation into how context influences the patterns and types of code-mixing.

Reasons of Code-Mixing

The findings of this study also show the data of reason of code-mixing with the details below:

Table 4. Frequency of Reason of Code Mixing in Jesselyn Lauwreen YouTube Channel

No	Reason for Using Mixing	Frequency	Percentage (%)
1	Talking about a particular topic	24	70.59
2	Quoting somebody else	1	2.94
3	Being emphatic about something	2	5.88
4	Interjection	1	2.94
5	Repetition used for clarification	0	0.00
6	Intention of clarifying the speech content for interlocutor	1	2.94
7	Expressing group identity	5	14.71
	Total	**34**	**100**

Table 5. Frequency of Reason of Code Mixing in Gita Savitri Devi YouTube Channel

No	Reason for Using Mixing	Frequency	Percentage (%)
1	Talking about a particular topic	113	81.88
2	Quoting somebody else	5	3.62
3	Being emphatic about something	1	0.72
4	Interjection	2	1.45
5	Repetition used for clarification	0	0.00
6	Intention of clarifying the speech content for interlocutor	8	5.80
7	Expressing group identity	9	6.52
	Total	**138**	**100**

Table 6. Frequency of Reason of Code Mixing in Sandy Kristian Waluyo YouTube Channel

No	Reason for Using Mixing	Frequency	Percentage (%)
1	Talking about a particular topic	30	62.5
2	Quoting somebody else	0	0.00
3	Being emphatic about something	3	6.25
4	Interjection	5	10.42
5	Repetition used for clarification	0	0.00
6	Intention of clarifying the speech content for interlocutor	1	2.08
7	Expressing group identity	9	18.75
	Total	**48**	**100**

In the findings above, the subjects exhibit a similar pattern in their reasons for code-mixing, with the most frequent reason being "talking about a particular topic." According to Trisna (2018), speakers often switch languages when discussing specific topics that are more easily or naturally expressed in one language. Another common reason is "expressing group identity," which could be attributed to the fact that the subjects are studying in international settings, where their peers frequently use English when speaking to one another. As a result, the subjects often refer to their group identity in English rather than Indonesian. These two reasons are the most prevalent among the subjects in this study. Trisna (2018) identifies seven reasons for code-mixing: talking about a particular topic, quoting somebody else, being emphatic about something, interjections, repetition used for clarification, intention of clarifying the speech content for the interlocutor, and expressing group identity. Based on the findings of this study, one reason for code-mixing was not observed in any of the subjects: "repetition used for clarification."

The findings of this study align with previous research by Simanungkalit and Putra (2021), who also found "talking about a particular topic" to be the most common reason for code-mixing. However, this contrasts with Nurhayati's (2021) study, which identified "intention to clarify" as the most frequent reason for code-mixing.

Conclusion

Based on their frequent usage of code-mixing, the study suggests that Jesselyn Lauwreen, Gita Savitri Devi, and Sandy Kristian Waluyo, three Indonesian YouTubers, use insertion as the most common kind of code-mixing. All three of the YouTubers follow this same trend, with talking about a particular topic appearing as the most often cited justification for code-mixing. The results show that code-mixing is used to promote more understandable communication, particularly when talking about complicated subjects that are better articulated in English. Furthermore, the noteworthy application of code-mixing for expressing group identity implies that these YouTubers employ language to either project a broader global identity to their audience or to establish a connection with certain social groupings.

The contribution of this study is to provide deeper insights into the phenomenon of code-mixing in the context of digital communication, particularly on the YouTube platform. By analysing the types and reasons for code-mixing by Indonesian YouTubers, this research adds to the understanding of how content creators interact with their audiences through language. In addition, the results of this study can serve as a reference for further studies in linguistics and communication, as well as a consideration for educational practitioners in understanding the language dynamics that develop among the younger generation on social media.

References

Achmad, S., Azkarahma, I., Permana, S. A., & Ashyya, S. (2023). Figuring code-switching and code-mixing use in the movie Ali & Ratu-Ratu Queens. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 11(2), 1529-1542.

Adnyani, P. R., Maharani, S. A. I., & Savitri, P. W. (2023). An analysis of code-mixing found in the Indonesian movie: Cek Toko Sebelah. Linguistika: Buletin Ilmiah Program Magister Linguistik Universitas Udayana, 30(1).

Aslim, M., Eliwarti, E., & Hadriana, H. (2024). An analysis of code mixing used by beauty influencers in TikTok. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 12(1), 89-97.

Astri, Z., & Fian, A. (2020). The sociolinguistics study on the use of code mixing in Gita Savitri Devi's YouTube channel video. Seltics Journal: Scope of English Language Teaching Literature and Linguistics, 83-92.

Creswell, J. W. (2009). Research design: Qualitative, quantitative, and mixed methods approaches (3rd ed.). SAGE Publications, Inc.

Dewi, N. M. Y. S., Titania, K. S., Kardana, I. N., & Muliana, I. N. (2022). Code-mixing used in the vlog on the Sacha Stevenson YouTube channel. KnE Social Sciences, 38-44.

Harya, T. D. (2018). Sociolinguistics (code: code switching and code mixing). LENTERA: Jurnal Ilmiah Kependidikan, 11(1), 87-98.

Herman, H., Van Thao, N., Purba, R., & Simanjuntak, N. S. U. (2022). Attracting viewers through advertisement by using code mixing: A sociolinguistics study. Anglophile Journal, 2(2), 80-88.

Holmes, J. (2013). An introduction to sociolinguistics (4th ed.). Routledge.

Holmes, J., & Wilson, N. (2022). An introduction to sociolinguistics. Routledge.

Hudson, R. A. (1996). Sociolinguistics (2nd ed.). Cambridge University Press.

Kothari, C. R. (2004). Research methodology: Methods and techniques. New Age International.

Lubis, S. E., & Hendrawan, D. (2022). Code mixing and switching used in the breakout music program of Net TV. Sinkron: Jurnal dan Penelitian Teknik Informatika, 6(3), 860-873.

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded sourcebook. Sage.

Muysken, P. (2000). Bilingual speech: A typology of code-mixing. Cambridge University Press.

Muysken, P. C. (2006). Code-mixing and multilingual communication. In P. Auer & L. Wei (Eds.), Multilingual communication (pp. 303–328). Cambridge University Press.

Muysken, P. (2011). Code-switching. In R. Mesthrie (Ed.), The Cambridge handbook of sociolinguistics (pp. 301–314). Cambridge University Press.

Nurhayati, S. (2021). A study of code switching and code mixing on YouTube talk show Curhat Bang Denny Sumargo with Cinta Laura. International Journal of English and Applied Linguistics (IJEAL), 1(3), 268-282.

Putra, I. M., et al. (2024). The analysis of code switching and code mixing in 99 Cahaya di Langit Eropa movie: A sociolinguistic study. Education Quarterly Reviews, 7(2).

Siahaan, J. B., & Johan, M. (2020). The study of code mixing on Melaney Ricardo's video YouTube channel: Sociolinguistics approach. Humanitatis: Journal of Language and Literature, 7(1), 13-22.

Simanungkalit, A. R., & Putra, E. E. (2021). Code mixing analysis on Daniel Manantaa's YouTube channel video: Sociolinguistic approach. IDEAS: Journal on English Language Teaching and Learning, Linguistics and Literature, 9(1).