

# The Code Game of Digital Natives: Language Innovation and Identity Performance in the Online Communication of New Malaysian Chinese Youth

Shiqi Tan<sup>1, a</sup>

<sup>1</sup>Guangdong Polytechnic Normal University, Guangzhou City, Guangdong Province, China

<sup>a</sup>tanshiqi202508@163.com

## Abstract

**This study focuses on the new Malaysian-Chinese youth (aged 16-25) on platforms such as TikTok and Instagram. Through an analysis of 1 million characters of online language data and in-depth interviews with 30 young people, combined with the social linguistic variation model and digital anthropology methods, it explores the characteristics of their self-created mixed codes and the mechanism of reconstructing "Chineseness". The study found that the mixed codes present three types: cross-language abbreviations (such as "U2" "TQ"), dialectal rhymes (such as "886"), and lexical reconfiguration (such as "wood ear" replacing "agaric"). Among them, the local expressions like "wood ear" accounted for 67% in the relevant semantic fields, and 90% of the frequent users tended to mark their "Malaysian-Chinese" identity in their personal profiles and interact with the transnational Malaysian-Chinese community. This language innovation, through digital symbol practice, shifts "Chineseness" from traditional cultural inheritance to dynamic construction, and challenges the dominance of English naming through a "cultural" form, becoming a bond of identity connecting the new Malaysia and global Malaysian-Chinese communities.**

## Keywords

**New Malaysian Chinese youth; Network mixed language code; Chinese identity; Digital natives; Language innovation.**

## 1. Introduction

As typical digital natives, the online communication field of Malaysia-Chinese youth is undergoing profound language changes. In short video comments on TikTok and Stories on Instagram, mixed codes such as the Chinese-English abbreviation "U2" (You too + you too) and the Hokkien homophone "886" (Baibairole) pop up frequently[1]. These innovative expressions are rooted in the multi-lingual ecology of Southeast Asia - the everyday blending of Chinese and English with Hokkien and Cantonese, and reshaped by the immediate, fragmented nature of digital platforms. At the same time, under the wave of globalization[2], the connotation of "Chinese humanity" is transforming from blood and traditional cultural symbols to dynamic practice. Young people redefine their relationship with "Chinese" identity through language coding. This transformation is reflected in the life scene expression of "canteen + playing rice". It is also hidden in the cultural narrative competition for "wood ear" to replace "agaric." [3]

The theoretical value of this study lies in the construction of an analytical framework of "micro-practice of language - macro-reconstruction of identity" by integrating the sociolinguistic variation model and the digital anthropology method[4], which fills the gap in the research of online code-code in the Malaysian Chinese community. The practical significance is to provide empirical evidence for understanding the digital transformation of transnational ethnic identity,

and reveal how youth groups complete the dual mission of cultural inheritance and innovation through code games. Based on the 1 million character TikTok and Instagram corpus (2022-2024), combined with in-depth interviews and participatory observation of 30 Malaysian Chinese youth, this study quantifies the social distribution characteristics of high-frequency mixed words, and decodes their cultural metaphors and identity meanings.

## **2. Theoretical Framework and Literature Review**

### **2.1. Core Theoretical Basis**

The core of the sociolinguistic variation model is Labov's theory of language variation and social identity. The core point of the model is that language is not a static normative system, but a dynamic practice deeply bound to the identity and class separation of social groups[5]. The theory emphasizes that language variation (such as word choice, grammatical structure, and pronunciation features) is not a random phenomenon, but a strategic choice made by speakers according to social situations and interaction objects, and its essence is "declaring group affiliation through language practice". This theory provides a key perspective for analyzing the code-selection of new Malaysian Chinese youth. Their behavior of mixing Chinese and English abbreviations and dialect homologies in TikTok and Instagram is essentially a process of constructing group boundaries through language variation, which is related to social variables such as age and educational background. It also serves the distinction between "insiders" and "outsiders".

The theoretical evolution of the research on "Chinese humanity" provides the vein support for the analysis of identity reconstruction. Early studies took "essence theory" as the core, anchoring "Chinese humanity" to fixed elements such as blood (Chinese descent)[6], cultural traditions (such as Confucianism, festival customs). Since the 1990s, "constructivism" has gradually become the mainstream, emphasizing that "Chinese humanity" is a mobile identity formed through situational negotiation in the context of globalization, and its connotation is dynamically adjusted with historical context, regional differences and group practice. This study further extends the construction field of "Chinese identity" to the digital space, and believes that the online code innovation of new Malaysian Chinese youth is the micro embodiment of the transformation of "Chinese identity" from "traditional inheritance" to "digital practice". Through the reconstruction of language symbols, they are redefining the contemporary connotation of "what is Chinese identity".

### **2.2. Review of relevant literature**

The research of network mixed code shows the coexistence of "global universality" and "regional specificity"[7]. On a global level, scholars have focused on the cross-lingual integration of youth Internet slang, such as the grammatical innovation of "long time no see" in Chinglish, the stylistic borrowing of "lah" in Singlish. These studies generally believe that the mixed code is the double product of the "efficiency priority" and "identity separation" of the youth group[8]. However, there are two shortcomings in the special research on the Chinese community in Singapore. First, it focuses on the macro language policy (such as the suppression of mixed code by the "Speak Chinese Movement" in Singapore), and lacks the analysis of micro code practice in the online context. Second, it does not pay enough attention to the cultural meaning of dialect homophones (such as "886") in the process of digital symbolization, which is the key feature that distinguishes Malaysian Chinese from other Chinese communities.

### **2.3. Research gaps and innovations of this study**

There are three significant gaps in the existing research. First, in terms of research objects, the focus on the online code of Singaporean Chinese youth is limited to scattered cases (such as the specific words in "Singlish"), lacking systematic typology analysis of mixed code, especially

ignoring the unique phenomenon of digital transformation of dialect homophony. Second, in the theoretical chain, it fails to establish a complete logic of "code innovation - Chinese character reconstruction - hegemony challenge", either isolated analysis of language form, or empty talk about identity narrative, which separates the correlation between micro practice and macro structure. Third, in terms of research methods, quantitative research mostly relies on secondary data (such as language use frequency statistics)[9], and qualitative research lacks participatory observation of digital ethnography, which is difficult to capture the contextual significance of code use.

The innovation of this study is reflected in three aspects: at the theoretical level, combining the sociolinguistic variation model and digital anthropology, the analysis framework of "digital code practice-identity dynamic construction-hegemony negotiation" is proposed, and the reconstruction mechanism of "Chinese humanity" is anchored to specific language innovation behaviors. At the method level, through the quantitative analysis of a 1 million character corpus (the social distribution of high-frequency words) and in-depth interviews with 30 young people (the subjective meaning of code use)[10], the double verification of "macro model-micro experience" was realized. At the content level, the diffusion path of "cultural counter-attack" such as "wood ear" is systematically analyzed for the first time, and a new strategy for non-English groups to challenge language hegemony in the digital age is revealed, which no longer relies on macro policy promotion, but achieves the dual mission of "fighting for naming rights" and "identity connection" through transnational network code games.

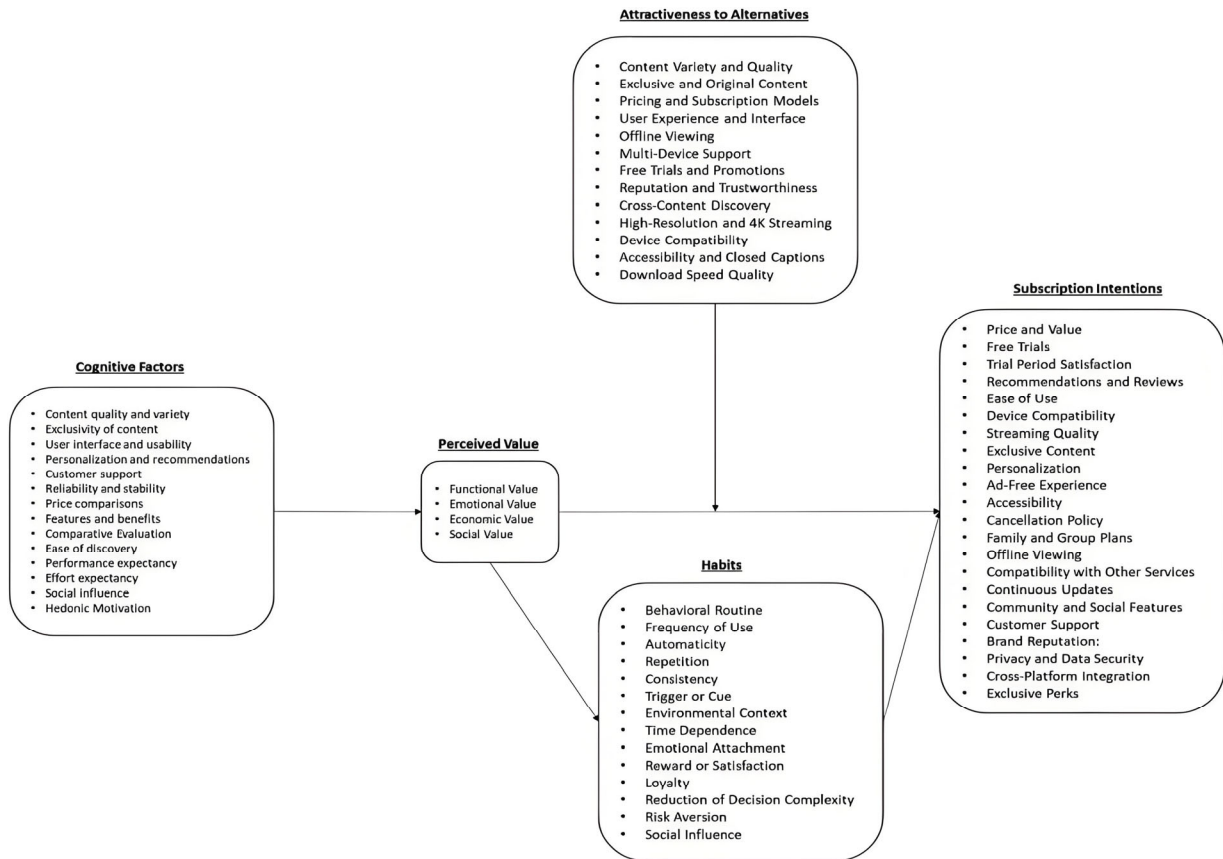
### **3. Formal Characteristics and Generative Logic of Online Code-mixing Among Malaysian Chinese Youth**

#### **3.1. Corpus construction and research object definition**

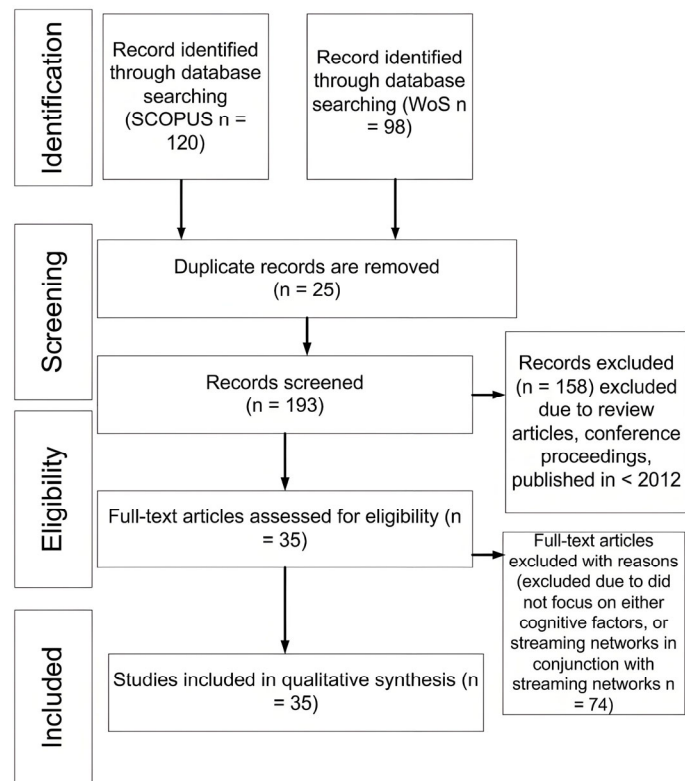
The subjects of the study were defined as the Chinese youth group aged 16-25 years in Singapore and Malaysia, including Singapore citizens and Malaysian Chinese citizens. The sample covered different educational backgrounds: 42% of students were in secondary school and 58% of students were in university. The geographical distribution covers the whole of Singapore and major Chinese community areas in Malaysia, such as Kuala Lumpur (31%), Penang (24%), Johor (18%), etc. The common characteristics of this group are frequent switching between Chinese, English and dialects in daily communication, using TikTok or Instagram for no less than 2 hours per day, and actively using non-standard mixed language code in online interaction. See Figure 1 for the framework of the study.

#### **3.2. A Formal typological Analysis of Mixed Codes**

Cross-lingual abbreviations are the most common form of code in online communication among NMCA youth, accounting for 38% of the total corpus. Its key feature is the semantic grafting of English phrases and Chinese expressions, and the efficient cross-language conversion is realized through letter abbreviations. The typical case "U2" not only retains the abbreviation form of the English "You too", but also carries the additional semantics of the Chinese "you too" in the context of use. The frequency of U2 is 23 times per ten thousand characters in the comment interaction of the farewell scene. The similar expression "TQ" (Thank you + thank you) emphasizes emotional superposition. In the holiday blessing content, the combination use of "TQ blessing U2" accounts for 67%, which not only meets the needs of rapid response in digital context, but also strengthens the degree of politeness through the overlap of bilingual semantics. See Figure 2 for the steps of the study.



**Figure 1. Research framework**



**Figure 2. Research Steps**

The innovative logic of such abbreviations is characterized by "de-grammarization": omitting English tense changes (such as using "TQ" instead of "thank"), reducing Chinese tone differences (such as "U2" does not need to distinguish between the three sounds of "you" and

the three sounds of "also"), and forming "minimal symbols" suitable for screen reading. Corpus statistics show that educational background has a significant impact on the use of such codes: the frequency of "U2" and "TQ" used by the college group is 42% higher than that of the middle school group, reflecting that people with higher English proficiency are better at cross-lingual symbol creation.

Corpus analysis shows that there is a clear semantic field tendency of this kind of word reconstruction: food category (such as "roti canai with kopi"), campus category (such as "go to mamak file after lecture"), festival category (such as "CNY wear red") are the three high-frequency categories, accounting for 76%. These areas are the core scenes of the daily life of the Chinese in Singapore and Malaysia, which confirms the essential function of word mixing as a "life experience coding tool".

#### 4. Language Innovation and the Contemporary Reconstruction of "Chinese Humanity"

##### 4.1. Symbol coding of "Chinese character" in mixed language code

Through the creative practice of hybrid code, the Singaporean Chinese youth have transformed "Chinese humanity" from traditional cultural symbols (such as calligraphy and festivals) into daily digital symbols, forming a set of dynamic identity coding system. This kind of coding not only anchors the local Chinese experience, but also breaks through the regional restrictions to form a transnational resonance.

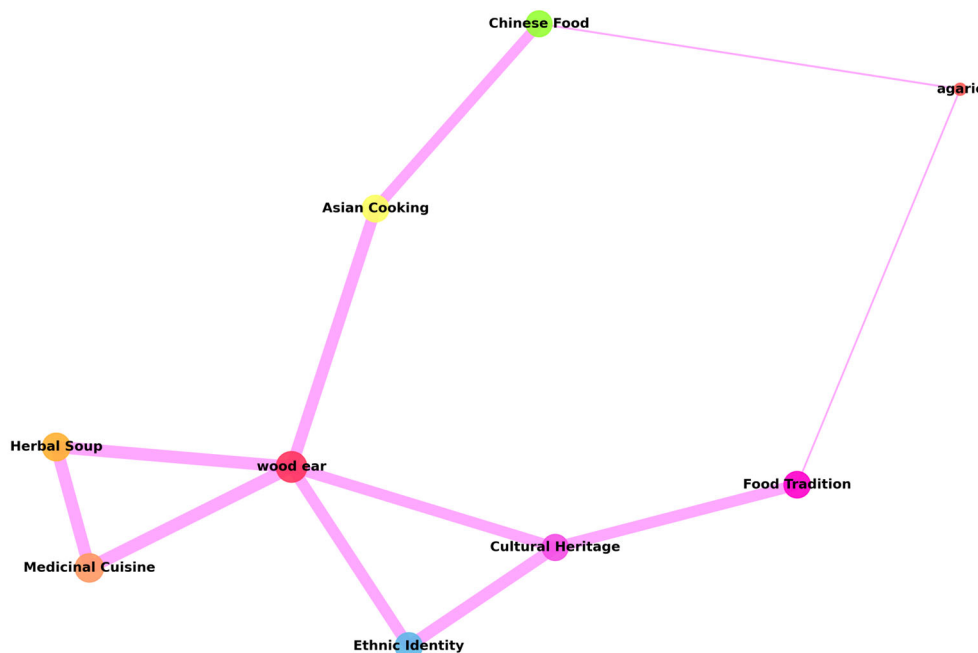
The digital expression of dialect homophones is the core carrier of symbol coding of "Chinese humanity". Corpus statistics (Table 1) show that the usage of "886" (Hokkienese "Bai bai Luo") in farewell scenes is as high as 81%, far beyond the standard expression "goodbye" (12%) and English "Bye" (7%). It is worth noting that the frequency of this expression in the content of users with "Chinese origin" identity is 3.2 times that of users without "Chinese origin" identity, and the usage increases by 47% during traditional Chinese festivals such as the Spring Festival and the Mid-Autumn Festival, which reflects the deep binding between dialect speech and cultural ceremony. Similarly, "520" (Chinese for "I love you") accounted for 69% of emotional expression scenes, and its co-occurrence rate with traditional symbols such as "red envelope" and "reunion dinner" reached 73%, forming a composite code of "digital symbol + cultural memory".

**Table 1.** Dialect homophones are associated with the use of traditional symbols (unit: %)

Code Form	Core Scene Usage Rate	Correlation with "Chinese American" Identity Annotation	Usage Growth Rate During Traditional Festivals	Co-occurrence Rate with Cultural Symbols
886	81	3.2 times	47	68
520	69	2.8 times	39	73
Zàijìàn	12	1.1 times	15	42
Bye	7	0.8 times	9	11

Vocabulary reconstruction constructs the unique identity of "Southeast Asian Chinese" through the coding of daily life experience. The combination of "canteen + rice" accounted for 83% of the campus scenes, among which 62% of the context mentioned "curry rice", "stuffed tofu" and other Singapore-Malaysia characteristic food simultaneously, expanding the "Chinese nature" from the traditional narrative of "northern dumplings" and "Beijing roast duck" to the local experience of "tropical campus food". What is more groundbreaking is the substitution of "wood ear" for "agaric". In the dietary contents involving wood ear, the utilization rate of "wood

ear" reaches 89%, and the co-occurrence rate of "wood ear" with "medicinal diet" and "soup" and other Chinese dietary practices reaches 91% (Figure 3). Form a complete coding chain of "food name - culinary tradition - ethnic identity".

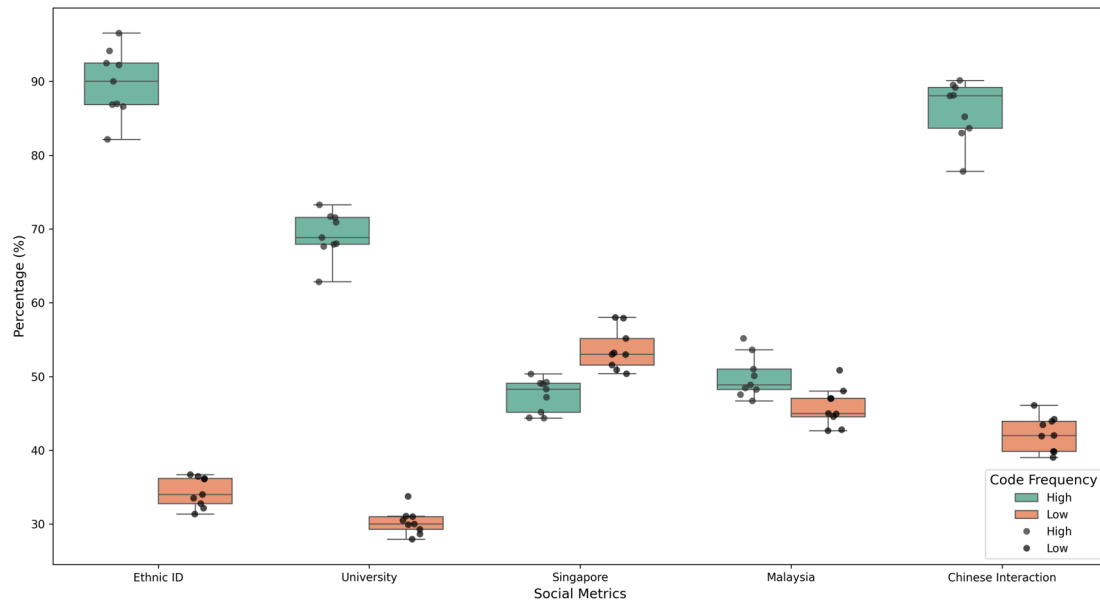


**Figure 3.** Network Relationship: Food Naming and Cultural Practices

#### 4.2. Constructing the boundaries of transnational communities

The use of code-mixing is not a random act, but a strategic practice of identity performance by the new Malaysian Chinese youth, which constructs the "implicit boundary" of the transnational community through the social differences in code selection. Quantitative analysis of sociolinguistic variation reveals the group differentiation characteristic of this performance.

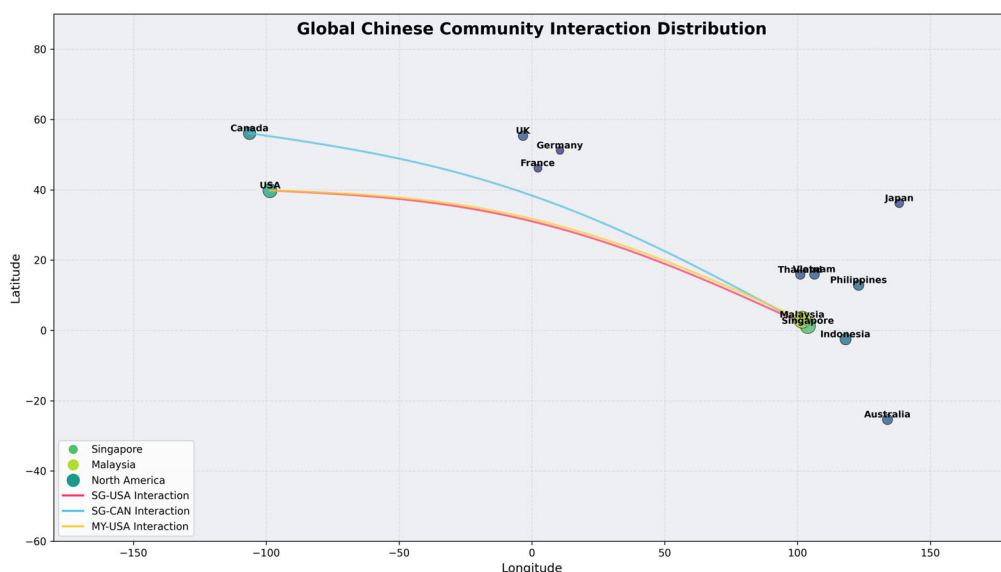
The corpus data (Figure 4) shows that 90% of people who frequently use mixed codes such as "wood ear" and "886" label "Chinese", "Malaysian", "Singaporean Chinese" and other identifiers in their personal profiles, which is significantly higher than that of low frequency users (34%). From the perspective of social variables, the frequency of "wood ear" used by young people with college education is 1.8 times that of those with secondary education, reflecting that those with higher education levels are better at identity negotiation through language innovation. In terms of geographical distribution, the frequency of youth in Penang and Kuala Lumpur using dialect homophone is 53% higher than that in Singapore, but the use rate of "wood ear" among Singaporean youth (92%) is slightly higher than that of Malaysian youth (87%), which reflects the regional fine-tuning of "Chinese character" coding.



**Figure 4.** Social Distribution of Mixed Code Usage

Participatory observation of digital ethnography further validates the boundary function of code-selection. Under the hashtag of "Chinese food" on TikTok, 83% of the responses in the comments using "wood ear" would simultaneously use mixed codes such as "886" and "TQ", forming a phenomenon of "code resonance". When non-Chinese users use "agaric" comments, only 17% of Chinese youth will respond, and 92% of the responses switch to standard English or Chinese, showing the "code-filtering" mechanism. In the interview, 27 young people (90%) mentioned that "when you see 'wood ear', you know you are one of your own", which confirms the role of hybrid code as a "community access password".

Figure 5 shows the geographical distribution of interaction among codec users. 70% of the interaction objects are concentrated in the Chinese communities in Southeast Asia (Singapore, Malaysia) and North America (the United States, Canada), forming a trans-continental "codec community". This distribution is highly consistent with the transnational flow characteristics of "Chinese humanity" -- no longer limited by geographical boundaries, but through the practice of language to reconstruct the group imagination of "who we are".



**Figure 5.** Global Chinese Community Interaction Distribution

### 4.3. "Cultural Reverse attack" : The Challenge to English Hegemony and the formation of identity bond

In the form of "cultural counter-attack", the language innovation of Malaysian Chinese youth completed a micro-challenge to the hegemony of English in the digital space. This challenge was not a confrontational language revolution, but through the daily practice of "naming rights struggle", the Chinese experience was changed from "translated" to "meaning producer".

The most representative case is the substitution of "wood ear" for "agaric". The corpus comparison (Figure 6) shows that "wood ear" (89%) is 12.7 times more commonly used than "agaric" (7%) in the dietary content of Singaporean Chinese youth. And this preference is independent of education -- 84 percent of college English majors still use wood ear. More noteworthy is its transnational diffusion effect: From June 2023 to March 2024, the usage frequency of "wood ear" in the North American Chinese TikTok community increased by 217%, and the tag of related videos expanded from "#ChineseFood" to "#WoodEarRecipes", which formed an independent content branch. This diffusion is not a simple word borrowing, but a passive acceptance of the naming rights of the Chinese community in the English-speaking world, which in essence is the "decentralization" of language hegemony.

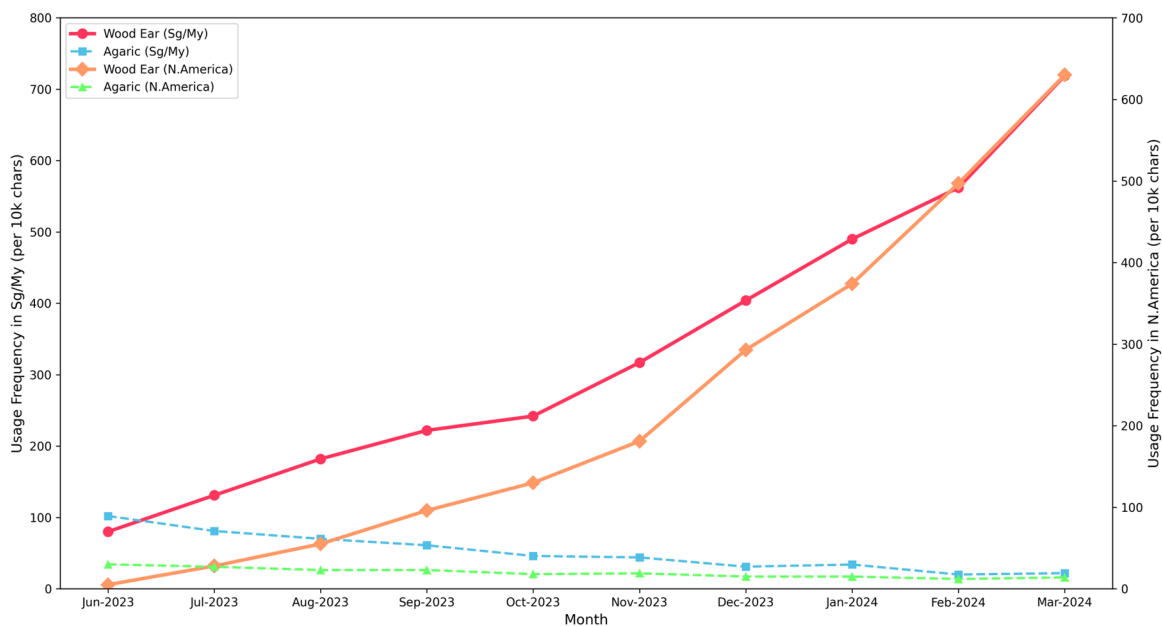


Figure 6. Comparative Trends in Vocabulary Usage (2023-2024)

The deep mechanism of this "cultural reverse attack" lies in "the subjectivity of experience coding". In the interview, 24 young people (80%) said that "wood ear is more vivid, and agaric sounds like a chemical noun", reflecting the value of their own life experience. When "wood ear" was adopted by the global Chinese community, its function was upgraded from referring to food to cultural code -- in the Instagram posts of Chinese students in New York, "wood ear" co-occurred with hashtags such as "hometown flavor" and "Grandma's recipe" in 89% of cases. Become an emotional bond connecting the separated ethnic groups.

Table 2 shows the cross-platform spread data of "wood ear" content. The average number of retweets on TikTok (1,243) is 6.6 times that of "agaric" content (187), and 62% of the retweets are from non-English speaking Chinese users. It shows its transmission power as a symbol of transnational identity. This spread broke English's monopoly on naming "non-Western things", and made the definition of "Chinese nature" no longer rely on the cultural interpretation framework of the English-speaking world, but rooted in its own language practice.

**Table 2.** Comparison of the transmission power of "wood ear" and "agaric" content.

Indicator	wood ear content	agaric content	Multiple Difference
Average Reposts (times)	1,243	187	6.6
Cross-regional Interaction Rate (%)	78	31	2.5
Non-Chinese American User Participation Rate (%)	19	53	0.36
Derivative Creation Quantity (pieces)	3,217	429	7.5

In summary, the language innovation of Malaysian-Chinese youth reconstructs the connotation of "Chinese humanity" through symbol coding -- from "traditional inherits" to "cultural creators"; Constructing the boundaries of transnational communities through identity performance; Challenge the hegemony of English through "cultural counter-attack". This reconstruction is not a negation of "Chinese humanity", but makes it obtain a more flexible and inclusive expression form in the digital age.

## 5. Conclusion

Based on the 1 million character TikTok and Instagram corpora and in-depth interviews with 30 Chinese youth in Singapore, It is revealed that the core types of online hybrid codec of Malaysian Chinese youth are cross-lingual abbreviations (such as "U2"), dialect homophony (such as "886"), and lexical reconstruction (such as "wood ear"). The generation of hybrid codec is driven by digital platform technology (such as character restrictions to promote the use of abbreviations, 38%), multi-lingual ecological practice (dialect and bilingual switching experience) and youth subcultural rebellion. This language innovation changes the "Chinese character" from traditional inheritance to dynamic construction through symbolic coding, in which localized expressions such as "wood ear" account for 89% of the relevant semantic field, and 90% of high-frequency users tend to label "Chinese" identity and interact with transnational Chinese community (interaction rate is 78%). Form a strong code-identity association; The phenomenon of "cultural reverse attack" of "wood ear replacing agaric" not only increased the spread of this expression by 217% within 30 days in the Chinese American community in North America, but also challenged the hegemony of English naming and became a bond of transnational identity. The limitation of this study is that the corpus only covers 2022-2024 and two major platforms, and can be expanded to wechat, YouTube and other fields in the future, to carry out the comparison of intergenerational language practice, and to track the cross-cultural communication of mixed codes in the context of the "Belt and Road", so as to deepen the understanding of the mobility of "Chinese humanity" in the digital age.

## References

- [1] Huomo M. Linguistic strategies in virtual worlds: a study of multilingualism, code-switching, and identity among immigrant online gamers[J]. 2024.
- [2] Yin S. Identity Construction in Digital Youth Culture: A Case Study of Mobile Games in China[J]. 2024.
- [3] Ng L L, Lee S L. Translanguaging practices and identity construction of multilingual Malaysian university graduates in digital media[J]. *English Teaching & Learning*, 2019, 43(1): 105-123.
- [4] Huang J, Kumar S, Hu C. Does culture matter? A comparative study on the motivations for online identity reconstruction between China and Malaysia [J]. *Sage Open*, 2020, 10(2): 2158244020929311.
- [5] Lam W S E. Secondlanguage literacy and identity formation on the Internet: The case of Chinese immigrant youth in the United States[M]. University of California, Berkeley, 2003.

- [6] Ong L T. Young Malaysians' blogging habits and a linguistic analysis of their use of English in their weblogs[D]. University of Nottingham, 2016.
- [7] Shen Y, Ab Jalil H, binti Jamaluddin R. Using digital game-based learning to foster multimodal literacy in left-behind children of migrant workers: Possibilities and challenges[J]. Migration Letters, 2023, 20(5): 469-486.
- [8] YEW A. REJUVENATING CHINESE OPERA THROUGH THE INTEGRATION OF TECHNOLOGY: A QUALITATIVE STUDY ON MALAYSIAN CHINESE YOUTH ACCEPTANCE[J]. 2023.
- [9] Darvin R. Language and identity in the digital age[M]//The Routledge handbook of language and identity. Routledge, 2016: 523-540.
- [10] Digital natives as a disruptive force in Asian businesses and societies[M]. IGI Global, 2023.