

Constructing Identity through Code-Switching in Social Media Contexts

TU Jie

Wuhan University of Technology, Wuhan, China

This study focuses on Chinese social media platforms (Weibo and Douyin), adopting a mixed-methods approach combining quantitative and qualitative research to explore patterns of code-switching in online language and its role in identity negotiation among internet users. Quantitative analysis of 208 valid questionnaires and subsequent qualitative discourse analysis reveal that 83% of users unconsciously engage in Chinese-English code-switching, primarily demonstrating intra-sentential switching patterns where English words embed as salient “figures” against the Chinese “ground.” Results of the Chi-square test indicate significant correlations between code-switching frequency, types, and users’ identity dimensions ($p < 0.05$). Data identified three identity functions: cultural belonging, in-group signaling, and persona curation. The study confirms that code-switching in digital contexts has transcended the level of linguistic efficiency to become an important social infrastructure for internet users to construct their identities.

Keywords: social media, code-switching, identity, Figure-Ground Theory, online language

Introduction

In the contemporary era, technological innovation and globalization have converged to redefine human communication paradigms. Social media platforms, with Weibo and Douyin at the forefront in the Chinese digital landscape, have emerged as dominant arenas for interpersonal interaction, information dissemination, and cultural exchange. According to the latest report from the China Internet Network Information Center (2024), over 90% of Chinese internet users actively engage with social media, generating vast amounts of user-generated content daily.

As a product of language contact and cultural collision, CS embodies the dynamic interplay between individual linguistic choices and broader social structures. Benjamín (2020) posits that code-switching is not merely a mechanical language shift but a strategic social practice deeply intertwined with speakers’ cultural capital, power relations, and self-cognition. Sociolinguistic research has long recognized the multifaceted social functions of code-switching. Gumperz (1982) was among the first to highlight its role in signaling social identities, marking social boundaries, and negotiating interpersonal relationships. Bucholtz and Hall (2005) further expanded on this concept, arguing that language use, including code-switching, is central to the

TU Jie, M.A. Candidate in Foreign Language and Literature, School of Foreign Languages, Wuhan University of Technology, Wuhan, China.

construction and performance of social identities, enabling speakers to project specific personas and align with particular social groups.

Despite the rich body of research on code-switching, a significant gap persists in understanding its role in digital social contexts. Most existing studies have concentrated on code-switching in face-to-face interactions, traditional media, or offline written texts, while the unique characteristics and functions of code-switching in social media remain under-explored. Social media platforms introduce novel communicative dynamics, such as asynchronous communication, multimodality, and the formation of virtual communities, which may significantly influence how users employ code-switching and the social meanings associated with it. In particular, the relationship between code-switching and identity negotiation in the digital realm—a process through which individuals actively construct, express, and manage their identities in online spaces—has received limited empirical attention. This study aims to address this research gap by systematically analyzing the patterns of CS in online language on Weibo and Douyin and exploring its role in internet users' identity negotiation.

Literature Review and Theoretical Framework

Code-switching has been a central topic in sociolinguistics for decades, attracting extensive research from multiple disciplinary perspectives. Hoffman (1991) defined code-switching as the alternation between two languages or variants in the same communicative process, emphasizing its role as a linguistic strategy in bilingual or multilingual settings. Early studies on code-switching focused primarily on its classification and structural analysis. Blom and Gumperz (1972) made a seminal contribution by distinguishing between “situational code-switching” and “metaphorical code-switching.” Poplack (1980) introduced a different classification system based on linguistic structure, categorizing code-switching into three types: “inter-sentential code-switching,” “intra-sentential code-switching,” and “tag switching.”

With the rise of digital communication, researchers have begun to explore code-switching in online contexts. Danet and Herring (2007) noted that the unique features of digital media, such as instantaneity, informality, and the use of non-verbal cues (e.g., emojis, abbreviations), have given rise to new forms of code-switching. Online code-switching often takes on simplified and symbolic forms, with users frequently employing abbreviations and emoticons to convey meaning efficiently. For example, in Chinese social media, terms like “yyds” (meaning “eternal god” in English, used to express admiration) and “emo” (short for “emotional,” indicating a mood of sadness) have become widely used code-switched expressions.

The relationship between code-switching and identity has been a key area of inquiry in sociolinguistics. Social constructivism posits that identity is not a fixed or inherent trait but a socially constructed and negotiated concept (Hall, 1996). From this perspective, language plays a crucial role in identity formation, as individuals use language to perform and assert their identities in social interactions.

Cognitive psychology also offers valuable insights into the mechanisms underlying code-switching and identity. The theory of common ground (Clark & Brennan, 1991) suggests that individuals tend to use more concise and efficient language when communicating with others who share similar knowledge backgrounds. In online communication, where users often interact with like-minded individuals within specific virtual communities, this cognitive principle may contribute to the prevalence of intra-sentential code-switching.

Myers-Scotton's (1993) Markedness Model, integrated with the Figure-Ground Theory from cognitive linguistics, provides a particularly useful framework for analyzing code-switching in relation to identity. According to the Markedness Model, speakers make conscious or unconscious choices about their language use based on the markedness values associated with different linguistic forms. In code-switching, the choice of which language to use as the "figure language" (the dominant language in an utterance) and which to use as the "ground language" (the inserted language) can carry significant social meaning. The Figure-Ground Theory, originally proposed by Talmy (2000), posits that in any cognitive scene, the "figure" stands out as the focus of attention, while the "ground" serves as the background or reference point. In the context of code-switching, the "ground language" provides the basic framework for communication, while the "figure language" draws attention and conveys specific meanings.

Research Methods

Research Design

This study employs a mixed-methods research design, combining quantitative and qualitative approaches to achieve a comprehensive understanding of the research topic. The quantitative phase involves the collection and analysis of survey data to identify patterns and correlations related to code-switching frequency, types, and identity dimensions. The qualitative phase focuses on in-depth discourse analysis of selected social media posts to explore the specific strategies and meanings associated with code-switching in identity negotiation. This integrated approach allows for the triangulation of data sources, enhancing the validity and reliability of the research findings.

Research Subjects

The research subjects are active social media users aged 15-35, a demographic that constitutes the majority of online social media users in China. This age group is characterized by its high digital literacy, active participation in online communities, and frequent use of code-switching in online communication. Through stratified sampling, 219 questionnaires were distributed across different regions, genders, and educational backgrounds, and 208 valid questionnaires were recovered, resulting in a validity rate of 94%. The sample demographics are as follows: 49.5% male and 50.5% female; 36.5% undergraduates, 28.8% postgraduates, 21.2% high school graduates, and 13.5% others.

Data Collection

Quantitative Data Collection: The questionnaire, developed based on a review of relevant literature and pilot testing, consists of four main sections: (1) basic information (age, gender, education level, etc.); (2) code-switching frequency (how often users engage in code-switching in different social media); (3) code-switching types (self-reported dominant types of code-switching, such as intra-sentential or inter-sentential); and (4) identity dimensions (measured using a Likert scale to assess users' perceptions of how code-switching relates to their cultural belonging, in-group signaling, and persona curation). The questionnaire's validity was ensured through expert review and a pre-test, and its reliability was confirmed by a Cronbach's α coefficient of 0.82, indicating a high level of internal consistency.

Qualitative Data Collection: After the quantitative data collection, 20 respondents were randomly selected for in-depth qualitative analysis. The selection criteria included diversity in code-switching frequency, types, and demographic characteristics. For each selected participant, approximately 60 social media posts (including texts, comments, and captions) from Weibo and Douyin were collected, resulting in a total corpus of 1,200 code-switching instances. The data collection period spanned three months to capture a representative sample of users' online communication behaviors.

Data Analysis

Correlation Analysis between Code-Switching Frequency and Identity Awareness

The survey results indicate that code-switching is a prevalent phenomenon among social media users. Specifically, 83% of users reported exhibiting unconscious code-switching habits, often using English words or phrases in their Chinese posts without deliberate thought. This high percentage reflects the deep integration of English into Chinese online language, driven by factors such as globalization, the influence of Western media, and the popularity of English-language internet culture. In contrast, 12% of users reported having bilingual usage awareness, consciously choosing to use code-switching to enhance their communication, while 5% actively used code-switching as a strategy to shape their online personas.

To explore the relationship between code-switching frequency and identity dimensions, a Chi-square test was conducted. The results showed significant correlations between code-switching frequency and the three identity dimensions of “cultural belonging,” “in-group signaling,” and “persona curation” ($\chi^2 = 18.72$, $p < 0.01$; $\chi^2 = 12.36$, $p < 0.05$; $\chi^2 = 15.41$, $p < 0.01$, respectively). For example, in the dimension of cultural belonging, the active usage group scored 27% higher than the unconscious group, suggesting that conscious code-switching is more strongly associated with a sense of cultural belonging. This finding aligns with the social constructivist perspective, which posits that language use can be a powerful means of expressing and negotiating cultural identity. Active code-switchers may use English words and phrases to signal their connection to global culture, while also maintaining their Chinese cultural roots, thus creating a hybrid cultural identity.

Distribution Characteristics of Code-Switching Types and Verification of Identity Functions

Based on the Figure-Ground Theory and Myers-Scotton's Markedness Model, code-switching in the collected corpus was classified into two main types: “Chinese ground + English figure” and “English ground + Chinese figure.” The analysis revealed a significant difference in the distribution of these two types, with 89 instances (78%) of “Chinese ground + English figure” and 23 instances (20%) of “English ground + Chinese figure” ($\chi^2 = 22.58$, $p < 0.001$).

The “Chinese ground + English figure” mode was predominantly used in emotional expression and cultural display scenarios. For example, in the post “这家餐厅的氛围超有vibe，拍照很好看！” (“The atmosphere of this restaurant is so vibey, and the photos turn out great!”), the English word “vibe” serves as the figure, emphasizing the emotional and aesthetic qualities of the restaurant. This type of code-switching often leverages the affective and cultural connotations of English words to enhance the expressiveness of the message. As Danet and Herring (2007) noted, English has become a globally recognized language of emotion and style among young people, and its use in Chinese social media can evoke a sense of modernity, trendiness, and cosmopolitanism.

In contrast, the “English ground + Chinese figure” mode mostly appears in academic discussions, such as “Final project的deadline要到了” (“The final project’s deadline is approaching”), where Chinese embedded words play a role in explanation and emphasis, highlighting professional identity attributes.

Conclusion

This study has provided a comprehensive understanding of the patterns and functions of code-switching in online language on Chinese social media platforms and its role in internet users’ identity negotiation. Through a mixed-methods approach, the study has demonstrated that code-switching is not merely a linguistic phenomenon but a complex social practice deeply intertwined with identity construction.

Empirical data confirm that code-switching serves multiple identity functions in social media contexts. Unconscious code-switchers, who constitute the majority of users, integrate into internet culture through habitual language use, reflecting the naturalization of code-switching in digital communication. Conscious code-switchers, on the other hand, strategically use code-switching to achieve specific identity goals, such as confirming cultural belonging, signaling in-group membership, and curating personal images. The significant correlations between code-switching frequency, types, and identity dimensions, as verified by Chi-square tests, further validate the theoretical framework of this study, demonstrating the applicability of the Figure-Ground Theory and Myers-Scotton’s Markedness Model in analyzing digital code-switching.

Limitations in this study include age-group concentration and platform restriction. Future research should expand generational sampling, incorporate multi-platform corpora (e.g., short video/messaging apps), and employ longitudinal designs to trace code-switching’s dynamic co-evolution with identity.

References

- Blom, J. P., & Gumperz, J. J. (1972). Social meaning in linguistic structure: Code-switching in Norway. In J. J. Gumperz & D. Hymes (Eds.), *Directions in sociolinguistics*. New York: Holt, Rinehart & Winston.
- Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4-5), 585-614.
- González Rojas, B. (2025). Code-switching as identity negotiation in multilingual communities. *Bulletin of Language and Literature*, 2(1), 13-23.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Hall, S. (1996). *Introduction: Who needs identity?* In S. Hall & P. du Gay (Eds.), *Questions of cultural identity* (pp. 1-17). London: Sage Publications.
- Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Oxford University Press.