Universitat Pompeu Fabra Barcelona

Treball de fi de màster de recerca

Code-Switching and Identity in Digital China: Language Ideologies in Mandarin–Cantonese WeChat Communication

Nom i Cognoms: Jingjing Tu

Master: Lingüística Teòrica i Aplicada

Edició: 2024-2025

Directors: Dr. Joan Costa Dr. Manuel Pavón-Belizón

Any de defensa: 2025

Col·lecció: Treballs de fi de màster

Departament de Traducció i Ciències del Llenguatge

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Dr. Joan Costa and Dr. Manuel Pavón-Belizón, for their invaluable guidance and support throughout this research process. Their professional insight, constructive feedback, and constant encouragement have played a crucial role in shaping both the content and the quality of this thesis. I feel extremely fortunate and honored to have been mentored by two such dedicated and inspiring scholars.

I would also like to thank my Chinese friends Zanjie and Pipi for their generous support and for helping me gather valuable data for this study. Their enthusiasm and practical assistance have made the research process much smoother and more meaningful. Special thanks go to my family, whose unwavering understanding and support have been my solid foundation throughout this journey. Their belief in me has given me the courage to persist even in the most challenging moments.

Finally, I am deeply grateful to all the friends and classmates who accompanied me during this time. Their emotional support, academic discussions, and mutual encouragement have provided both strength and joy during the long hours of writing and revising. I feel especially thankful to Facundo, Anita, Cynthia, and Anastasia, whose presence and kindness have meant a great deal to me. It has been a privilege to share this journey with such warm and supportive people.

This work would not have been possible without the contribution and kindness of all those mentioned above. Thank you all, from the bottom of my heart.

Abstract

This study explores how digital code choice and code-switching between Mandarin and Cantonese shape identity construction, interactional strategies, and ideological positioning in WeChat group communication. Using a mixed-methods approach, the analysis draws on 370 chat messages (3,568 characters) and 74 questionnaire responses collected from two Cantonese-speaking groups. The results show that Mandarin is dominant, especially in formal or task-oriented exchanges, while Cantonese serves key social and emotional functions. Code-switching often signals shifts in tone, intimacy, or social stance. Users adapt code use according to group type, topic, and interpersonal relationships, balancing clarity with relational goals. The study also reveals ideological perceptions that associate Mandarin with rationality and standardization, while Cantonese expresses warmth and regional identity. These findings highlight how users navigate digital contexts by managing symbolic code values, contributing to broader understandings of multilingual interaction online.

Keywords: Code-switching, Identity, Language ideology, Digital communication, Mandarin, Cantonese

Contents

1. Introduction

1.1 Coexistence of Mandarin and Dialects

Modern Chinese society is characterized by the parallel use of a national standard language—Putonghua (Mandarin)—and a rich variety of regional dialects. Mandarin, based on Beijing pronunciation and northern dialects, was officially adopted in 1956 and has since been widely promoted through education, government policy, and national media. Its institutional role is firmly established by the 2000 Law on the Standard Spoken and Written Chinese Language. Today, Mandarin dominates in schools, workplaces, and public communication, often symbolizing authority, formality, and modernity.

At the same time, Chinese dialects remain vital in everyday life. According to the Ministry of Education (2021), China's dialects fall into ten major groups, including Yue (Cantonese), Wu, Min, Jin, Xiang, Gan, Hakka, Hui, Pinghua, and Mandarin regional varieties. These dialects are used in more intimate and informal domains—especially family, peer interaction, and community life. Many speakers are bilingual, shifting between Mandarin and their native dialect depending on topic, relationship, or communicative setting.

In recent decades, especially in urban areas, an increasing number of speakers grow up with Mandarin as their first language. However, even among monolingual Mandarin speakers, regional features—such as accent, intonation, and vocabulary—often reflect local linguistic ecologies. These Mandarin regional varieties, shaped through long-term contact with local dialects, are commonly used in both primary and secondary social relationships, sometimes blurring the boundary between "standard" and "local." A dialect distribution map from the

Language Atlas of China (2012 edition) is included in Appendix A to visually contextualize this linguistic diversity.

Among the dialects, Cantonese stands out due to its strong regional identity and wide user base—over 60 million speakers primarily in Guangdong, Hong Kong, and parts of Guangxi. In cities like Guangzhou, people often use Mandarin in formal or mixed-group contexts and Cantonese in familial or local settings. As a result, code-switching is common and socially meaningful. It reflects not just linguistic convenience, but also identity positioning, emotional expression, and language attitudes.

This study focuses on such coexistence between Mandarin and Cantonese in digital environments, especially WeChat group interactions. While Mandarin generally holds a dominant position, Cantonese plays a vital role in indexing closeness, authenticity, and local belonging. The dynamics of language choice and switching in this context offer insights into how users manage identity and negotiate social meanings online.

1.2 Language Practice in the Context of WeChat

WeChat is one of the most used digital platforms in China. People use it for everyday conversations. They can chat one-to-one or in groups. In this field, Mandarin is the main language. But dialects are also frequently used, especially in informal or intimate contexts. On WeChat, people use the written representation of dialects. Users often type dialectal expressions using standard Chinese characters that sound similar to express dialectal words. This creative use of writing allows dialects to become visible in text-based interaction on WeChat. For example, Cantonese expressions like “系啊” (“yes”) use characters based on

meaning, since “系” is the copula “to be” in Cantonese.

This kind of “digital dialect writing” is not only for convenience. It is also a social choice. The mixed codes, emojis, and voice clips to express dialects show identity, emotions, and belonging. In our data, we see some users mixing Mandarin and dialects. Some use dialect to show feelings. Others only use Mandarin in all situations. Their language choices in group chats show how they present themselves. These choices also reflect their language attitudes.

1.3 Research Questions and Hypotheses

In general, our study focuses on: How do users switch between Mandarin and Cantonese in WeChat group chats? How does this language choice reflect their identity and language attitudes? More specifically, we propose 16 hypotheses across six thematic categories. These hypotheses reflect the expected patterns of language use, code-switching, and identity performance in WeChat group communications:

Group 1. Code Distribution across Group Types:

H1. Mandarin is more frequently used in work groups.

H2. Cantonese is more frequently used in friend groups.

Group 2. Code-Switching and Identity Variables:

H3. Code-switching practices are shaped by users’ social identity.

H4. Code-switching practices vary depending on the topic of conversation.

H5. Older users tend to switch codes more frequently.

Group 4. Language Ideologies Reflected in Attitudes:

H8. Mandarin is perceived as a formal and appropriate language in serious contexts.

H9.Users tend to use Mandarin in formal or institutional contexts, reflecting its perceived role as an authoritative code.

Group 5. Functional Tendencies in Language Use:

H10. Mandarin is more commonly used for giving commands.

H11. Mandarin is more commonly used for asking questions.

Group 6. Identity Indexicality throgh Code Choice

H12. Cantonese is perceived as warmer and is often used to express emotions and closeness.

H13. Users construct and display their social identity through language choice.

H14. Cantonese shows where they are from.

H15. Mandarin shows professionalism.

H16. Mandarin shows authority.

2. Literature Review

This chapter reviews previous research on language use, focusing on four areas. First, it examines how language choice relates to identity and power, drawing on sociolinguistic and interactional theories. Second, it explores code-switching and diglossia, outlining their types,

functions, and examples in the Chinese context. Third, it addresses language attitudes and ideologies, including perceptions of Mandarin and dialects. Finally, it discusses language practices on digital platforms—especially in WeChat groups—and how these reflect users' identities and social meanings.

2.1 Language, Power, and Identity in Language Choice

Language use is inherently tied to power. As Bourdieu (1991, p. 37) points out, "linguistic exchanges are also relations of symbolic power" that actualize existing social hierarchies between speakers and groups. In multilingual societies, this symbolic value is unequally distributed: standard languages tend to be institutionally privileged, while other varieties—including regional languages and dialects—may be minoritized or assigned lower status depending on sociopolitical ideologies. As O'Rourke (2021, p. 271) notes, "the power relations between different language varieties and their speakers […] is not always equally distributed, with some more valued than others." This imbalance is clearly reflected in the Chinese context. Mandarin, promoted as the national language since 1956 and reinforced by the 2000 Language Law, has become a marker of authority, education, and institutional legitimacy. In contrast, dialects are often discouraged in formal settings or perceived as "less proper" (Chen, 2013, pp. 29–30). According to Kircher and Zipp (2021, p. 12), this reflects a broader cross-cultural trend: standard varieties are globally associated with higher status than non-standard ones.

On digital platforms, these symbolic hierarchies persist. As Liu (2011) observes, in online settings where paralinguistic cues are absent, users often rely on written language to build social roles and project identity. For instance, some users consistently use Mandarin

even when others are typing in dialects, possibly to display a more “professional” or “rational” persona or to conform to perceived norms of correctness.

At the interactional level, language choice also serves as a strategic tool. Gumperz’s (1982) “we-code / they-code” model shows how speakers navigate intimacy and distance through code-switching. This becomes particularly relevant in WeChat group settings, where users alternate between Mandarin and dialects not arbitrarily, but to signal hierarchy, solidarity, or group membership. Yus (2011, pp. 176–177) extends this view into digital communication through his cyberpragmatics model. He argues that users actively compensate for the lack of visual and paralinguistic cues by manipulating linguistic forms—modal particles, emojis, dialect spellings—to express affective proximity or distance. These stylistic choices function as identity management strategies, reflecting users’ negotiation of social relationships. Similarly, Androutsopoulos (2006, pp. 426–427) highlights that language use in computer-mediated discourse is context-sensitive and socially motivated. Users reshape offline norms to construct discursive identities through code alternation and stylistic mixing.

Finally, Bourdieu (1991, p. 30) reminds us that power is embedded in language use itself. A speaker’s linguistic behavior reflects their social positioning, including class, ethnicity, and institutional capital. In this sense, code-switching is not only a pragmatic act, but also a response to symbolic hierarchies. Those with greater communicative capital navigate language choices more freely, while others may conform or self-censor.

Altogether, these perspectives reveal that language alternation in WeChat is neither random nor purely expressive. It is a reflection of users’ social strategies, identity performances, and the power dynamics embedded in everyday interaction.

2.2 Diglossia and Code-Switching

In contemporary China, Mandarin is used nationwide, while Cantonese is widely spoken in southern regions such as Guangdong and Hong Kong in daily life. This creates a diglossic system, in which two language varieties exist within the same community. Each language has its own role. People often switch between them, which is called code-switching.

2.2.1 Diglossia in the Chinese Context

Ferguson (1959, pp. 325–337) introduced the concept of diglossia to describe speech communities where two language codes coexist in a structured relationship: a high variety (H-variety) used in formal domains (education, government, media) and a low variety (L-variety) used in informal, everyday communication. He emphasized three core features: functional separation, prestige asymmetry, and acquisition pattern—where the L-variety is learned at home, and the H-variety through schooling. In China, Mandarin has long served as the H-variety—promoted through state policies and standardized education—while dialects such as Cantonese function as L-varieties, reserved for localized and intimate contexts.

Recent studies have shown that this diglossic structure is evolving under modernization and digital communication. Wang, Chen, and Zhang (2012) highlight the pervasive functional coexistence of Mandarin and dialects in daily life, forming dual linguistic repertoires shaped by communicative domain. Zhou and Liu (2024) further reveal that while Mandarin is recognized for its instrumental value in mobility, education, and employment, dialects are emotionally tied to cultural belonging, especially in southern regions. They emphasize that diglossic boundaries are increasingly shaped by identity stances and communicative contexts,

rather than rigid formal/informal domains. Moreover, digital platforms foster "virtual speech communities" that extend dialect usage beyond geographic boundaries, creating hybrid zones of negotiation between H and L codes.

Woolard's (2016) distinction between anonymity and authenticity offers a valuable ideological lens to interpret these dynamics. Mandarin, as a language of anonymity, is perceived as neutral, standardized, and institutionally endorsed, enabling broad communication. Cantonese, by contrast, exemplifies authenticity—anchored in locality, heritage, and group identity. On platforms like WeChat, users often adopt Mandarin to project neutrality or authority, while dialects are used to express intimacy, regional pride, or cultural alignment.

Integrating Ferguson's structural model with Woolard's ideological framework, and building on empirical insights from Chinese research, this study approaches diglossia as a dynamic and context-sensitive system. Rather than a fixed hierarchy, it reflects flexible identity performances mediated by setting, social relationships, and platform affordances.

2.2.2 Types and Pragmatic Functions of Code-Switching

Code-switching refers to the alternation between two language codes—such as Mandarin and a local dialect—within the same conversation. Beyond linguistic form, it conveys identity, stance, and interactional meaning. Gumperz (1982, pp. 77–84) distinguished situational code-switching, which reflects shifts in context or setting, and metaphorical switching, which signals changes in social meaning or speaker alignment. These types emphasize the communicative rather than grammatical role of switching.

Building on Gumperz's model, this study proposes a finer categorization adapted to

Chinese digital contexts, particularly WeChat group chats. While Gumperz highlighted social indexing through switching, he did not elaborate detailed structural distinctions. Our coding framework (see Section 3.3.1) refines this with four functional types designed for online discourse.

Auer (1998, p. 4) further developed this perspective, viewing code-switching as an interactional resource that contextualizes shifts in speaker role, topic, or stance. Switching thus reflects dynamic social positioning and narrative structuring rather than random language mixing.

Code-switching also mediates social distance. Following Gumperz (1982) and Woolard (2016), dialects often serve as we-codes, marking emotional closeness in intimate settings like family or friend groups. This aligns with Woolard's (2016, p. 21) notion of the ideology of authenticity, where dialects index local identity and relational warmth, while Mandarin represents neutrality and institutional legitimacy.

Methodologically, we drew on Siebenhaar's (2006) study of Swiss German online chats, which offered a structured framework for identifying code-switching types and pragmatic functions in digital interaction. We adapted his typology for the Chinese context, focusing on Mandarin–Cantonese alternation in WeChat. Our analysis explores how switching operates not only as a linguistic shift but as a socially embedded act shaped by relationship, purpose, and communicative norms.

2.3 Language Attitudes and Language Ideologies

Language is not only a tool for communication. It also expresses attitudes and positions. When people choose a certain code, it often shows their feelings and value judgments. These

attitudes are not always spoken directly. But they strongly shape how people use language.

2.3.1 Language Attitudes and the Construction of "Standard" Language

Language attitudes refer to people's cognitive, emotional, and behavioral responses toward specific language varieties (Kircher & Zipp, 2021). These can be explicit—such as openly preferring Mandarin—or implicit, like associating it with formality or prestige without conscious reflection. For instance, one may believe Mandarin is more "correct" (cognition), prefer using it at work (behavior), yet feel emotionally closer to dialects (affect).

Such attitudes are socially constructed, not inherent. As Bourdieu (1991) argues, a language gains value through its association with institutional power—education, policy, media—not from intrinsic features. "Standard" languages emerge when the social elites who use them hold symbolic capital, which is reinforced via schooling, exams, and media. Mandarin exemplifies this: its state-led promotion has established it as the "official" and "correct" language. In contrast, dialects are often marginalized—mocked, restricted, or labeled as inappropriate in formal settings.

Julià Munné (2019), comparing China and Catalonia, highlights that language standardization—though often justified for intercomprehension—also serves political and ideological aims. In both cases, official languages expanded in public life while local varieties were relegated. This symbolic hierarchy extends into digital spaces: Mandarin tends to dominate in serious or institutional messages, while dialects are used for intimacy, humor, or local bonding.

2.3.2 Signs of Language Ideology

Language ideology often works in hidden ways. Gal (Gal, 2023, Chap 3) emphasizes that language ideologies often operate through a mediated process of semiotic interpretation and social judgment. Rather than being openly declared, these ideologies emerge from implicit associations between how people speak and how they are socially evaluated. This process is not neutral but ideological: speakers are judged not only for what they say, but for how their language aligns with dominant norms of intelligence, respectability, or credibility. Such ideologies, though often invisible in casual interaction, influence how speakers are positioned in both informal and formal settings.

Liu Jin (2011) noted that language attitudes in digital spaces are more complex and flexible. Because most online conversations are written, users often adjust their writing style and code depending on who they are talking to and in what setting. This helps them express identity, manage social distance, and follow group norms. Among young users, dialects are often used to create humor or show a local identity. These choices may follow group expectations, but they can also respond to or challenge dominant beliefs about what counts as the "standard" way to write or speak (Liu, 2011, pp. 62–67).

2.4 Identity Construction and Language Practices in Digital Contexts

As a form of computer-mediated communication (CMC), WeChat plays a central role in daily interaction. The absence of non-verbal cues leads users to adjust their linguistic forms to convey tone, stance, and identity (Yus, 2011, p. 255). In our WeChat data, users often employ sentence-final particles like "ha", "sai", or "lo" (哈，噻，咯), which function as pragmatic

markers from Chinese dialects. Though semantically empty, they enhance expressiveness and reinforce social closeness. Gumperz (1982, pp. 75–78) noted that code-switching and variation serve to construct group identity, express emotions, and manage social distance—functions amplified in digital settings.

WeChat groups, with their semi-private nature, stable membership, and continuous context, offer a rich environment for observing code choice and identity performance. Frequent exchanges promote self-disclosure and build trust, turning language into a tool not just for conveying information but for expressing social alignment (Yus, 2017, p. 82). In multilingual contexts, code choice becomes central to identity construction (Chen, 2013, pp. 27–28).

Studies by Sun (2018) and Mei (2024) show that digital dialect use supports identity continuity and innovation. Sun's work on Fujianese youth highlights how dialect maintains "tribal" belonging across geographic distance, while Mei shows how WeChat and Bilibili users co-construct "network dialects" that transcend regional norms. These findings underscore that CMC platforms not only transmit language but foster new forms of identity through linguistic creativity and group negotiation. As Yus (2018, p. 10) argues, digital code use is shaped by both user expectations and the semiotic environment—including group norms, platform tools, and atmosphere—making language choice a site of both structure and strategy.

While existing research has emphasized how dialect signals intimacy or locality in digital interaction, few studies compare different dialects or group types. Much Chinese-language literature focuses on single-dialect groups or general youth code-switching

trends. Consequently, the interplay between pragmatic motivations, group type (e.g., family vs. work), and language ideologies remains underexplored. This study builds on previous work by analyzing Cantonese use in real WeChat group data across two group types. By examining code-switching as a strategy for performing roles and expressing stances, we reveal how language practices index beliefs about formality, professionalism, and authenticity—thus connecting micro-level interaction to broader ideologies.

3. Research Methodology

By combining corpus data, questionnaire and example analysis, this study proposes a Triadic Analytical Model to study WeChat language practices. This model includes three key elements: identity, code, and language ideology. Unlike models that focus only on code-switching or interaction strategies, we highlight the social meaning of codes.

3.1 Research Design and Strategy

This study uses a small-sample and mixed-method design. It includes both quantitative and qualitative analysis. We counted data from a WeChat group chat corpus (See Section 3.2.1 for details) and a questionnaire (See Section 3.3.4 for details).

We aimed to use corpus coding to get measurable data (see Appendix B for coding examples). This includes language use, frequency of switching, and the function of different codes. The questionnaire helped us check user attitudes. These answers offered a subjective view to support what we found in the chats. At the same time, we used qualitative analysis of

selected chat segments from the WeChat corpus. By examining users' language choices in context, we aimed to understand processes of identity construction and to identify implicit language ideologies. This step-by-step approach asked: What code is used? Why is it used? What kind of identity or attitude does it express?

Taken together, this triangulated approach allows us to examine language use on three levels: structure (what codes and switching patterns are used), function (what pragmatic roles they play), and social meaning (what identities and ideologies are reflected). The pragmatic path focuses on how people use language in chat, and the ideological path focuses on the values behind language use. We asked about some perspectives about Mandarin and Cantonese, which helped us see people's attitudes about codes. The questionnaire thus serves as a supplement to the corpus, helping to compare what users say and what they actually do in conversations.

3.2 Data Sources and Corpus Composition

3.2.1 Background and Diversity of WeChat Group Samples

A WeChat group is usually created for a reason, such as family, work, or friendship. The group members vary in familiarity depending on group type: family and friends groups consist of primary relationships with strong personal ties, while work groups involve secondary relationships, marked by role-based interaction and less emotional closeness. They speak in a natural way. So, this setting is good for observing how people use language in daily life. Compared to public platforms, WeChat shows how people communicate in private or close relationships. It shows how they express who they are. The natural way they speak, the

tone they use, and the type of language (Mandarin, dialect, or mixed) is often the result of shared habits, not just message content. In these small online communities, people may keep their real-life identity or try to build a new one. Also, all chats are saved automatically. So, it is easy to collect and build a corpus from real, natural conversations.

The chat data comes from two Cantonese WeChat groups (work group & friend group) where the researcher has been granted access with informed consent. Each group has a stable group of members and a complete chat history. The original research design included four WeChat groups: two using Cantonese and two using Sichuanese. All four datasets underwent complete quantitative and qualitative analysis following the same coding framework presented in this thesis. However, due to the strict page limit set for this master's dissertation, only the two Cantonese groups are presented and discussed in the following chapters. The Sichuanese dataset remains available for future comparative research, especially to investigate cross-dialectal identity practices in similar digital settings.

Below are the basic details of the four groups:

Group A (C1): Cantonese Work Group. A work group with 12 members (3 male, 9 female), all from Guangdong. They are between 18–36 years old. There is a clear workplace hierarchy. Most messages are in Mandarin, but Cantonese appears in relaxed conversations. This group is useful to study the link between language choice, power, and formality.

Group B (C2): Cantonese Friends Group. A friend group with 14 members (4 male, 10 female). One member is not from Guangdong but knows some Cantonese. All are aged 18–36. They chat often and joke a lot. Mandarin is the main code, but there is frequent code-switching into Cantonese, especially in voice messages, jokes, or emotional content.

This group gives rich data for studying identity and emotion in informal settings.

Overall, the group members fall into three main types based on their language background and usage patterns. First, balanced bilinguals: these are speakers who grew up with both Cantonese and Mandarin as native languages and switch codes flexibly depending on context. Second, dialect-dominant users: they mainly use Cantonese in daily life and shift to Mandarin in formal or written contexts. Third, Mandarin-dominant users: these are typically younger participants who primarily speak Mandarin, especially in school and online, and have only limited level of Cantonese.

3.2.2 Corpus Collection and Selection Criteria

We collected group chat messages from March to May 2025, and built a WeChat group corpus (Appendix B). The chats include scenes such as daily life, work talk, and casual chatting among friends. The language is natural and the data is rich, which is good for pragmatic analysis.

Before collecting any data, we explained the research goals to all group members through the group administrators. All members were told that participation was voluntary and anonymous. Then we sent the online questionnaire link, which included a consent form on the first page (Appendix D). This form explained the purpose of the study and how the data would be used. Only after they read and agreed to the consent form could they continue to the survey. The chats were originally written before formal consent. But we only started to export and use the records after receiving this written agreement through the survey. This ensured that all data use followed ethical guidelines. Personal information in the chat data was removed or anonymized. For the corpus, we forwarded chat records and manually transcribed

the content. We turned voice messages into text, one by one. If there were stickers or pictures related to the meaning, we added a short explanation.

In the end, we collected 370 valid messages, about 3568 Chinese characters (tokens counted as individual characters). The distribution is as follows:

Group	Messages	Tokens
Group A “Cantonese Work Group”	157	1292
Group B “Cantonese Friends Group”	213	2276
Total	370	3568

Table 1: Corpus distribution of WeChat groups

3.3 Coding System and Variable Categories

3.3.1 Code Types and Switching Patterns

In this study, we treat Mandarin and Cantonese as distinct codes. Mandarin refers to the standardized written form taught in schools and used in media and official contexts. It follows national norms in grammar, vocabulary, and uses simplified Chinese characters. Cantonese includes Cantonese-specific vocabulary, sentence structures, expressive particles, and, when transcribed, may appear in Jyutping or Chinese characters—either simplified or traditional, depending on user preference. While most participants used simplified characters, occasional traditional characters appeared for stylistic or habitual reasons. We also identify a “mixed code” category, where both Mandarin and Cantonese appear in the same message. Although we did not further classify code dominance, most mixed messages in our corpus appear Mandarin-dominant (see Section 4.1).

About types of code-switching, we defined four types: inter-sentential switching, intra-sentential switching, tag insertion, and lexical borrowing (Examples are from the author's WeChat corpus):

A. Inter-sentential Switching. Different codes are used in separate sentences.

Example:"他在群裡破防了。***笑到我聲都變埋***。" ("He broke down emotionally in the group chat. ***I laughed so hard that even my voice changed***.") The first sentence is in Mandarin, and the second sentence is a Cantonese.

B. Intra-sentential Switching. A single sentence contains words from two or more codes, fully mixed in grammar. To code this, the researcher must know the dialect. If needed, we consulted native speakers.

Example:"他跟那个***慢慢浸佢***好像***啊***。" ("He sounds similar to that one guy who says ***soak it slowly***"). The sentence mixes Mandarin grammar with Cantonese words like "浸佢" ("soak") and "啊" (a Cantonese expressive particle). The code-switching happens inside the sentence, blending both syntax and vocabulary.

C. Tag Switching. Dialect particles or exclamations are added to a Mandarin sentence. This type is often used for humor or closeness.

Example:"电话暂时无法接通***嘎***。" ("The phone cannot be reached at the moment ***ga***."). "嘎 (ga)" is a Cantonese particle, expressing emphasis or slight surprise. The main sentence is in Mandarin, but the ending tag adds Cantonese expressive particle.

D. Lexical Borrowing. Only one or two words from another code are used, without grammar mixing. This type shows how words move across codes. It is useful in studying contact between languages.

Example:“你可以要个***蒜蓉蒸***” (“You can order a ***suan-rong-zheng***.”) The rest of the sentence is in Mandarin, showing how a dialect word is borrowed into a Mandarin frame.

When a message contained multiple types of switching, we coded only the most prominent one to ensure consistency. For example, if both tag switching and lexical borrowing appeared, we marked it as lexical borrowing, since lexical items usually carry more semantic and stylistic weight than particles. While this approach may overlook some co-occurring features, it simplifies analysis and enables clearer frequency counts. Given the study’s limited scale and focus on socially salient patterns, we believe this method effectively highlights dominant code-switching strategies. Future research could adopt multi-tag coding to capture greater complexity.

3.3.2 Framework for Pragmatic Functions

Pragmatic function refers to what social meaning is carried by how people use language in real situations. To analyze these functions in a CMC context like WeChat, we built a seven-category coding system. It is based on Francisco Yus's (2011) theory of cyberpragmatics and Gumperz's (1982) sociolinguistic views on interactional strategies. Examples are from the author’s WeChat corpus:

A. Information exchange. This refers to messages that give or ask for facts. It includes updates, questions, or reports about daily matters. The tone is usually neutral. This is the most basic function in WeChat group chats. For example: “车祸原因堵了一会儿车” (“The car accident caused a short traffic jam.”)

B. Expressing emotion. This function shows feelings or attitudes. People use it to express happiness, care, anger, or worry. These messages often include exclamation marks,

emojis, or voice notes. They are more subjective than information exchange. For example: “太好笑了” (“That’s so funny!”).

C. Creating intimacy. This function builds closeness and group belonging. It often uses dialects. One common way is using dialect nicknames or personal expressions to call others. These messages show strong in-group identity and emotional connection. For example: “舍不得姨回去了啊” (“I don’t want auntie to leave!”).

D. Expressing distance / politeness. This is used to show respect, keep social distance, or mark social roles. People usually use Mandarin in these cases. This happens more often in formal settings or when people have unequal power. For example: “哦，我搞错了” (“Oh, I made a mistake.”) – the use of a polite correction in Mandarin shows deference and maintains formality.

E. Humor / irony. This function includes jokes or indirect criticism. It helps to create a relaxed mood or say something critical in a softer way.

F. Quoting others. This means repeating someone else’s words. It often adds emphasis or signals a topic shift. It may also lead to a code switch.

G. Group alignment. This means people choose language based on who they talk to or the type of group. For example, using Cantonese in a Cantonese-speaking group. This shows a wish to fit in. It also reflects how speakers are sensitive to relationships and group identity.

We aimed to explore how the use of a certain code relates to specific pragmatic goals, such as politeness, emotional tone, or speaker alignment. In cases involving dialect expressions whose meanings were uncertain, the author consulted native speakers of Cantonese for clarification.

3.3.3 Identity Indexing and Ideological Clues in Code Use

In WeChat group chats, people's choice of Mandarin or Cantonese can reflect their identity performance and their underlying language ideologies. In this study, we coded each message from two dimensions: (1) how the speaker indexes identity ("who they are"); and (2) what the code choice reveals about their language ideologies.

Although identity is often fluid and layered, we followed an operational coding scheme adapted from Yus (2011) and Bucholtz & Hall (2005, pp. 591–593). The latter proposed that identity emerges in interaction, is relational and indexical, and always partial and situated. These principles help explain how speakers in digital contexts use language to position themselves in relation to others. Our identity categories aim to capture the most salient role being performed at that moment in the group conversation. we define eight types of identity expression in the chats:

A. Emphasizing regional identity: Using dialect to show or confirm local background.

B. Family/kinship role expression: Code reflects family roles like parent, child, elder.

C. Generational identity: Code shows the speaker is part of an older or younger generation.

D. Social closeness: Code shows intimacy or solidarity with close contacts.

E. Social distance: Code shows caution, respect, or social gap.

F. Asserting authority: Code signals control, leadership, or decision-making power.

G. Passive adaptation to group norm: Speaker follows the main group code without clear intention.

We also look at signs of language ideology. Based on Woolard (2016)'s "authenticity"

and “anonymity”, we marked several ideologies, for examples, “Positive/Nagetive dialect evaluation”, “Standard language supremacy”.

3.3.4 Questionnaire Design and Survey Details

To complement the corpus analysis and assess the generality of observed patterns, we designed a structured questionnaire (see Appendix C) to explore users’ language choices and attitudes in WeChat groups across different backgrounds. It included both multiple-choice (e.g., Likert scales, frequency items) and open-ended questions. The former provided quantifiable data; the latter captured participants’ views in their own words. We distributed two survey versions: one for group members and one for external participants. The results helped verify whether corpus patterns reflected broader trends.

To meet ethical standards, the survey began with a consent form outlining the research purpose, contact information, and participant rights. Anonymity and voluntary participation were ensured. Only after agreeing to the terms could participants proceed. Data were collected via Wenjuanxing, with all personal identifiers removed before analysis—only coded variables were retained.

4. Quantitative Analysis Results

This chapter presents the quantitative findings of the study, based on a systematic coding of 370 WeChat group chat messages and 74 questionnaire responses (See appendix E for sample background of questionnare). This chapter focuses on the linguistic patterns observed

in the chat corpus and survey data, including code type distribution, switching patterns, pragmatic functions, and identity construction. Together, these data offer a comprehensive picture of how language choice, identity, and ideology intersect in digitally mediated group communication.

4.1 Code Types and Group Differences

Chart 1 compares the proportion of code types used in four WeChat groups (Cantonese work group C1, Cantonese friends group C2.)

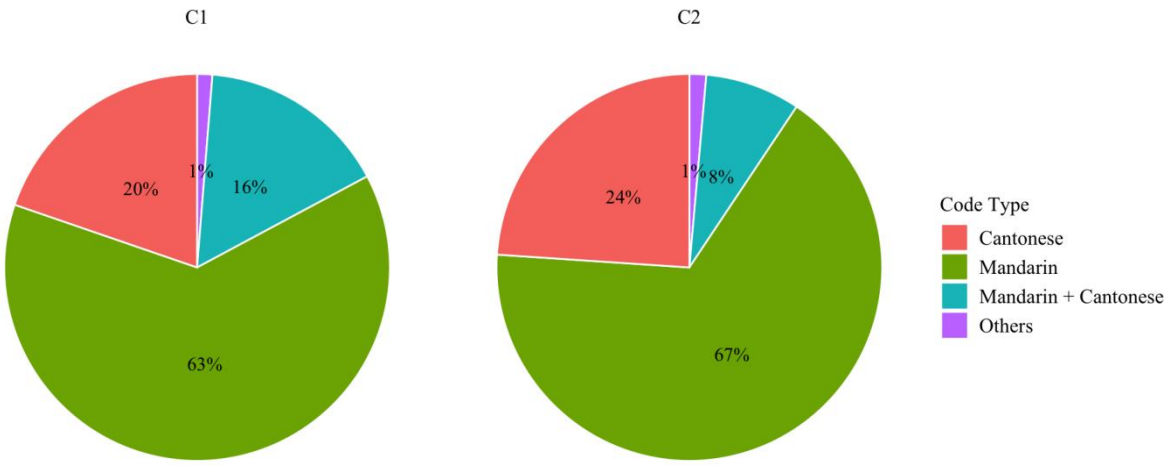


Chart 1: Code Type Distribution by Group

Overall, Mandarin dominates all groups, accounting for 63% in C1 and 67% in C2. As the national standard, Mandarin is the most frequently used code in WeChat chats. Dialect usage reflects group features: in C1, Cantonese (pure + mixed) totals 36%, and in C2, 32%. Surprisingly, the work group (C1) shows slightly more Cantonese than the friend group (C2), possibly because C1 members are all local Cantonese speakers, while C2 may prioritize efficient communication in Mandarin. Thus, H1 is partially supported, but H2 is not fully

confirmed.

Questionnaire responses further confirm these patterns. For Q9 ("When typing in WeChat, which code(s) do you use most?"), 43.2% chose "Mainly Mandarin," while only one chose "Mainly dialect." About 25.7% said it depends on the group (e.g., Mandarin in work groups, dialect in family groups), and 28.3% selected "Mandarin+Cantonese," indicating habitual code-mixing. These results highlight Mandarin's dominance in written chats, especially among younger users, though dialects persist in close-knit groups.

4.2 Types of Code-Switching and Triggering Features

4.2.1 Code-Switching Patterns across Age and Contexts

Chart 2 shows the distribution of code-switching types by age group.

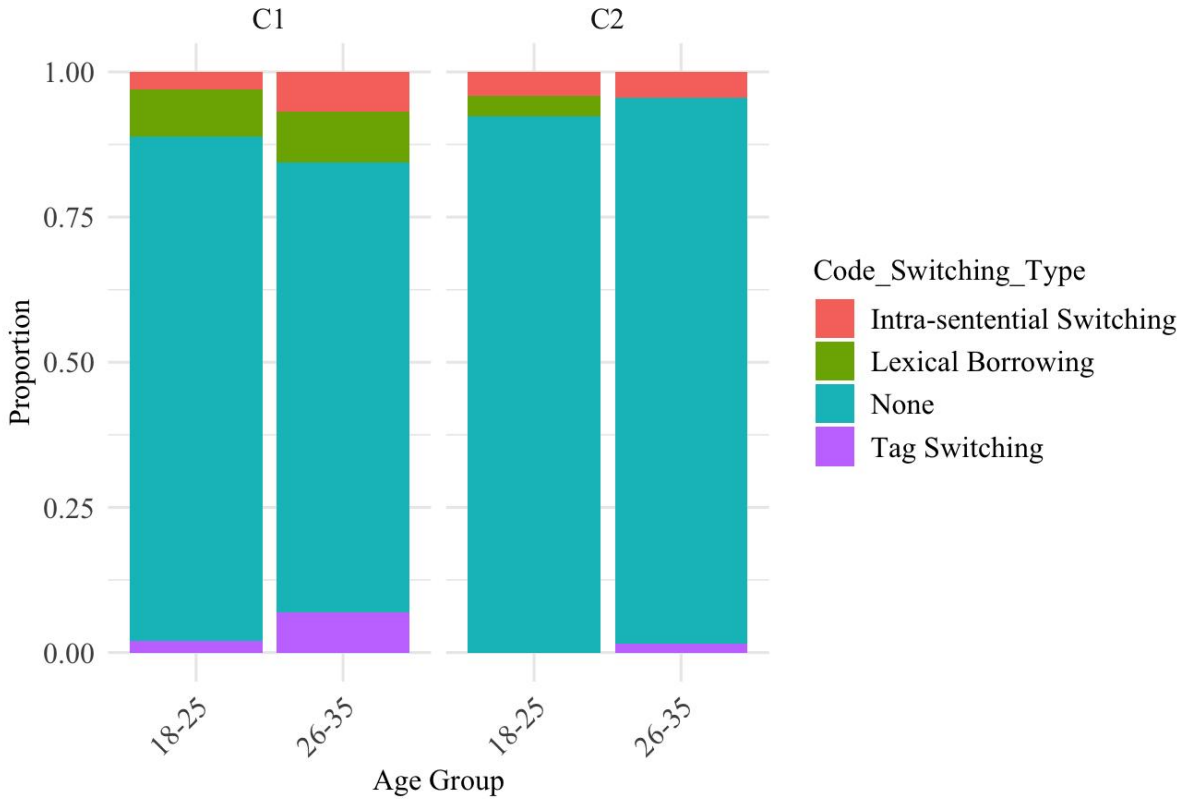


Chart 2: Proportional Code-switching Types by Age and Groups

Users aged 18–25 communicate almost entirely without code-switching ("None"), typically exceeding 90%, which supports H6: older users (26–35) switch codes more frequently. Among younger users, Tag Switching is slightly more common in Group C1 (work group) than C2 (friend group), suggesting that even when using Mandarin, Cantonese particles are retained. In contexts involving emotion, transitions, or quotations, sentence-level switching is more likely. The presence of Tag and Inter-sentential Switching in relaxed or intimate topics supports H4, confirming that code-switching patterns vary with topic type.

Question 10 of the questionnaire asked how users adjust language when interacting with different people. Most participants reported aligning their code with that of their interlocutor. In elder/authority interactions, 45% switch to dialect if the other does, 30% switch to Mandarin if the elder uses it, and 16% mix both. Only 9% always default to Mandarin. In contrast, during peer interactions, only 7% avoid switching, while 93% frequently use dialect or mix codes. These findings support H3, showing that switching reflects social identity and relational roles.

Code choice also shifts with context. In formal, serious-topic work groups, 74% adjust based on interlocutor background, while 26% use only Mandarin. In informal friend groups, 85% report frequent switching, and only 15% never switch. These results confirm H5 and further support H6–H7: serious topics are linked to Mandarin; lighter topics lead to more Cantonese switching.

In sum, corpus and survey data show that code-switching is more frequent in informal contexts. It is shaped by users' social roles and relationship dynamics (Group 2 hypotheses) and triggered by topic informality and social closeness (Group 3 hypotheses).

4.2.2 Pragmatic Functions of Code Usage

Chart 3's heatmap shows that pragmatic functions are closely related to code types and code-switching behaviors. The color intensity indicates how common each combination is.

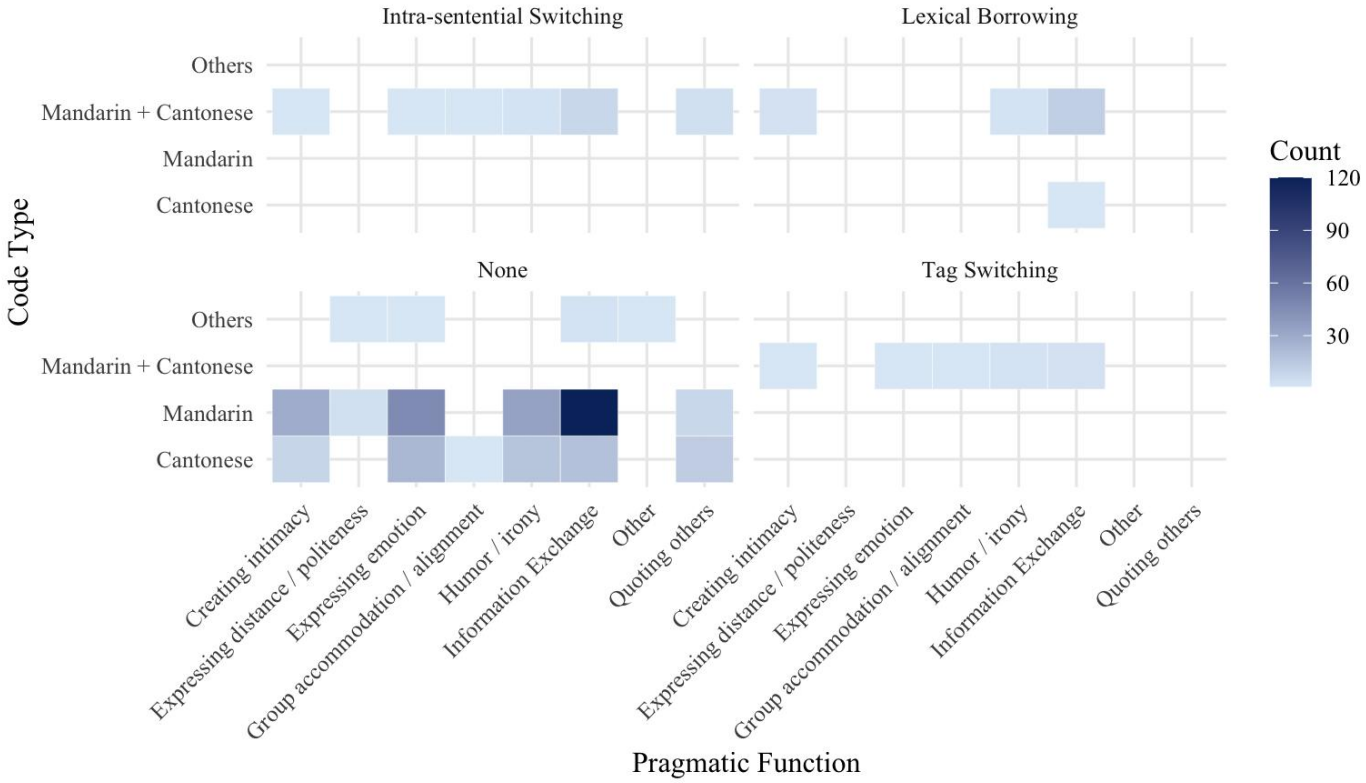


Chart 3: Pragmatic Functions by Code types and Code-switching

The darkest block is concentrated on "No switch + Mandarin + Information Exchange", indicating that in highly functional and formal communication, users overwhelmingly use pure Mandarin. This confirms H10 and H11, suggesting that Mandarin is more commonly used for giving commands and asking questions. Likewise, for the function "Expressing distance / politeness", Mandarin again shows high intensity, supporting H8 and H9, which propose that Mandarin is perceived as a formal and appropriate code in serious or institutional contexts. Furthermore, in the quertionnare section 3, Q17 and Q18 asked in which types of WeChat groups Mandarin seems more appropriate or respectful. The results are quite

concentrated: 73% selected categories “work-related groups,” “school-related groups,” and groups with members from different regions. In other words, in these formal or mix-background groups, sticking to Mandarin is considered more polite and solid. Only 15% of people said “I don’t think Mandarin is more appropriate/respectful,” indicating few oppose this view. This strongly supports that mandarin is perceived as a formal and appropriate language in serious contexts, reflecting its perceived role as an authoritative code (H8, H9), showing professionalism and authority (H15, H16).

In expressive functions like “Humor/Irony”, “Expressing Emotion”, and “Creating Intimacy”, we observe more instances of mixed codes (Mandarin + Cantonese), especially in Tag Switching. This supports H6 and H7, indicating that lighter topics are more likely to trigger code-switching. It also supports H12, as Cantonese is associated with emotional and intimate expression.

4.3 Identity Construction and Language Ideology

Chart 4 shows the distribution of identity construction categories associated with different code types in each age group. Overall, when using pure Mandarin compared to dialect or mixed codes, users more frequently construct identities related to “authority/distance/neutrality.” By contrast, using dialect or mixed codes (Mandarin+Cantonese) tends to express “closeness”, “family roles”, or “regional identity.”

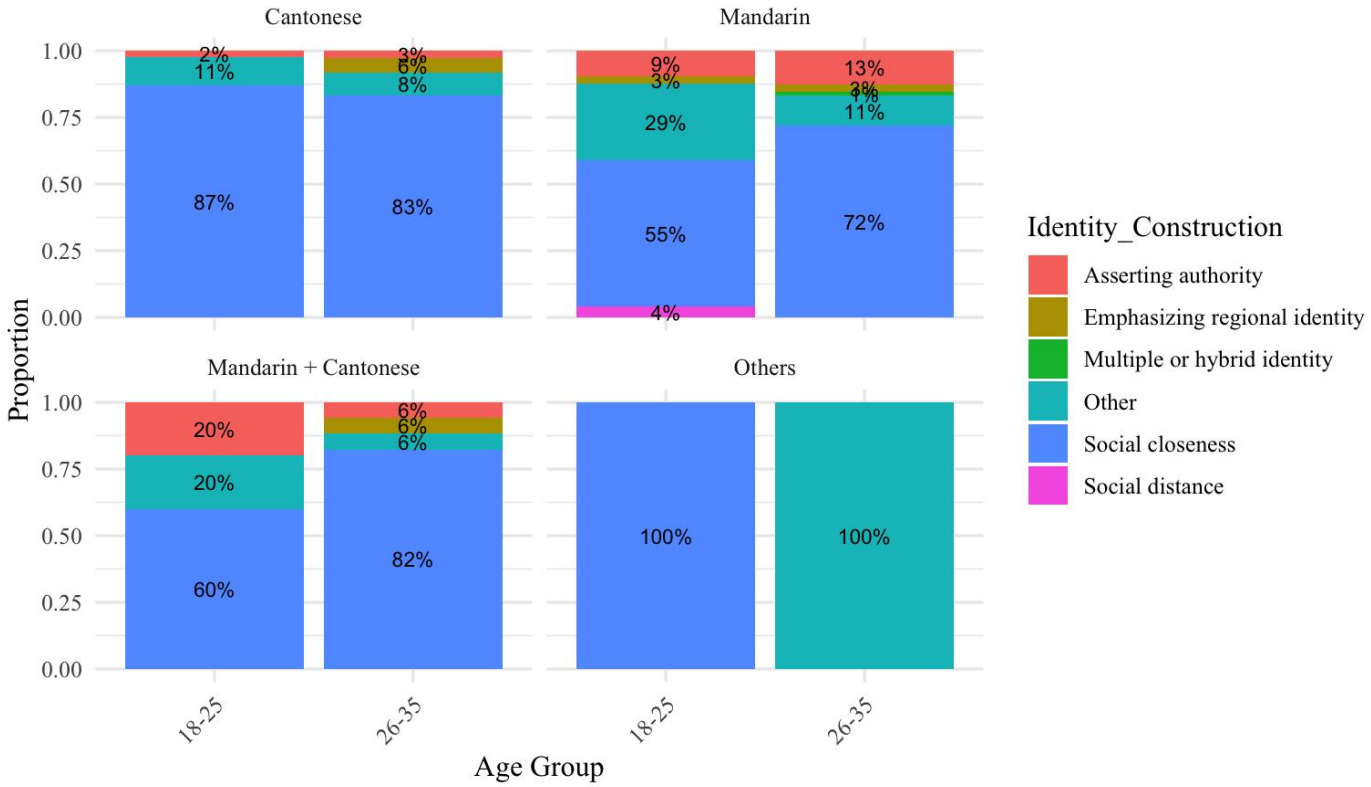


Chart 4: Identity Construction by Age and Code Type

For instance, in the Mandarin category, over 70% of utterances are linked to a neutral or distanced tone, and 13% (in the 26–35 age group) are linked to authority construction, supporting H16 that Mandarin indexes authority. By contrast, in the Cantonese category, over 80% of cases are associated with social closeness, while expressions of authority are nearly absent. This supports H14, as Cantonese clearly functions as a marker of group membership and regional belonging.

Chart 5 illustrates the distribution of responses across seven language ideology indicators, segmented by age groups.

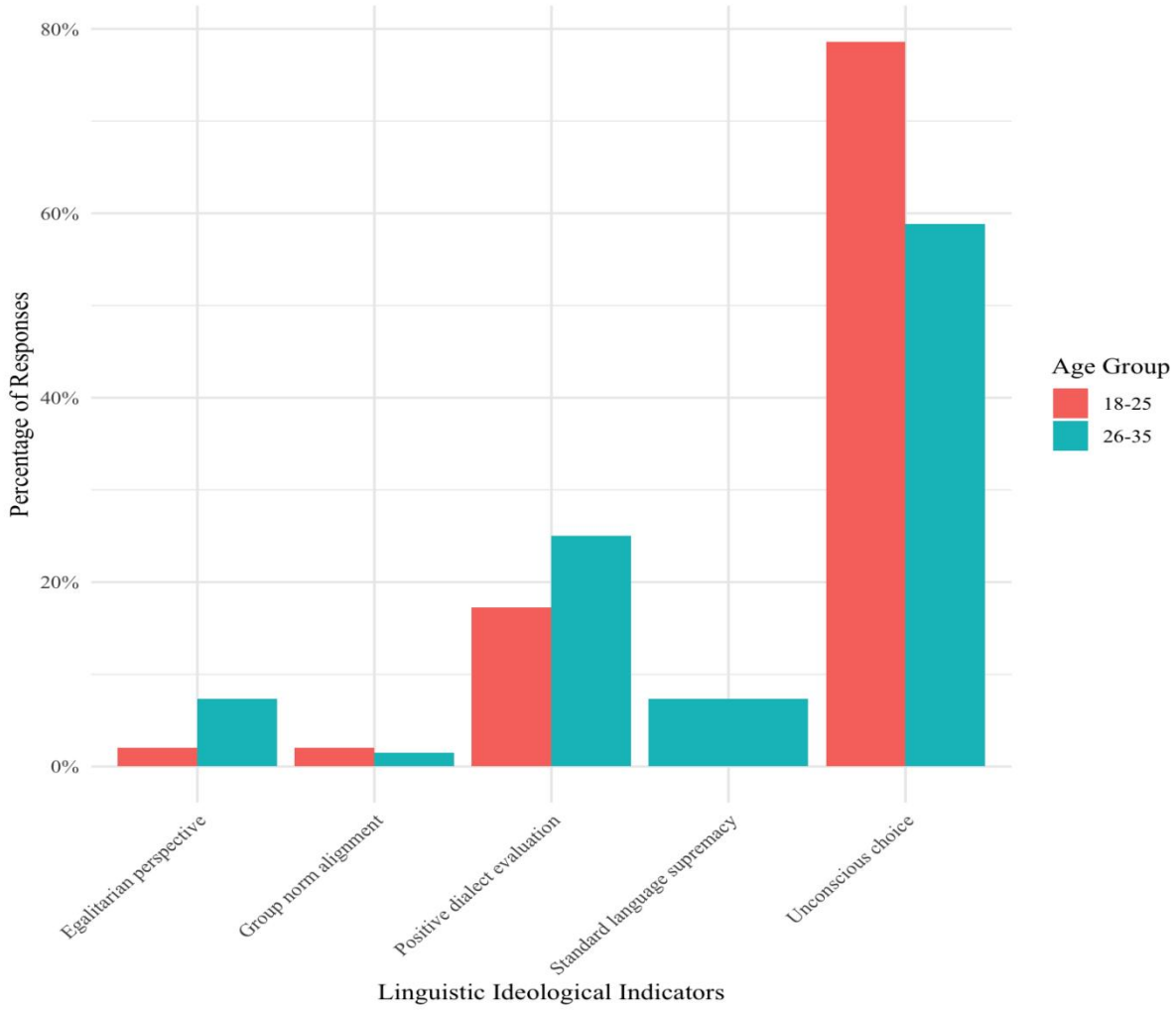


Chart 5: Ideologies by Age Group

Overall, a prominent trend emerges in the Cantonese context: those aged 18–25 and 26–35, exhibit a disproportionately high number of responses under the "Unconscious choice" category. Notably, the 18–25 age group contributes nearly 200 instances, accounting for the majority of the total responses in that setting. This suggests that, for younger users in Cantonese-speaking environments, the choice between Mandarin and Cantonese is often driven by routine habits rather than deliberate ideological stances. In other words, the use of Mandarin tends to be perceived as a “default” or “natural” option. Moreover, in the Cantonese context, responses under “Positive dialect evaluation” are noticeable. This indicates that although these users may predominantly rely on Mandarin in practice, they still perceive

Cantonese as emotionally valuable. This is also proved through the questionnare data. regarding emotional expressiveness, 46% respondents strongly agree and 38% respondents somewhat agree that dialects convey feelings better than Mandarin in online communication, while only 12% show neutral attitude and 4% select “Not sure”.

Our survey also explored social evaluation and dialect prejudice. In Q15, 57% of respondents were unsure whether dialects face discrimination; about 25% said “No,” while 43% named specific dialects, notably Cantonese (38%) and Henan dialect (16%). This awareness can shape language use—for instance, speakers may avoid revealing their dialect background in unfamiliar groups. In Q16, 56% mentioned dialects are mocked by “young people joking,” and 51% noted mockery “on social media.” These findings suggest that fear of negative judgment may discourage dialect use in formal settings.

Q19 and Q20 addressed whether Cantonese is seen as less “proper” or “educated” than Mandarin, especially in formal or written contexts. Responses showed moderate agreement: 20% answered “often,” and 22% “sometimes.” When combined with chat data, these results indicate that social evaluation pressures influence users’ language behavior. Even those who value dialects may avoid them in formal group chats to prevent misjudgment. This supports H13: users adapt code choice to manage social identity under perceived language ideologies.

5. Qualitative Analysis and Interpretive Insights

Based on the quantitative analysis, this chapter provides a qualitative interpretation of specific chat excerpts, exploring the interaction between code choice and social identity,

pragmatic function, and language ideology. In each section, we select representative chat excerpts from the WeChat groups (with English translations), and then conduct an in-depth discussion integrating code usage and speaker dimensions such as identity, authority, pragmatic functions.

5.1 Identity Construction

Language choice is an important means of expressing identity. Cantonese often serve as markers of local identity and intimacy, contrasting with Mandarin's official, formal role. In social group chats, speakers use dialect or Mandarin to convey closeness/distance, regional identity, and social role (such as kinship roles or group organizers).

Example 1: C1 Group Chat — Expressing emotions and group identity in Cantonese.

① C101: 晒死人*甘* ("It's so hot it could kill someone.")

② C102: *好晒* ("Super hot.")

③ C103: *落狗屎我都上班* ("Even if it's raining dog poop, I still have to work.")

④ C103: *翻公司* ("This damn company.")

This example is from the C1 Cantonese work group. Members are complaining about the hot weather, speaking entirely in Cantonese, which highlights the group's shared identity and emotional resonance. Line ① is a typical Cantonese exaggeration, where "甘" is Cantonese for "like this." This expression is both humorous and emotionally charged, serving to express emotion (humor/irony) and naturally indexing the speaker's Cantonese background, reinforcing a tacit atmosphere of closeness within the community. Line ② is brief and direct, also in Cantonese, expressing the feeling of heat. Through dialect it strengthens the sense of community and the authenticity of emotional expression ("social closeness" + "expressing

emotion"). Line ③ further uses a vulgar, exaggerated Cantonese saying to describe the nasty weather, expressing resignation with a humorous effect. This shows how Cantonese facilitates the expression of negative emotions in this context, making use of its rich local features to enhance emotional resonance and group identity. What the members express in Cantonese is not only their feeling about the weather, but also a "we're all in this together" sense of group identity.

Example 2: C2 Group Chat — Cantonese humor as a means of positive emotional expression.

① C202：太好笑了淦 ("That's so funny.")

② C202：粤语废话文学 ("Cantonese nonsense.")

③ C204：笑死我了 ("I'm dying of laughter.")

④ C206：*顶唔顺* ("Cannot handle it.")

In this lighthearted excerpt, the first few lines are in Mandarin and the final line is in Cantonese, highlighting the emotional amplification function of the dialect in humorous contexts. Lines ① and ③ use Mandarin to convey emotion, while line ④ "I can't handle it" (literal Cantonese "顶唔顺") is entirely in Cantonese. This Cantonese exclamation is both strongly expressive and phonetically compact,which emphasizes the intensity of emotion. Its use here is a classic example of emotional expression: the code choice is not for information transfer, but to amplify and release the positive emotion already generated by the preceding lines. Notably, line ② "Cantonese nonsense" is written in Mandarin but contains a playful self-mockery, implicitly referring to Cantonese style: the speaker is familiar with Cantonese features and uses a teasing tone to emphasize cultural belonging. This attitude reflects a

positive dialect-evaluation ideology: using ironic humor to show closeness rather than disparage. In this relatively informal and pleasant context, Mandarin clearly states the emotion, while Cantonese then intensifies it. This code-use strategy reflects Cantonese's strength in expressing positive emotions and community identity.

The excerpts above show that when WeChat group members want to emphasize intimacy and local belonging, they naturally use Cantonese. This context-driven code-switching not only reflects the identity structure within the group but also aligns with our hypotheses group 6 (Identity Indexicality throgh Code Choice), namely that local dialects help construct intimacy and regional identity.

5.2 Pragmatic Functions

Code choice and switching serve multiple pragmatic functions in communication, such as making information requests, venting emotions, showing politeness, or conveying humor. Code-switching is an interactional strategy used to signal closeness or distance. In our chat data, the use of Mandarin and Cantonese in different scenarios follows similar patterns: Mandarin is often used for information exchange and directive communication, while Cantonese is used to enhance humor, soften tone, or express emotion.

Example 3: C1 Group Chat — Cantonese particles softening tone in announcements.

① C107: 下午讲的宠物领养活动，*有无人有兴趣*~还有一个名额~ ("The pet adoption event I mentioned this afternoon—anyone interested? There is one spot left~")

② C107: *无就截葛啦~吾等20号啦~* ("If not, we'll close it off la~ We're not waiting until the 20th la~")

③ C102: 还有吗 (“Is there still one left?”)

④ C107: 没了 (“No more.”)

⑤ C102: 哦好吧 (“Oh, okay then.”)

⑥ C107: 有同事认领*左* (“A colleague already claimed it zuo.”)

This example is from the C1 work group and involves announcing and responding to an event signup. Although the topic is work-related, the tone is light, and embedding Cantonese particles makes the exchange friendly. The sign “~” is used in Chinese as a mark of a kind of joyous or playful, singing tone. Line ① is constructed as a Cantonese sentence structure. Line ② is entirely in Cantonese; the particle “啦” (la) is a typical Cantonese expressive particle that softens the imperative (in this context, meaning “we’ll close it off”). This indicates that even in a work context, the speaker is willing to use dialect to build rapport and soften the announcement’s formality. In the replies (lines ③–⑤), members respond briefly in Cantonese and Mandarin. Notably, line ④ returns to Mandarin when confirming information, contrasting with line ②’s Cantonese. Overall, this example demonstrates a combination of Mandarin as the neutral communication code and Cantonese particles softening tone. Speaker C107 uses Mandarin as the main medium for conveying information, supplemented by Cantonese tagging (“la”, “zuo”) to create intimacy, thus keeping a relaxed atmosphere even while posting a work notice.

Example 4: C2 Group Chat — Tone negotiation in a request and buffering with Cantonese.

① C206: C205 你能不能再去回一句 (“C205, could you reply to him once more?”)

② C205: 我不知道回他什么 (“I don’t know what to reply to him.”)

③ C205: 真的 (“Really.”)

④ C202: 然后他再回你十几句 (“Then he’ll reply to you with another dozen messages.”)

⑤ C205: 一会说那一会说这的 (“One moment he says this, the next he says that.”)

⑥ C201: 你发个表情包 (“Just send him an emoji.”)

⑦ C202: 捡点垃圾回一下 (“At least send some silly reply.”)

⑧ C201: ***唔使***回什么实质性的东西 (“No need to reply with anything substantial.”)

In this complex exchange, group members ask C205 to send a reply to someone, but C205 is reluctant. Each person employs different strategies to soften pressure and maintain group harmony, with code choice playing a key role in modulating tone and relationships. In line ①, C206 makes a suggestive request in polite Mandarin. This shows politeness among acquaintances on one hand, and Mandarin’s role as the medium for explicit guidance on the other. Lines ②–③ are Mandarin responses showing clear avoidance and unwillingness (“social distance”), functioning to give a feedback and polite refusal. Then other members intervene: lines ④ and ⑥–⑦, still in Mandarin, use humor and suggestions to ease the tension. They form a kind of soft command strategy: on the surface they sound like suggestions, but in effect they are gentle requests intended to nudge C205 to comply without applying strong pressure. Finally, line ⑧ is a classic insertion of Cantonese “唔使” (literally “don’t need to”). This Cantonese phrasing is more colloquial and tactful than the Mandarin equivalent “不用回复什么实质内容” (“no need to reply with anything substantial”), making

the negation less direct. The result is that the overall request appears casual, yet effectively encourages C205 to comply. This exemplifies a key pragmatic function of dialect in everyday digital interaction: enhancing intimacy, softening commands, and maintaining group cohesion.

The above analyses of different contexts spporte H10 and H11 that Mandarin is used to ensure communication efficiency and authority in secondary relationships, for example, in information requests or commands; whereas dialectal elements are used in primary relationships to increase social closeness or reduce tension, for example, in humorous teasing or softening requests.

5.3 Language Ideologies

Language ideologies refer to people's beliefs and attitudes about the values of different languages or dialects. In our data, the asymmetric use of Mandarin and local dialects often implies judgments about language status. Participants' reliance on Mandarin often reflects their ideological alignment with its perceived authority, correctness, and appropriateness. The following excerpts show how such ideologies are manifested in natural conversation:

Example 5: C2 Group Chat — "Teaching Cantonese to someone who doesn't speak it": Mandarin as 'teaching authority'.

① C206: 好恨我我不会粤语啊 ("I really wish I could speak Cantonese.")

② C208: 今天 ("Today.")

③ C208: 七仔老师教你一个四字成语。 ("Professor Qizai will teach you a four-character idiom.")

④ C208: ***Lai Yat Fong Cheung 来日方长***。 (“There are many days left to do it.”)

⑤ C206: 哈哈哈哈哈哈哈 (“Hahahahaha.”)

⑥ C208: 你跟老师读一遍。 (“Repeat after the teacher.”)

This example presents a lighthearted Cantonese lesson interaction where language choice highlights different ideological functions of Mandarin and Cantonese. Line ① is in Mandarin text, but it conveys the speaker’s positive attitude toward Cantonese and desire for local identity (“emphasizing regional identity + positive dialect evaluation”). However, from line ② onward, C208 uses only Mandarin to play the role of a Cantonese teacher. He gives directives in Mandarin as if teaching in a classroom. Even though he’s teaching a Cantonese idiom “Lai Yat Fong Cheung”, his explanation and organization rely on Mandarin. This situation shows that even when the content is dialect (the idiom), the entire teaching process is presented through Mandarin as the “authoritative discourse,” highlighting participants’ acceptance of Mandarin’s institutional status. This is exactly an example of “standard language supremacy” in an informal context: the dialect is treated as the “object” to be taught, while Mandarin is the default medium for instruction. Although members demonstrate interest in Cantonese, their accepted this mode of teaching is through Mandarin. This again reflects language ideologies: dialect is seen positively as a symbol of identity (“positive dialect evaluation”), but during its transmission and learning, Mandarin maintains the core role.

Through qualitative analysis of these fiver representative excerpts, we find that in WeChat group chats, code choice is jointly influenced by identity strategies, pragmatic purposes, and ideological beliefs. The use of Cantonese often marks intimacy and regional identity, whereas Mandarin carries out information delivery and authoritative expression.

Group members flexibly switch codes according to communicative context: in situations involving kinship, humor, or sentiment, they insert dialectal words to reinforce closeness and emotional resonance; when needing to express clearly, give advice, or resolve misunderstandings, they use Mandarin to convey rigor and credibility. At the same time, the dialogue practices reflect participants' implicit views of language status: Mandarin is assumed to be a neutral and standard code, while dialects convey affect and identity. In summary, code choice in group chat is not only an efficient communication strategy but also an expression of identity and ideology.

6. Discussion

6.1 Language Choice, Identity, and Attitudes in WeChat Groups

This study, using WeChat groups as the digital context, investigates how language use, specifically code choice and switching between Mandarin and Cantonese, participates in identity construction, interaction negotiation, and ideology reproduction. The data come from two actual WeChat groups (370 chat entries) and 74 questionnaires, forming a convergent quantitative and qualitative research design. The data reveal that Mandarin consistently dominates in all four groups, but different group types show clear differences. In the work group (C1), Mandarin is almost the sole choice; and in the friends group (C2), there is more mixed use of Cantonese and Mandarin. This indicates that dialect use is more common in contexts with closer interpersonal relationships, whereas in situations that require maintaining a formal, professional image, Mandarin becomes the default option.

Language choice is influenced not only by group type but also by interaction goals. Mandarin is frequently used when posting notices, making requests, or confirming information that require clear communication, reflecting its “institutional authority” role in the digital context. Dialects are used more often in situations of emotional expression, joking, or kinship interaction, serving to build familiarity and intimacy. Code-switching often serves shifts in pragmatic function: for example, switching from a command to an emotional comfort is often accompanied by a switch from Mandarin to dialect. In terms of identity construction, some members use different codes with different interlocutors. This language adaptation reflects both social role awareness and a strategic expression of identity in context. In the Cantonese groups, younger members emphasize regional belonging by inserting Cantonese words into Mandarin speech while avoiding excessive formality, thus creating a more relaxed interaction.

6.2 Mandarin Dominance and Coexistence of Codes

Although dialects are actively used in daily interactions across the four groups, the overall data reveal that Mandarin holds a dominant position. In the questionnaire, participants generally regard Mandarin as a formal and rational communication tool, whereas dialects are more often associated with being warm authentic, or fun. Through the survey, we also found that dialects are frequently given negative labels in society. Some people, fearing negative evaluation, consciously reduce their use of dialect.

WeChat, as a major social platform, has a strong tendency toward textual communication. This CMC environment weakens the emotional cues and social signals present in face-to-face conversation, making language choice rely more heavily on clarity in writing. The advantages

of Mandarin, which are being clear, standard, and unambiguous, are further amplified in this context, whereas dialects must appear in marked ways (for example, by inserting dialectal vocabulary, particles, or syntactic patterns into Mandarin sentences). This strategic switching and embedding reflect how the digital context regulates language practice and illustrates the choices users make between identity, emotion, and communicative efficiency.

7. Conclusion and Future Research Directions

7.1 Summary of Findings

This study examined how language use in WeChat groups, specifically code choice and code-switching between Mandarin and dialects, participates in identity construction, interaction negotiation, and ideology reproduction. The findings are as follows: Mandarin dominates most interactions, but Cantonese have irreplaceable social and emotional functions in certain contexts. Code-switching is highly functional and strategic, and language choice is often embedded in group identity, relationships, and context types. Through the triangulated model, we tested each of the 16 proposed hypotheses:

Hypothesis Group	Sections	Result
Group 1: Code Distribution across Group Types	4.1	Partially supported: Mandarin dominates across all groups, with slight variation.
Group 2: Code-Switching and Identity Variables	4.2.1	Supported: Age and identity correlate with switch types.
Group 3: Topic Sensitivity and Code-Mixing	4.2.1, 4.2.2	Partially supported: Formal/informal contexts influence switching, but not always.
Group 4: Language Ideologies Reflected in Attitudes	4.2.2, 5.3	Supported: Clear ideological alignment by age and region.
Group 5: Functional Tendencies in Language Use	4.2.2, 5.2	Supported: Specific code types serve pragmatic functions.
Group 6: Identity Indexicality through Code Choice	4.2.2, 4.3, 5.1	Supported: Code choice reflects identity, especially in dialect use.

Table 2: Hypothesis Groups and Results

7.2 Limitations and Future Directions

At the theoretical level, this study expands understanding of identity negotiation mechanisms in digital contexts and enriches how language ideologies manifest in real interactions. By triangulating corpus and survey data, we validated multiple hypotheses about the relationships between code types, pragmatic functions, and identity variables. The study suggests that writing dialect online is not only a language practice but also an expression of cultural attitudes and social identity. On the practical side, the research reveals consistent patterns in language use across different social contexts in digital communication. The results can inform multilingual community management and strategies for incorporating dialects in educational contexts, and may help promote dialect preservation starting from textual practice.

This study has several limitations. First, focusing on four specific WeChat groups means the sample size is limited, which may not fully represent broader community differences, especially in terms of geographic scope and social structures. Second, the data are primarily text-based. Although we constructed a detailed coding scheme, due to space constraints we did not code more granular information such as the direction of code-switching. This somewhat limits our ability to capture more micro-level aspects of language interaction. Additionally, all data collection and analysis were conducted remotely from Barcelona. This geographical distance added a layer of difficulty in accessing participants and managing contextual interpretation, but also demonstrates the feasibility and relevance of conducting digital ethnography across locations.

Future research could expand the data scale to include more diverse dialect regions and group types, particularly cross-regional and cross-generational interactions, and even groups where multiple dialects coexist. As the original research also included two fully analyzed Sichuanese WeChat groups (one family group and one friend group), future studies could revisit this data to enable cross-dialectal comparison between Cantonese and Sichuanese users in digital communication. Additionally, incorporating spoken data analysis or participant observation could enrich a multi-modal perspective of language interaction. Such extensions would further deepen our understanding of multilingual communication in digital communities.

Bibliography

Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of sociolinguistics, 10*(4), 419-438.

Auer, P. (Ed.). (2013). *Code-switching in conversation: Language, interaction and identity.* Routledge.

Bourdieu, P. (1991). *Language and symbolic power.* Harvard University Press.

Bucholtz, M., & Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse studies, 7*(4-5), 585-614.

Chen, X. R. (2013). Pragmatic identity: Dynamic selection and discursive construction. *Foreign Language Research*, (4), 27–32.

Ferguson, C. A. (1959). Diglossia. *word*, *15*(2), 325-340.

Gal, S. (2023, June 21). Language Ideologies. *Oxford Research Encyclopedia of Linguistics.* Retrieved 11 Jun. 2025.

Gumperz JJ. Conversational code switching. In: *Discourse Strategies*. Studies in Interactional Sociolinguistics. Cambridge University Press; 1982:59-99.

Julià i Muné, J. (2019). *The Standardization of Chinese and Catalan: Two Eurasian parallel processes (International edition)*. Edicions de la Universitat de Lleida.

Kircher, R., & Zipp, L. (2021). *An introduction to language attitudes research*. Routledge.

Liu, J. (2011). Deviant writing and youth identity: Representation of dialects with Chinese characters on the internet. *Chinese Language and Discourse*, 2(1), 58-79.

O'Rourke, B. (2022). Researching language attitudes in multilingual communities. *Research methods in language attitudes*, 271-281

Siebenhaar, B. (2006). Code choice and code-switching in Swiss-German Internet Relay Chat rooms. *Journal of sociolinguistics*, 10(4), 481-506.

Sun, L. (2018). Returning to tribalization: How migrant youth sustain and redefine dialect communities in the new media era. *Modern Communication (Journal of Communication University of China)*, (7).

Wang, Y., Chen, J., & Zhang, J. J. (2012). Dialect and Putonghua coexistence: A review of diglossia research in psychology. *Advances in Psychological Science, 20*(8), 1243–1250.

Woolard, K. A. (2016). *Singular and plural: Ideologies of linguistic authority in 21st century Catalonia*. Oxford University Press.

Yang, C. (2007). Chinese Internet language: A sociolinguistic analysis of adaptations of the Chinese writing system. *Language@ Internet, 4*.

Yus, F. (2011). *Cyberpragmatics: Internet-mediated communication in context*. John Benjamins Publishing Company.

Yus, F. (2017). Contextual constraints and non-propositional effects in WhatsApp communication. *Journal of pragmatics*, 114, 66-86.

Yus, F. (2018). Identity-related issues in meme communication. *Internet Pragmatics*, 1(1), 113-133.

Zhou, Y. K., & Liu, H. T. (2024). The relationship between dialects and the common national code in China's modernization: With remarks on the role and function of virtual speech communities. *Journal of Zhejiang University—Humanities and Social Sciences, 54*(8), 124–137.

Appendices

Appendix A: Language Atlas of China (2012 edition)

Appendix B: Coding Examples

Group	ID	Message (Original)	Message (English)	Age Group	Language Background	Text Type	Code Type	Code-Switching Type	Pragmatic Function	Identity Construction	Linguistic Ideological Indicators
C1	01	晒死人甘	It's so hot it could kill someone.	18-25	Mandarin + Cantonese	Text	Cantonese	None	Humor / irony	Social closeness	Unconscious choice
C1	02	好晒	Super hot.	26-35	Mandarin + Cantonese	Text	Cantonese	None	Expressing emotion	Social closeness	Unconscious choice
C1	03	落狗屎我都上班	Even it's raining dog poop, I still have to go to work.	18-25	Mandarin + Cantonese	Text	Cantonese	None	Humor / irony	Social closeness	Positive dialect evaluation
C1	03	翻公司	This damn company.	18-25	Mandarin + Cantonese	Text	Cantonese	None	Expressing emotion	Social closeness	Positive dialect evaluation
C1	01	@开摆 公司救过你咩	@C103 Didn't the company save you before?	18-25	Mandarin + Cantonese	Text	Mandarin + Cantonese	Tag Switching	Humor / irony	Social closeness	Unconscious choice
C1	05	建议大家必要时在家办公，珍爱生命	Everyone, I suggest working from home if necessary. Life is precious.	26-35	Mandarin + Cantonese	Text	Mandarin	None	Humor / irony	Social closeness	Unconscious choice
C1	01	你要翻嚟救公司	You need to come back and save the company!	18-25	Mandarin + Cantonese	Text	Cantonese	None	Humor / irony	Social closeness	Positive dialect evaluation

Appendix C: Questionnaire

Questionnaire

Section 1: Demographic Information

1. Gender: ☐ Male ☐ Female ☐ Other ☐ Prefer not to say

2. Age group: ☐ 18–25 ☐ 26–35 ☐ 36–50 ☐ Over 50

3. Place of birth: _____

4. Current residence: _____

5. Highest Education level: ☐ Middle school or below ☐ High school ☐ Associate degree ☐ Bachelor's ☐ Postgraduate or above

6. Employment status: ☐ Student ☐ Employed ☐ Self-employed ☐ Homemaker ☐ Retired ☐ Other

7. (The following questions will appear depending on this question)Self-assessed language proficiency (Native/ Fluent / Intermediate / Basic / None):

- Mandarin: _____ - Cantonese: _____ - Sichuanese: _____

Section 2: Language Use in WeChat

8a. (For people who speaks Cantonese) In what contexts do you usually use the following languages? (Select all that apply)

- Mandarin: ☐ Home ☐ Friends ☐ Work ☐ School ☐ Online ☐ Other: _____

- Cantonese: ☐ Home ☐ Friends ☐ Work ☐ School ☐ Online ☐ Other: _____

8b. (For people who speaks Sichuanes) In what contexts do you usually use the following languages? (Select all that apply)

- Mandarin: ☐ Home ☐ Friends ☐ Work ☐ School ☐ Online ☐ Other: _____

- Sichuanese: ☐ Home ☐ Friends ☐ Work ☐ School ☐ Online ☐ Other: _____

9a. (For people who speaks Cantonese) When typing in WeChat group chats, which language(s) do you use most often? Please specify:

☐ Mainly Mandarin

☐ Mainly Cantonese

☐ Mandarin and Cantonese

☐ Other dialect (please specify): _____

□ Depends on the group (Please specify: e.g., use Mandarin in work group, Cantonese in friend group...) → _____

9a. (For people who speaks Sichuanese) When typing in WeChat group chats, which language(s) do you use most often? Please specify:

□ Mainly Mandarin

□ Mainly Sichuanese

□ Mandarin and Sichuanese

□ Other dialect (please specify): _____

□ Depends on the group (Please specify: e.g., use Mandarin in work group, Cantonese in friend group...) → _____

The following questions aim to understand whether and how you adjust your language use in WeChat group chats depending on social situations or group members. Please select the option that best describes your typical behavior.

Q1 1. (For people who speaks Cantonese) When interacting with elders or people with authority (e.g., parents, teachers, supervisors), how do you usually adjust your language use?

□ If they speak Mandarin, I switch from Cantonese to Mandarin

□ If they speak Cantonese, I switch from Mandarin to Cantonese

□ Although they speak Cantonese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Cantonese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Cantonese

□ If they speak Cantonese, I use a mix of Mandarin and Cantonese

□ Other (please specify): _____

Q1 2. (For Sichuanese speakers) When interacting with elders or people with authority (e.g., parents, teachers, supervisors), how do you usually adjust your language use?

□ If they speak Mandarin, I switch from Sichuanese to Mandarin

□ If they speak Sichuanese, I switch from Mandarin to Sichuanese

□ Although they speak Sichuanese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Sichuanese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Sichuanese

□ If they speak Sichuanese, I use a mix of Mandarin and Sichuanese

□ Other (please specify): _____

Q2 3. (For people who speaks Cantonese) When chatting with peers or close friends:

□ If they speak Mandarin, I switch from Cantonese to Mandarin

□ If they speak Cantonese, I switch from Mandarin to Cantonese

□ Although they speak Cantonese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Cantonese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Cantonese

□ If they speak Cantonese, I use a mix of Mandarin and Cantonese

□ Other (please specify): _____

Q2 4. (For people who speaks Sichuanese) When chatting with peers or close friends:

□ If they speak Mandarin, I switch from Sichuanese to Mandarin

□ If they speak Sichuanese, I switch from Mandarin to Sichuanese

□ Although they speak Sichuanese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Sichuanese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Sichuanese

□ If they speak Sichuanese, I use a mix of Mandarin and Sichuanese

□ Other (please specify): _____

Q3 5. (For people who speaks Cantonese) When some group members do not understand dialects:

□ If they speak Mandarin, I switch from Cantonese to Mandarin

□ If they speak Cantonese, I switch from Mandarin to Cantonese

□ Although they speak Cantonese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Cantonese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Cantonese

□ If they speak Cantonese, I use a mix of Mandarin and Cantonese

□ Other (please specify): _____

Q3 6. (For people who speaks Sichuanese) When some group members do not understand dialects:

□ If they speak Mandarin, I switch from Sichuanese to Mandarin

□ If they speak Sichuanese, I switch from Mandarin to Sichuanese

□ Although they speak Sichuanese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Sichuanese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Sichuanese

□ If they speak Sichuanese, I use a mix of Mandarin and Sichuanese

□ Other (please specify): _____

Q4 7. (For people who speaks Cantonese) When some group members have limited Mandarin speaking ability but understand and tend to use dialects more, do you switch to dialects to accommodate them?

□ Switch from Mandarin to Cantonese

□ Keep using Mandarin, do not switch

□ Mix of Mandarin and Cantonese

□ Other (please specify): _____

Q4 8. (For people who speaks Sichuanese) When some group members have limited Mandarin speaking ability but understand and tend to use dialects more, do you switch to

dialects to accommodate them?

☐ Switch from Mandarin to Sichuanes

☐ Keep using Mandarin, do not switch

☐ Mix of Mandarin and Sichuanese

☐ Other (please specify): _____

Q5 9. (For people who speaks Cantonese) In formal groups (e.g., work or school-related WeChat groups):

☐ If they speak Mandarin, I switch from Cantonese to Mandarin

☐ If they speak Cantonese, I switch from Mandarin to Cantonese

☐ Although they speak Cantonese, I continue using Mandarin without switching

☐ Although they speak Mandarin, I continue using Cantonese without switching

☐ If they speak Mandarin, I use a mix of Mandarin and Cantonese

☐ If they speak Cantonese, I use a mix of Mandarin and Cantonese

☐ Other (please specify): _____

Q5 10. (For people who speaks Sichuanese) In formal groups (e.g., work or school-related WeChat groups):

☐ If they speak Mandarin, I switch from Sichuanese to Mandarin

☐ If they speak Sichuanese, I switch from Mandarin to Sichuanese

☐ Although they speak Sichuanese, I continue using Mandarin without switching

☐ Although they speak Mandarin, I continue using Sichuanese without switching

☐ If they speak Mandarin, I use a mix of Mandarin and Sichuanese

☐ If they speak Sichuanese, I use a mix of Mandarin and Sichuanese

☐ Other (please specify): _____

Q6 11. (For people who speaks Cantonese) In informal groups (e.g., close family or friend

groups):

□ If they speak Mandarin, I switch from Cantonese to Mandarin

□ If they speak Cantonese, I switch from Mandarin to Cantonese

□ Although they speak Cantonese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Cantonese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Cantonese

□ If they speak Cantonese, I use a mix of Mandarin and Cantonese

□ Other (please specify): _____

Q6 12. (For people who speaks Sichuanese) In informal groups (e.g., close family or friend groups):

□ If they speak Mandarin, I switch from Sichuanese to Mandarin

□ If they speak Sichuanese, I switch from Mandarin to Sichuanese

□ Although they speak Sichuanese, I continue using Mandarin without switching

□ Although they speak Mandarin, I continue using Sichuanese without switching

□ If they speak Mandarin, I use a mix of Mandarin and Sichuanese

□ If they speak Sichuanese, I use a mix of Mandarin and Sichuanese

□ Other (please specify): _____

11. In which of the following situations do you tend to use dialects more often in WeChat group chats? (You may select more than one)

□ When joking or making humorous comments

□ When expressing strong emotions (e.g., anger, excitement, affection)

□ When chatting with close friends

□ When talking with older family members

□ In informal groups or casual conversations

☐ To quote someone else's speech

☐ When you want to sound relaxed or personal

☐ I don't use dialects more often in any specific situation

☐ Other (please specify): _____

Section 3: Language Attitudes and Ideologies

12. Do you enjoy it when others use dialects to speak with you in WeChat group chats?

☐ Yes, always

☐ Yes, especially with close friends

☐ Yes, especially with older family members

☐ Yes, in informal or casual groups

☐ Neutral / I don't mind

☐ No, I prefer Mandarin in group chats

☐ Other (please specify): _____

13. Do you think dialects are more expressive or emotionally rich than Mandarin in digital communication?

☐ Strongly agree

☐ Somewhat agree

☐ Neutral

☐ Somewhat disagree

☐ Strongly disagree

☐ Not sure

14a. (For people who speaks Cantonese) Do you think Cantonese is appropriate in formal/public WeChat groups (e.g., work groups, school, etc)?

☐ Appropriate ☐ Inappropriate ☐ Depends (Please specify: _____)

14b. (For people who speaks Sichuanese) Do you think Sichuanese is appropriate in formal/public WeChat groups (e.g., work groups, school, etc)?

☐ Appropriate ☐ Inappropriate ☐ Depends (Please specify: _____)

15. Do you think people with certain dialect accents are discriminated against in China?

☐ Yes ☐ No ☐ Not sure. If yes, which dialects? _____

16. In which social situations do you think dialect is sometimes mocked, imitated, or laughed at? (You can select more than one option)

a. For people who speaks Cantonese

☐ In formal settings (e.g., workplace, public speaking)

☐ On social media (e.g., WeChat Moments, Douyin, memes)

☐ Among young people joking around

☐ In educational settings (e.g., classroom, university chat groups)

☐ In mixed-dialect group chats

☐ Never noticed

☐ Other: _____

b. For people who speaks Sichuanese

☐ In formal settings (e.g., workplace, public speaking)

☐ On social media (e.g., WeChat Moments, Douyin, memes)

☐ Among young people joking around

☐ In educational settings (e.g., classroom, university chat groups)

☐ In mixed-dialect group chats

☐ Never noticed

☐ Other: _____

17. In which WeChat groups do you think Mandarin is more appropriate than dialects? (You may select more than one)

□ Work-related groups

□ School-related groups

□ Extended family groups

□ Mixed-background friend groups

□ Close friend groups

□ I don't think Mandarin is more appropriate

□ Not sure

18. In which WeChat groups do you think Mandarin is more respectful than dialects? (You may select more than one)

□ Work-related groups

□ School-related groups

□ Extended family groups

□ Mixed-background friend groups

□ Close friend groups

□ I don't think Mandarin is more respectful

□ Not sure

19. (For people who speaks Cantonese) Do you think Cantonese is ever perceived as less "proper" or "educated" than Mandarin, especially in writing or formal WeChat groups?

□ Yes, often

□ Sometimes

□ No

□ Not sure

If yes, why do you think this perception exists? (Optional short answer): _____

20. (For people who speaks Sichuanese) Do you think Sichuanese is ever perceived as less "proper" or "educated" than Mandarin, especially in writing or formal WeChat groups?

□ Yes, often

□ Sometimes

□ No

□ Not sure

If yes, why do you think this perception exists? (Optional short answer): _____

21. Do you avoid using dialects in WeChat because others may not understand?

a. For people who speaks Cantonese: □ Often □ Sometimes □ Never

b. For people who speaks Sichuanese: □ Often □ Sometimes □ Never

Section 4: Open-ended Questions (Optional)

23. Have you ever had an experience in which using Mandarin or a dialect led to misunderstanding or judgment in a WeChat group?

Please describe (Mandarin) : _____

Please describe (Cantonese) : _____

Please describe (Sichuanese) : _____

Appendix D: Screenshot of Online Consent Form

关于普通话和方言的调查

问卷说明

亲爱的参与者：
您好！
您被邀请参与一项关于在微信群聊中使用普通话与方言的社会语言学研究。本研究由涂晶晶开展，作为其在庞培法布拉大学（Universitat Pompeu Fabra）“理论与应用语言学硕士项目”的一部分。

在决定是否参与之前，请仔细阅读以下信息。如您有任何疑问，欢迎随时联系研究者（邮箱：jingjing.tu01@estudiant.upf.edu）。请根据自己的意愿决定是否参与，本调查完全自愿，您可在任何阶段自由退出，无需承担任何后果。

研究目的：
本研究旨在探讨普通话与方言（粤语/四川话）使用者在微信群聊中的语言选择行为，以及这些选择如何体现社交身份与权力关系等社会语言学特征。研究致力于深入理解数字语境下的多语实践及其社会意义，属于社会语言学、计算机中介传播（CMC）与语言意识形态研究范畴的一部分。

调查方式：
如果您同意参与，将需要填写一份简短的在线问卷（预计耗时约5–10分钟），内容涉及您的语言背景及您在微信群中的语言使用习惯。**所有问卷均为匿名填写，仅用于学术研究。**

开始作答

关于普通话和方言的调查

保密原则：
本研究将**严格遵守保密与匿名原则**。问卷**不会收集任何身份识别信息**（如姓名、电话号码、IP地址等）。所有数据将在庞培法布拉大学信息部安全保存，仅供研究者本人用于分析与撰写学术论文，研究结果将以整体形式呈现，不会识别到任何具体参与者。研究结束后，数据将被销毁。

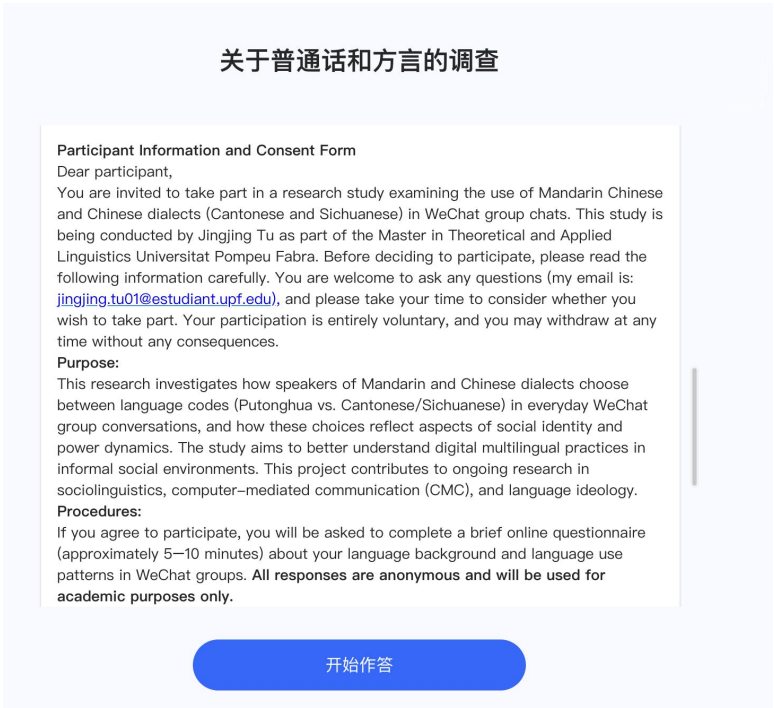
风险与收益：
本研究不会对您造成任何已知风险。您不会获得直接利益，但您的参与将有助于我们更好地理解中文数字交流中的社会语言现象。

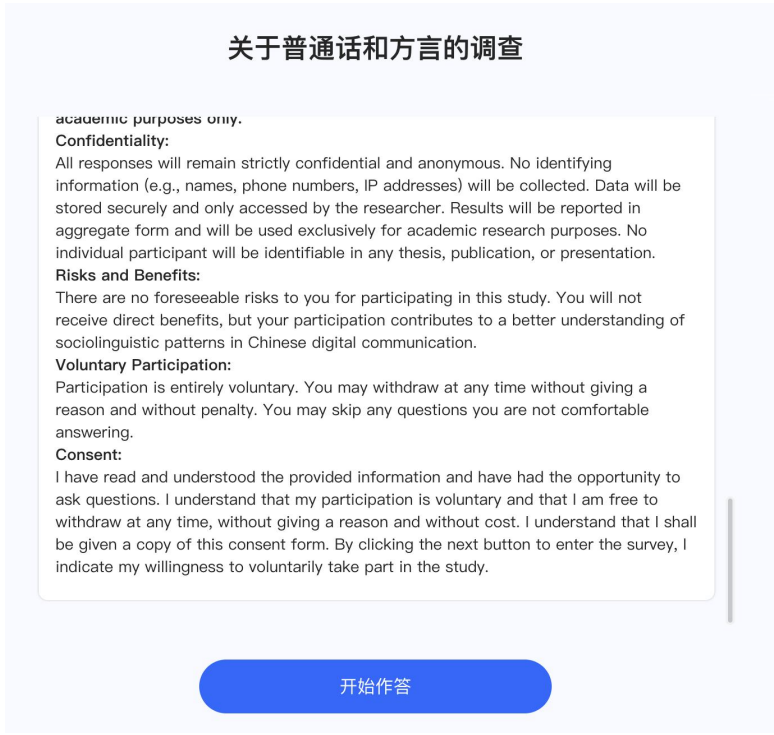
自愿参与：
参与本研究完全出于自愿。您可在任何时候选择退出，无需说明理由，也不会产生任何不良影响。

知情同意：
我已阅读并理解以上所提供的信息，并已获知可联系研究者进行咨询。我明白参与本研究完全出于自愿，我可随时退出，不需说明理由，也不承担任何费用。

点击“开始作答”进入问卷，代表“我同意自愿参与本研究”，并进行授权。

开始作答


关于普通话和方言的调查
Participant Information and Consent Form
Dear participant,
You are invited to take part in a research study examining the use of Mandarin Chinese and Chinese dialects (Cantonese and Sichuanese) in WeChat group chats. This study is being conducted by Jingjing Tu as part of the Master in Theoretical and Applied Linguistics Universitat Pompeu Fabra. Before deciding to participate, please read the following information carefully. You are welcome to ask any questions (my email is: jingjing.tu01@estudiant.upf.edu), and please take your time to consider whether you wish to take part. Your participation is entirely voluntary, and you may withdraw at any time without any consequences.
Purpose:
This research investigates how speakers of Mandarin and Chinese dialects choose between language codes (Putonghua vs. Cantonese/Sichuanese) in everyday WeChat group conversations, and how these choices reflect aspects of social identity and power dynamics. The study aims to better understand digital multilingual practices in informal social environments. This project contributes to ongoing research in sociolinguistics, computer–mediated communication (CMC), and language ideology.
Procedures:
If you agree to participate, you will be asked to complete a brief online questionnaire (approximately 5–10 minutes) about your language background and language use patterns in WeChat groups. All responses are anonymous and will be used for academic purposes only.
开始作答


关于普通话和方言的调查
academic purposes only.
Confidentiality:
All responses will remain strictly confidential and anonymous. No identifying information (e.g., names, phone numbers, IP addresses) will be collected. Data will be stored securely and only accessed by the researcher. Results will be reported in aggregate form and will be used exclusively for academic research purposes. No individual participant will be identifiable in any thesis, publication, or presentation.
Risks and Benefits:
There are no foreseeable risks to you for participating in this study. You will not receive direct benefits, but your participation contributes to a better understanding of sociolinguistic patterns in Chinese digital communication.
Voluntary Participation:
Participation is entirely voluntary. You may withdraw at any time without giving a reason and without penalty. You may skip any questions you are not comfortable answering.
Consent:
I have read and understood the provided information and have had the opportunity to ask questions. I understand that my participation is voluntary and that I am free to withdraw at any time, without giving a reason and without cost. I understand that I shall be given a copy of this consent form. By clicking the next button to enter the survey, I indicate my willingness to voluntarily take part in the study.
开始作答

Appendix E: Participant Demographics and Survey Sample Background

1. Group Chat Participant Information

Group	Total	Female	Male
C1	12	9	3
C2	14	10	4

Table 1. Gender Distribution by Group

Age Group	C1	C2
18–25	7	8
26–35	5	6

Table 2. Age Distribution by Group

2. Questionnaire Respondent Overview

All members in these groups filled out the survey, which means we got a 100% response rate from group participants. A total of 74 questionnaire responses were collected, including 26 from group chat members (Q1) and 48 from an external comparison sample (Q2).

Age Range	Group Members (Q1)	Comparison Sample (Q2)	Total
18–25 years	15	44	59
26 years and above	11	4	15

Table 3. Age Distribution.

Language	Group Members (Q1)	Comparison Sample (Q2)
Cantonese	21	24
Mandarin	14	35

Table 4. Number of Native Speakers (Cantonese/Mandarin).

These data indicate differences in age and regional composition between the two samples. The group member sample (Q1) tends to span multiple age groups, especially including older generations, while the comparison sample (Q2) is concentrated among youth. Self-assessment data show that among group members, 21 consider Cantonese their mother tongue and in the external sample, 24 are native Cantonese speakers. In other words, there is a substantial proportion of Cantonese users in the external sample, which helps in comparison with the Cantonese group. For Mandarin ability, almost all respondents report a good level, 3 group members and 1 external participant rated their Mandarin only as “average”.