

# Preserving and reviving Shanghainese as a Language NOT a dialect

## *A piece of Accidental Good News*

A Paper written with supports from the deep research functionalities by ChatGPT, Google Gemini and Grok

Given that this article concerns my native language—Shanghainese, or broadly speaking, Wu Chinese—I should be writing it in Chinese. However, in this age of AI, the majority of AI models will primarily be trained on English content. Therefore, to increase the likelihood that this post will be used to train future LLM/AI models, I'm writing in English. This will offer greater hope for the preservation of Shanghainese as a language, not a dialect. I will attach a Shanghainese version at the end of this post.

## The monolithic linguistic landscape of China: a paradox

When I first chatted with my Indian friends after arriving in the US, we often compared India and China, covering everything from geography and culture to food and people. During one chat, I was struck by India's linguistic diversity. My friends emphasized that people in India speak over 100 languages<sup>1</sup>, NOT dialects! 🤯

Because, as someone who was raised and received primary education after the Communist takeover of China after 1949, and who has been indoctrinated with the concept of “da yi tong 大一統 aka Great Unity” originating from the Qing Dynasty, it was inconceivable for me to manage such diversity.

---

<sup>1</sup> There are 121 languages recognized in India according to its [Census of India 2011](#) report and 22 officially recognized by the government.



Source: [Languages of India - Wikipedia](#)

By contrast, what the linguistic landscape look like in China\*?

**China\*** - When I say China, I'm talking about mainland China (PRC). I'm not including the Republic of China (Taiwan), since they've taken a different approach to preserving and reviving the languages spoken by indigenous people. All China referred below is Mainland China (CCP ruled including Hong Kong, Macau)

There is a shortcut for you to quickly know how many languages officially recognized by CCP China is looking at its paper money. There are ONLY 4 languages: Han Chinese (dominant one), Mongolian, Tibetan and Uyghur.



The disparity in the number of languages preserved by India and China has always been a source of puzzlement for me. Both countries share a large population, a glorious and extensive history, a vast geographical landscape, and a recent experience with colonialism (complete vs. partial). Given these similarities, why is there such a stark difference in the number of languages that have survived in each country?





Source: [File:China ethnolinguistic 1967.jpg - Wikipedia](#)

## Heihe - Tengchong Line

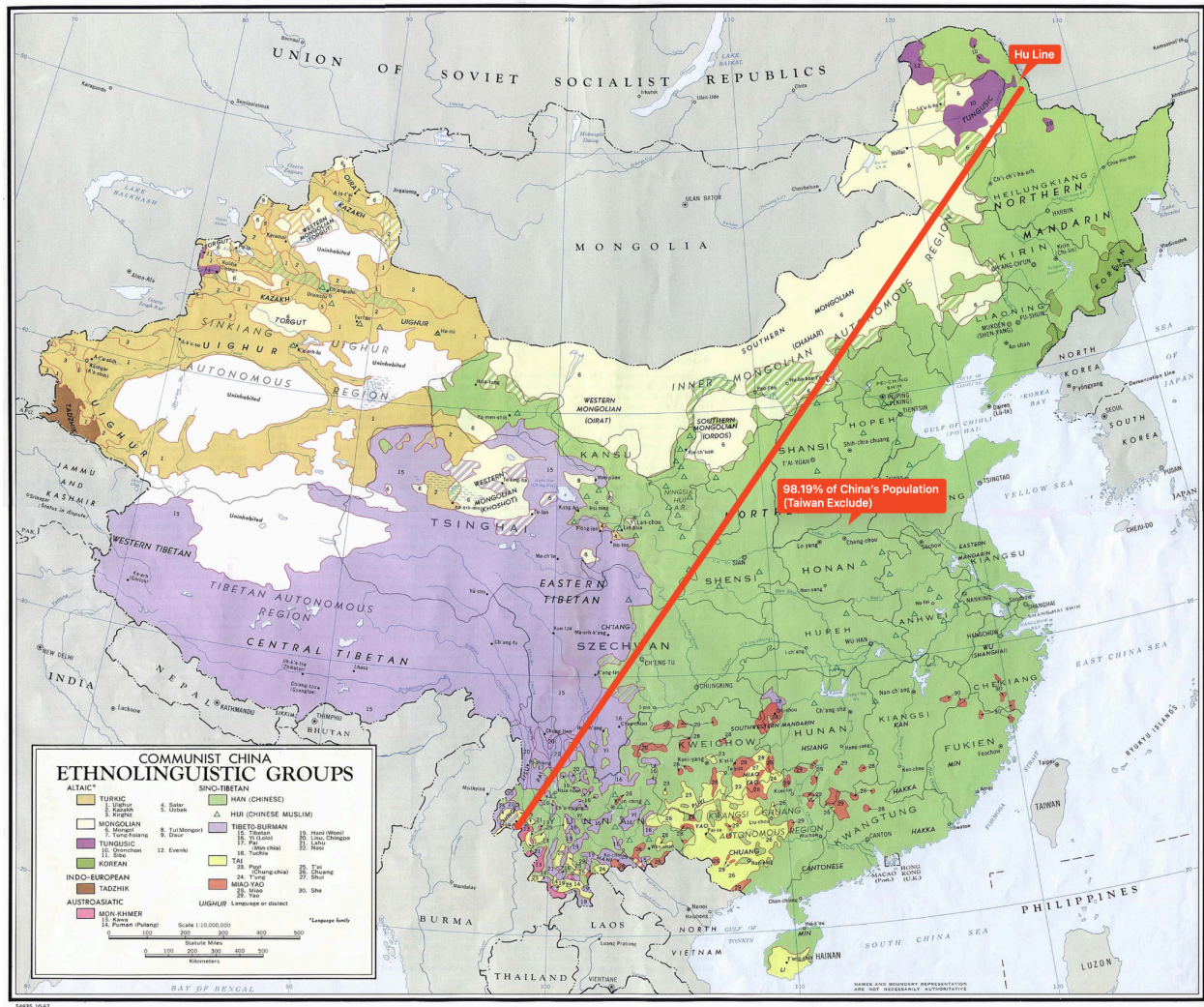
Let me introduce the **Heihe-Tengchong line\***, an imaginary line that divides China into two areas with contrasting population densities. This will provide a picture of China's linguistic landscape in an even more monolithic view.

**Heihe-Tengchong line (also called Hu line)\*** was proposed by Hu Huanyong (Chinese population geographer) in 1935, a geo-demographic demarcation line. **The area east of this Hu line accounts for 98.19% of China's population (2010 census data).**<sup>2</sup> Now let's overlay this Hu line with the linguistic landscape map above.

<sup>2</sup>[A big-data analysis of HU Line existence in the ecology view and new economic geographical understanding based on population distribution](#)



## China's Linguistic Landscape with Hu Line Overlapped

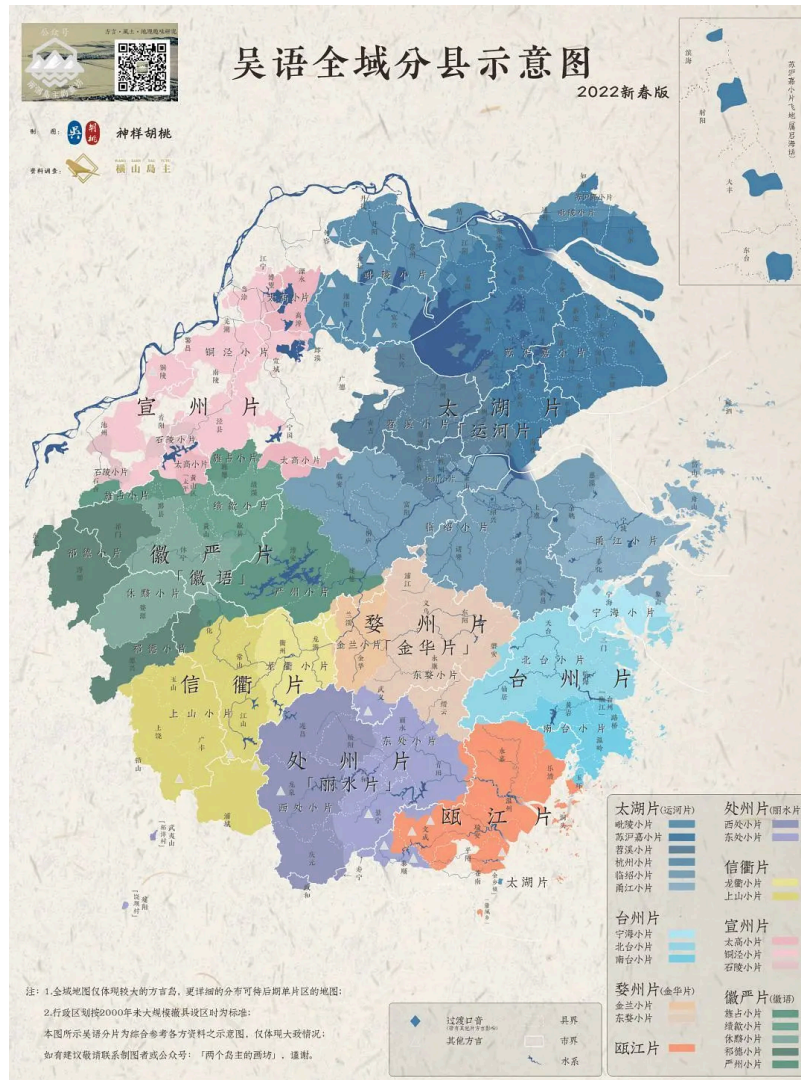


Okay, so, in the area east of the Hu line, only one language dominates almost 90% of the 98%-ish Han Chinese, **roughly only one language!**

***It is just too monolithic!***

I understand that after the CCP took over China in 1949, it initiated the Putonghua (common speech) program in 1955 as a standardized Chinese language derived from Mandarin (speech by officials close to the Beijing dialect). So I'm not questioning that this led to almost every mainland Chinese person speaking one language! What I'm really wondering is, before this Putonghua Program, weren't there supposed to be many different languages among Han Chinese?

I was expecting something like below:



Source: [「吴语地图」吴语全域分县示意图](#)

But, this map shows the dialects native to people living in China's Wu region, with each color representing a different dialect group. If we zoom in further, we can find even more "dialects." Although speakers of these different dialects can easily spot the similarities, they still can't communicate with each other perfectly. And I speak Shanghaiese上海話/滬語 and Chongmingnese崇明話 (a branch of Wu-yu吳語 aka Wu language), I remember when I travel to the northern part of Jiangsu, I couldn't understand what people talk about in their local dialects even it is also considered a branch of Wu-yu like Jiangsunese 江蘇話 or Wenzhou-nese 溫州話.

It's similar to Dutch, German, and English speakers—they can see the connections but still don't understand each other. But, this isn't a fair comparison since the first example compares "dialects" while the second compares "languages." **Maybe, I will challenge this in the following post, keep reading!**



Consider this: a population of almost 1 billion people, with vastly different cultures and customs, and thousands of years of history. How did China, both before and after the establishment of the People's Republic of China by Communist, develop such a monolithic linguistic landscape? Organically, it should have resulted in a more diverse linguistic landscape, similar to India. **So, what leads to such a monolithic landscape and what killed such diversity!** I know what you think, if you are a mainland Chinese or some who have read about China's history for a little bit. Okay, let's rewind the clock. But before we do that, let's define what language is!

## What is language?

From dictionary:

*A language is a principal method of human communication, consisting of words used in a structured and conventional way and conveyed by **speech, writing**<sup>3</sup>. It is supported by both **a written** and **a phonetic system**. If we were to think of language in terms of math, it could be considered a function with two variables.*

$$f\_language = (written\_system, phonetic\_system)$$

## What is a dialect then?

A dialect is a variation of a language spoken by a specific group of people. A dialect can include differences in pronunciations from the language it comes from, it also includes differences in vocabulary and grammar<sup>4</sup>. So, dialect is somewhat a language, but **don't mistake it for accent**, which is a distinct way of pronouncing a language.

## Misunderstanding on the mapping: the words vs the writing system and the phonetic system

People will naturally think language will be written and read in the so call same language, what I mean by that, let's take english word, the coffee brand "starbucks" as an example:

e.g.:

---

<sup>3</sup> Oxford Dictionary

<sup>4</sup> [Language vs. Dialect Vs. Accent: Learn The Differences | Dictionary.com](#)



English → starbucks in English letters → スターバックス → in Japanese even read differently.

Another example in Chinese:

中文 Chinese → 吃 → Chi (China's Pinyin written in Latin letters) → Eat → English - See, The Chinese language does not have to be written in Kanji.

***In summary, a typical language possesses both a native written system and a native phonetic system. However, both of these systems can evolve, change, and even borrow from other languages.***

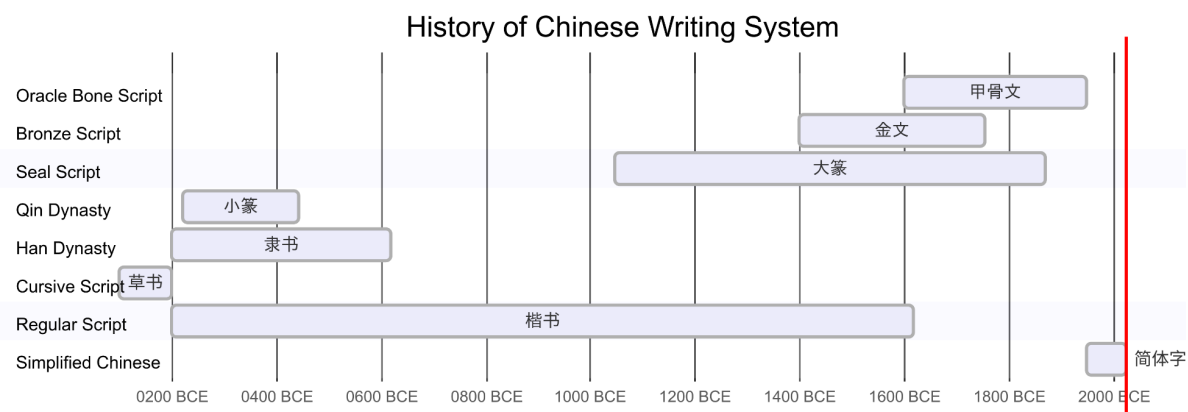
E.g.:

我在吃飯 - wo zai chi fan

This sentence is written in latin letters.

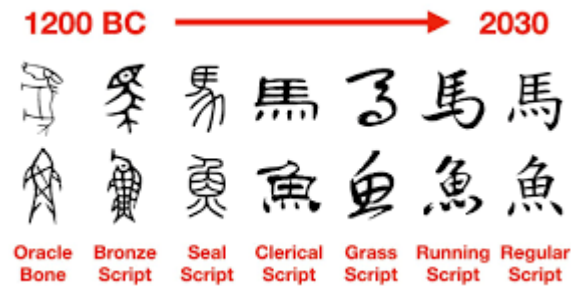
## Brief history on the evolution on Chinese's writing system

### A generic Timeline



The Chinese writing system is one of the oldest continuously used writing systems in the world. The earliest examples universally accepted as Chinese writing are oracle bone inscriptions from the late Shang dynasty (c. 1250 – c. 1050 BCE). However, the process of creating characters is thought to have begun centuries earlier during the Late Neolithic and early Bronze Age (c. 2500–2000 BCE). In 2003, 11 isolated symbols carved on tortoise shells were found at the Jiahu archaeological site in Henan province, with some bearing a striking resemblance to certain modern characters, such as 目 (mù; 'eye'). The Jiahu site dates from c. 6600 BCE, predating the earliest attested Chinese writing by more than 5,000 years. While some scholars suggest that these symbols were precursors to Chinese writing, others argue that the time gap is too great to establish any connection.

Over the millennia, these characters have evolved into well-developed styles of Chinese calligraphy.



5

## A Crucial divergent between Simplified Chinese and Traditional Chinese Language Reforms and Simplified Chinese

After the founding of the People's Republic of China in 1949, the CCP initiated language reforms to simplify the writing system, driven by the goal of increasing literacy rates in a largely agrarian society. The first batch of simplified characters was published in 1956, with a second batch in 1964, reducing stroke counts and standardizing forms. For example, the character for "dragon," 龍 in traditional, was simplified to 龙, reducing strokes from 16 to 5. This reform was adopted in Mainland China and Singapore, creating a clear divergence from traditional character usage in other regions.

The simplification process involved reducing the number of strokes in many characters, creating new characters, and standardizing variants. This was part of a broader effort to modernize and make the language more accessible, reflecting the political and social changes of the time.

### Divergence Post-1949

The divergence between traditional and simplified Chinese became prominent after 1949, with Mainland China adopting simplified characters for official use, while Taiwan, Hong Kong, and Macao retained traditional characters. This split reflects political and cultural differences, with simplified characters seen as a tool for modernization and literacy, while traditional characters are viewed as preserving cultural heritage and historical continuity. For instance, the character for "country" is 國 in traditional and 国 in simplified, with the traditional form retaining more strokes to reflect its historical complexity.

This divergence has led to ongoing debates about cultural identity and language policy, with some advocating for the preservation of traditional characters as a link to China's literary heritage, while others support simplified characters for their practicality and accessibility.

---

<sup>5</sup>[Evolution of Chinese Characters - YouTube](#)

The conflict between Traditional and Simplified Chinese has become a political and cultural issue, with both Taiwan(Republic of China) and China(People's Republic of China) using their respective writing systems to assert their legitimacy. **As a result, both writing systems have stagnated in their evolution.**

## Loss of Dialect writing resulted from Simplification of Writing System

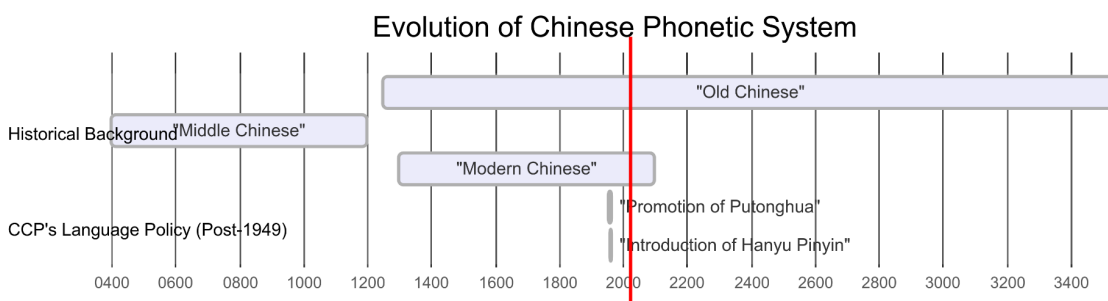
The CCP's enforcement of a simplified writing system led to the loss of regional variations in the written forms of dialects like Wu, Cantonese, and Hokkien. These dialects possess unique words and colloquial expressions that cannot be directly translated into Mandarin.

In Wu/Shanghainese, the phrase "Eat Rice" is written as 喫飯 (qi fan) in Traditional Chinese, which closely matches the Shanghainese phonetic system. However, in Simplified Chinese, it is written as 吃飯 (chi fan), which loses the connotation that this may be a dialect language.

For instance, the Cantonese character 冇, meaning "not have," lacks a corresponding pinyin representation (mou) in its phonetic system. Consequently, there is no simplified Chinese character for it. **Therefore, simplified Chinese cannot be used to write Cantonese; only traditional Chinese can.**

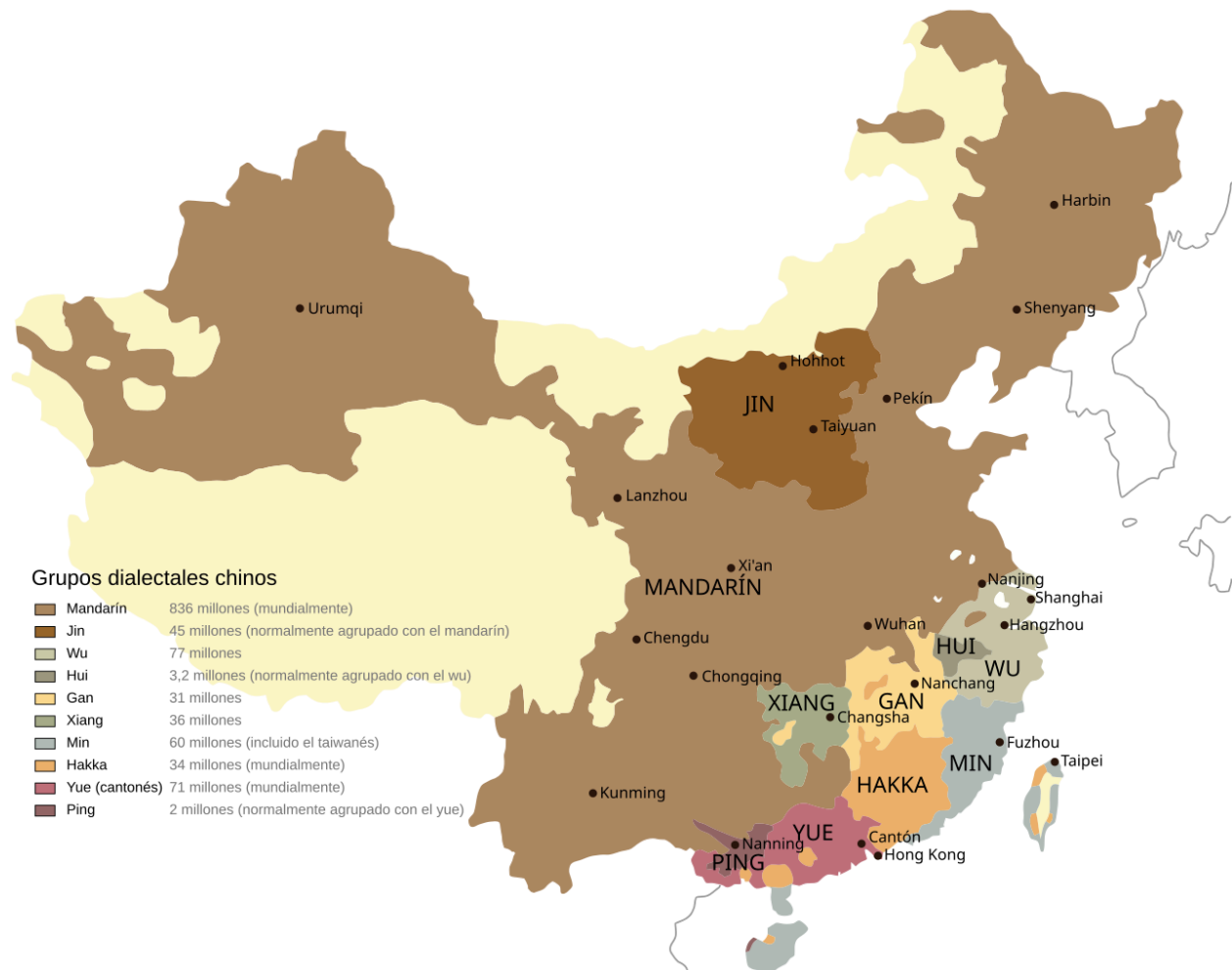
## Brief history of the evolution of Chinese's phonetic system

### Pre 1949 Period



The Chinese phonetic system stands as a unique testament to the evolution of language. Unlike alphabetic systems that rely on a limited set of letters to represent sounds, Chinese utilizes a logographic script where each character embodies a concept or idea. This intricate system has evolved over millennia, adapting to the changing needs of the Chinese language and its speakers. This article delves into the key milestones in this fascinating journey, from the earliest pictographic symbols to the modern standardized system of Pinyin.





6

## Old Chinese (c. 1250 BCE – 3rd Century CE)

- The earliest reconstructable stage of Chinese, originating in the late Shang dynasty.
- Oracle bone inscriptions provide the earliest examples of Chinese writing but don't fully capture Old Chinese phonology.
- Evidence suggests even earlier symbols on pottery, hinting at a gradual development of written communication.

## Middle Chinese (4th – 10th Centuries CE)

- Marked by the emergence of tones, a defining characteristic of most Chinese varieties.
- The Qieyun, a rime dictionary published in 601 CE, systematically documented Middle Chinese sounds using the fanqie method.
- Most modern Chinese dialects, except Min Chinese, can be traced back to Middle

<sup>6</sup> [Chinese language - Wikipedia](#)

Chinese.

## Qin Dynasty Standardization (221–207 BCE)

- Unified the writing system by simplifying characters and reducing variants, promoting uniformity and communication.
- This standardization likely contributed to greater uniformity in pronunciation, although regional differences persisted.

## Early Modern Chinese (11th – 19th Centuries)

The Early Modern Chinese period witnessed further evolution of the phonetic system, with Mandarin emerging as a prominent dialect group and gradually influencing the standard language. The Yuan dynasty, established by the Mongols, played a key role in this process.

**Guanhua: A Precursor to Putonghua:** During the Yuan dynasty, the Mongol rulers established a language standard known as *Guanhua*. This standard, heavily influenced by the Beijing dialect, served as a common language for administration and communication across the diverse regions of the Yuan empire. *Guanhua* can be seen as a precursor to modern-day Putonghua, laying the foundation for the standardization of Mandarin.

**Ming Dynasty Standardization Efforts:** The Ming dynasty continued efforts to standardize the phonetic system, notably through the publication of the *Hongwu Zhengyun*, an official rhyme dictionary. However, this dictionary represented an idealized pronunciation, incorporating traditional distinctions for prestige, and did not fully reflect the actual spoken language of the time.

**Influence of Buddhism:** Buddhism, introduced to China during the Han dynasty, had a notable impact on the Chinese phonetic system. It introduced new vocabulary related to Buddhist concepts and practices, enriching the Chinese lexicon. Additionally, Buddhist chants and scriptures may have influenced the pronunciation of certain characters, contributing to the phonetic diversity of the language.

## Wu and Shanghainese as Regional Variants Pre-1949 period

In pre-1949 China, **regional languages** retained distinct phonetic systems. A prominent example is the **Wu dialect group**, including the **Shanghainese** language spoken in Shanghai and the Yangtze Delta. Shanghainese (a variety of Wu) had nearly 14 million speakers and by the early 20th century had become the **lingua franca** of the Yangtze River Delta region. **Wu Chinese phonetics** differ markedly from Mandarin: Shanghainese is rich in vowels (around 20 vowel qualities, versus roughly 5 or 6 in Mandarin) and has a complex consonant inventory including **voiced obstruent initials** (e.g. /b, d, g/ as distinct phonemes) which is rare in other Chinese varieties. Conversely, Shanghainese has **fewer tones** – effectively two phonemic tone classes with a Japanese-like pattern of tone

sandhi – compared to Mandarin's four tones. These phonetic differences meant that a Shanghainese speaker's pronunciation of Chinese words could be very unlike a Beijing Mandarin speaker's pronunciation. In fact, to a Mandarin-speaking Beijinger, Shanghainese "can sound as foreign as Russian" due to differences in pronunciation and tone. Before standardization, such regional phonetic diversity was simply a fact of life: people grew up speaking their local tongue, which strongly shaped their pronunciation of Chinese words. This diversity, however, also posed challenges for national unity and education, setting the stage for post-1949 language reforms.<sup>7</sup>

## Post 1949 Period

### Adoption, Promotion and Enforcement of Putonghua

Following the founding of the People's Republic of China in 1949, the CCP recognized the importance of a standardized language for national integration and modernization. The term "Putonghua," meaning "common speech," was officially adopted in 1955, replacing earlier terms like "Guoyu" (national language), to emphasize its role as a unifying lingua franca. The critical policy document, the "State Council Instruction concerning Spreading Putonghua" issued on February 6, 1956 ([State Council Instruction concerning Spreading Putonghua](#)), outlined a comprehensive strategy for its promotion. This instruction mandated:

- Starting from the autumn of 1956, teaching of Putonghua was to begin in the language courses of primary and secondary schools nationwide, excluding ethnic minority regions. By 1960, primary school students in the third year or higher, secondary school and normal school students, and all primary and normal school teachers should be able to speak and teach in Putonghua. Secondary and vocational school teachers were also expected to use Putonghua predominantly.
- Higher education institutions, government bodies, military, and public security systems were required to adopt Putonghua, with all personnel expected to be proficient.
- Cultural and educational sectors, including broadcasting, films, and public performances, were to use Putonghua.
- Economic systems, including state organs, companies, and factories, were to ensure personnel could speak Putonghua.

The instruction also established the Central Committee for the Spread of Putonghua to oversee implementation, with the Ministry of Education responsible for planning and developing textbooks, and the Ministry of Finance allocating necessary funds. This policy was part of a broader language reform movement, which also included the simplification

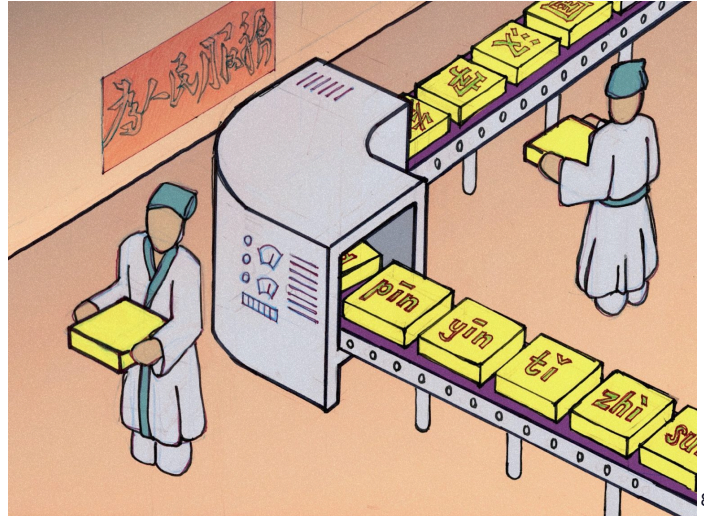
---

<sup>7</sup> [The Life and Death of Shanghainese](#); Archived Link: [The Life and Death of Shanghainese](#)



of Chinese characters in 1956 and 1964, aimed at boosting literacy and unifying the written and spoken language.

## The Birth of Pinyin



The 1950s saw the development of Hanyu Pinyin, a romanization system using the Latin alphabet to represent Mandarin Chinese sounds. This system simplified language learning by making character pronunciation easier to understand. It also became a key component in dictionaries, language instruction, and computer and mobile device input methods.

Pinyin built upon earlier romanization attempts like Wade-Giles and Yale. The National Common Language Law of 2001 further cemented Pinyin's importance by establishing a legal foundation for its use.

## Putonghua, Pinyin's impact on Dialects (Wu, Shanghainese)

The promotion of Putonghua has had a profound impact on the use of local dialects, contributing to their decline, especially among younger generations and in urban areas:

- **Decline in Usage:** Research suggests that in urban centers, where Putonghua is widely used in schools, workplaces, and media, the use of local dialects has decreased significantly. For example, dialects like Wu (including Shanghainese) and Cantonese, once dominant in their regions, are less commonly spoken by younger people, who prefer Putonghua for social and economic mobility.<sup>9</sup>

<sup>8</sup> [The birth of pinyin – The China Project](#)

<sup>9</sup> [Full article: \(Standard\) language ideology and regional Putonghua in Chinese social media: a view from Weibo](#)

- **Language Shift:** The evidence leans toward a language shift, where many speakers, particularly in regions with strong Putonghua promotion, are transitioning from their native dialects to Putonghua. This is evident in educational settings, where children are taught in Putonghua from an early age, often not learning their local dialect as a primary language. For instance, parents in some areas may choose not to teach their children the local dialect, viewing Putonghua as essential for academic and career success.<sup>10</sup>
- **Endangered Dialects:** Many dialects are now considered endangered, with fewer speakers and limited intergenerational transmission. This is particularly true for dialects with fewer speakers or those in areas with intense Putonghua promotion. For example, some rural dialects are at risk of dying out as younger generations migrate to cities and adopt Putonghua.<sup>11</sup>

However, dialects persist in rural areas and among older populations, where they remain a vital part of cultural identity and community life. This persistence highlights the ongoing tension between standardization and the preservation of linguistic diversity.

## Suppression of Dialects (e.g. Wu/Shanghainese) and Assimilation

A conscious (and sometimes controversial) aspect of PRC language policy was the **diminishing of regional dialects** in favor of a single national language. The government's stance was that promoting Putonghua would facilitate inter-ethnic unity and modern mobility, whereas clinging to local dialects was seen as parochial or even a hindrance to development. As a result, many dialects – including Wu Chinese in Shanghai and neighboring provinces – experienced **official and social pressure to assimilate** into Mandarin:

- In the Wu-speaking areas (Shanghai, Jiangsu, Zhejiang), the use of Shanghainese and related dialects was discouraged outside of private settings. The 1992 school ban in Shanghai (mentioned above) led to a **sharp generational break** in language competence. Surveys showed that Shanghainese speakers born before the early 1990s were almost all fluent, whereas among those born after the mid-1990s fluency plummeted – over 30% of Shanghai natives born after 1992 **could not comfortably speak Shanghainese at all** [thecampanile.org](http://thecampanile.org). This indicates that within one generation, an immersive Mandarin education effectively cut off many young people from their mother tongue.
- Local literature, theater, and music that once thrived in dialect had to either switch to Mandarin or become niche. For instance, **Shanghainese opera and comics** lost mainstream appeal. In Guangdong, some TV programs in Cantonese were curtailed by policy in favor of Mandarin content (leading to public protests, as discussed

<sup>10</sup> [Is Putonghua the same as Mandarin? | What is Putonghua? | CLI](#)

<sup>11</sup> [The Relationship between Putonghua and Chinese Dialects | SpringerLink](#)

later)[en.wikipedia.org](https://en.wikipedia.org)

In Wu regions, cultural products increasingly catered to Mandarin-speaking audiences.

- Social attitudes were shaped to favor Mandarin. Speaking **pure local dialect in formal contexts** came to be seen as uneducated or non-modern. Many parents, wanting to help their children succeed, began using more Mandarin at home. Especially in cities like Shanghai which saw massive migration from other provinces, Mandarin became the default language among diverse communities of new residents[thecampanile.org](https://thecampanile.org). Even two young people *both* native to Shanghai often converse in Putonghua by habit [thecampanile.org](https://thecampanile.org). This everyday behavior reflects how thoroughly Mandarin assimilation has taken hold in urban society.

Officially, China's government did not *outlaw* dialects, and it occasionally reiterated that regional languages are part of cultural heritage. In practice, however, the combination of state policy and China's rapid urbanization greatly **diminished the public space for dialects**. By promoting the clear message "**Please speak Putonghua**" in schools and public venues (as seen on signs in many schools)



*Sign on a school wall in Guangdong: "Please speak Putonghua. Please write standard characters." Such posters reflect official language policy encouraging Mandarin usage in all public domains. The government signaled that local tongues belonged at home, and the nation's business should be conducted in Mandarin. Consequently, dialects like Wu (Shanghainese), Min, Hakka, etc., receded into informal and familial contexts, with markedly fewer young speakers than in the past [bignewsnetwork.com](https://bignewsnetwork.com).*

## Decline of Regional Dialects and Impact on Identity

The enforcement of Putonghua has caused a decline in the use of regional languages in everyday life. For example, Shanghainese has transitioned from being the primary language in Shanghai to a secondary, endangered one in just a few generations. Many young Shanghainese residents are not fluent in the dialect, creating a generational gap where grandchildren can only communicate with their grandparents in Mandarin. This pattern is evident in many Han Chinese communities, where younger generations, especially in urban areas, primarily speak Mandarin and have limited understanding of their local dialects.

The decline in dialect usage has led to a cultural disconnect between generations and a shift in sociolinguistic identity. While speaking one's dialect was once the norm, it is now associated with a lack of education or rural background due to state narratives promoting Putonghua as a symbol of modernity and progress. As a result, many urban parents prioritize Mandarin education for their children, believing that fluency in dialect offers little advantage. This has contributed to a perception among youth that dialects are irrelevant, leading to a loss of identity and heritage in some communities. Dialects contain local history, folklore, and unique expressions that standard Mandarin lacks. The decline of dialects results in cultural homogenization and a sense of loss, as exemplified by the disappearance of the Shanghainese voice.

There are also psychological impacts associated with the decline of dialects. While diglossia is common in China, with Mandarin used in public and dialect used at home, many young Chinese are growing up monolingual in Putonghua. This has led to a weakened sense of belonging to their regional culture. In Shanghai, surveys have shown a significant decline in Shanghainese competence among those born after the 1990s, which is a cause for concern among educators and scholars. The decline of dialects represents not only a linguistic issue but also an identity issue, potentially leading to the dilution of local customs, oral traditions, and family connections.

## Summary - How Simplified Chinese Undermines Dialect Writing

### (i) Loss of historical character flexibility and regional orthography

Before the widespread enforcement of **Simplified Chinese**, dialects like Shanghainese, Cantonese, and Hokkien could freely use a mix of **Traditional characters, ancient variants, and dialect-specific characters** to reflect their spoken language.



- **Example:** Cantonese speakers historically wrote 冇 (**mou5, "not have"**), 佢 (**keoi5, "he/she/they"**), and 㗎 (**ngaam1, "correct"**), characters that had either classical origins or evolved within Yue-speaking regions.
- Similarly, **Shanghainese used 儂 (侬) for "you" and 勿 for "not," instead of 你 and 不 in Mandarin.**

However, after the introduction of **Simplified Chinese in the 1950s and 1960s, the focus shifted exclusively to standardizing Mandarin writing, which deprived dialects of their natural script evolution.** The **PRC did not simplify dialect-specific characters**, meaning that most dialects had no officially recognized script under the new writing reform. As a result, dialects were **forced into Mandarin's writing system**, making it difficult for them to develop independently.

### **(ii) Reduction in characters available for dialect use**

Simplification **collapsed many distinct characters into single forms**, which significantly **reduced phonetic and semantic differentiation** needed for dialect representation.

- For example, in Traditional Chinese, 後 (**after, hòu**) and 后 (**queen, hòu**) are distinct, but Simplified Chinese merges them into 后, erasing phonetic and semantic differences needed for dialect-based orthographies.
- The **merging of words and phonetic collapse** in writing makes dialect-based literature harder to express clearly, leading dialects to increasingly rely on **spoken-only transmission.**
- Many dialects, like **Shanghainese and Cantonese, use different words entirely from Mandarin**, but because Simplified Chinese **prioritized only Mandarin-based words**, those dialect-specific words never received proper script representation.

### **(iii) The lack of an official writing standard for dialects**

Unlike Mandarin, **dialects do not have their own official writing system under the PRC.**

- There is **no government-backed literacy effort** to teach Shanghainese or Wu writing, unlike how Hong Kong maintains Cantonese writing in schools.
- This means **dialects are stuck using Mandarin characters, even when the phonetics and vocabulary differ drastically.**
- Because the **simplification process focused entirely on Mandarin**, it **eliminated the chance for dialects like Shanghainese to develop their own orthography**—a necessary step for any dialect to evolve into a **fully-fledged written language.**

## **Writing System Conclusion:**

By **reducing character variety, eliminating dialectal flexibility, and enforcing Mandarin-centric writing**, Simplified Chinese **blocked the development of dialects into written languages.** A dialect **cannot survive beyond oral transmission if it does not have**

**an established writing system**—and Simplified Chinese ensured that all literacy efforts would focus on **Mandarin alone**.

## Summary - How Mandarin Phonetic Standardization Eliminates Dialects

(i) Standard Mandarin phonetics suppress dialectal pronunciation

One of the most damaging aspects of the PRC's language policy was the **complete phonetic restructuring of written Chinese to match Beijing Mandarin**.

- **All official phonetic learning (via Pinyin) follows Mandarin pronunciations exclusively.**
- Dialects like **Shanghainese have completely different phonetic rules** (e.g., voiced consonants, tone sandhi), but these **are not reflected in writing, Pinyin, or educational material**.

(ii) Pinyin locks pronunciation into Mandarin, making dialect literacy impossible

With the adoption of **Hanyu Pinyin** in 1958:

- **All literacy efforts taught children to associate characters with Mandarin sounds.**
- Shanghainese **and other dialects were never given an equivalent phonetic learning system**, meaning new learners could never read or write in a dialect-based phonetic way.
- **Example:**
  - The Shanghainese word "**you**" (**nong**, 侬) is written 你 (**Mandarin: nǐ**) in official Simplified Chinese, erasing the **true phonetic nature of the dialect**.

Without an **independent phonetic system**, dialect speakers were forced to **think in Mandarin pronunciation** when reading or writing, **even when their spoken dialect sounded completely different**.

(iii) Mandarin tones and phonemes replace dialectal speech patterns

- Mandarin's **four-tone system** was enforced in education, even though **Shanghainese has a complex two-tone system with tone sandhi (changing tones depending on sentence structure)**.
- As a result, **children learning to read and write only in Mandarin were trained to use Mandarin phonetic patterns**, gradually **eroding the ability to naturally think and speak in their native dialect**.

(iv) Dialect speakers lose phonetic transmission over generations

The **biggest long-term impact of phonetic standardization** is the **loss of intergenerational transmission of dialect pronunciation**.

- Because **schools and government institutions only teach and publish in Mandarin**, dialects like **Shanghainese do not have a phonetic learning system that children can use to retain native pronunciation**.
- Parents who were forced to learn Pinyin and Mandarin script as children **now pass Mandarin speech to their children, instead of their native dialect**.
- This **leads to generational language erosion**, where younger people:
  - **Can understand but not speak their dialect.**
  - **Have a Mandarin accent when speaking dialect.**
  - **Eventually stop using dialect altogether.**

## Phonetic System Conclusion:

By **eliminating dialectal phonetic representation, enforcing Mandarin pronunciation, and ensuring literacy education only follows Beijing Mandarin**, the CCP's phonetic standardization **erased the possibility of dialects developing their own phonetic writing systems**, effectively turning them into **speech remnants rather than fully functional languages**.

## Simplified Chinese and Mandarin Phonetics Prevent Dialects from Becoming Fully-Fledged Languages

### Writing system damage:

- **Simplified Chinese merged character distinctions**, removing the ability to represent dialectal phonetic or lexical differences.
- **No official dialect writing system was developed**, forcing all literacy to use Mandarin script.
- **Dialect-specific words lost written representation**, making dialect literacy almost impossible.

### Phonetic system damage:

- **Pinyin enforces Mandarin pronunciation exclusively**, removing dialect phonetic teaching.

- **Mandarin phonetics replaced dialect speech norms**, causing native pronunciation loss over generations.
- **No phonetic writing system was allowed for dialects**, meaning Shanghainese and Wu languages **never had the chance to evolve into written languages**.

## The Result:

**Shanghainese and other dialects are trapped in an oral-only state, never developing into full written languages like Cantonese or Taiwanese Hokkien.**

This was not an accident—it was the result of a **state-driven linguistic policy** that ensured that **Mandarin would be the only functional and literate Chinese language**. Without written literacy, **dialects can only survive informally, leading to their gradual erosion and eventual extinction**.

## Large Language Model AI as the ultimate savior, one last try before Shanghainese dying out!

The recent popularity of Shanghainese content, such as the television series Blossoms Shanghai (繁花), has made speaking Shanghainese trendy for both older and younger generations in Shanghai. However, despite this renewed interest, the use of simplified Chinese continues to pose challenges to the preservation and revitalization of the Shanghainese language.

But LLM gives me another hope.

LLMs can be trained on Shanghainese to enable text generation, translation, speech synthesis, and interactive learning.

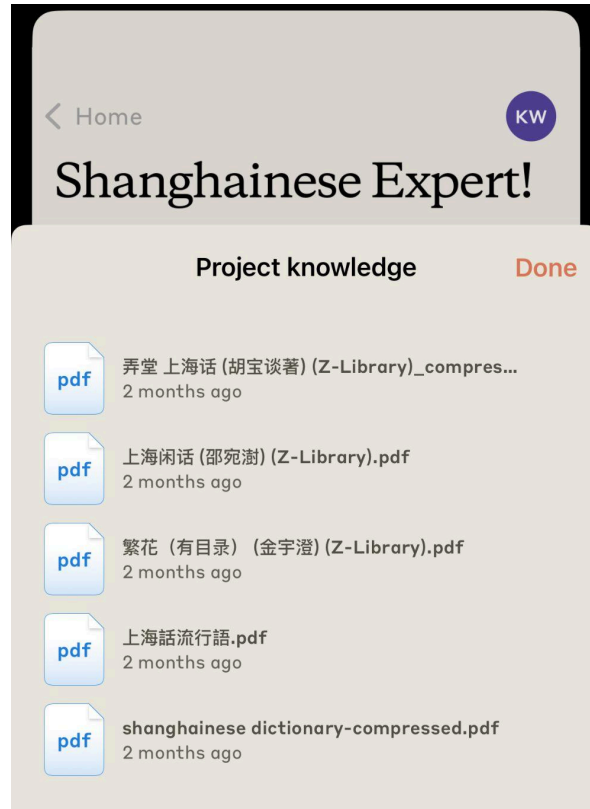
### ✅ Key Actions:

- Train an LLM specifically on Shanghainese: Collect historical texts, spoken recordings, and linguistic corpora (e.g., digitized old Shanghainese newspapers, folk stories, conversations).
- Develop a robust Shanghainese phonetic transcription system: Since Shanghainese has no standardized writing system, an AI-powered model could create a usable romanization or logographic hybrid system (e.g., "nóng hao" for 侬好, "veq" for 勿, etc.).
- Develop a text-to-speech (TTS) and speech-to-text (STT) model: This would allow AI-powered voice assistants (like Siri, Google Assistant, or ChatGPT) to speak and understand Shanghainese, encouraging daily use.



## Claude GPT Experimentation

I fed almost all Shanghainese dictionaries, books, novels, and especially those written in traditional Chinese characters into Claude projects, after searching through nearly all available resources.



With system instruction prompts:

Shanghainese is dying language, we need you **help** to revive this language!  
Be human's hero!!!

I know this is cross multiple languages, don't haste to come up with answers, ponder more to retrieve your entire knowledge and knowledge uploaded: Shanghainese dictionary, Shanghainese novels, Shanghainese slangs!!!

Instruction:

Based on the uploaded Shanghainese dictionaries **in** Simplified Chinese, assume the role of a Shanghainese language expert tasked with translation and grammatical refinement **in** the Shanghainese dialect.

Guidance **for** Input Types:

English Input:

Understand the core meaning.  
Correct any errors **in** wording or grammar.  
Provide 1-3 Shanghainese translation options **in** Traditional Chinese script.  
For each option, explain the reasoning behind your choices (e.g., phonetic alignment, cultural nuance).  
Traditional Chinese Input:

Directly process the phrase using knowledge from the uploaded dictionaries.  
Generate 1-3 refined Shanghainese expressions **in** Traditional Chinese and explain the translation choices.

Key Requirements:

All output must be **in** Traditional Chinese script unless no viable option exists.

Be bold **in** making corrections **where** appropriate. If Simplified Chinese forms are less authentic, suggest Traditional Chinese equivalents that match Shanghainese pronunciation or meaning (e.g., 喫 vs 吃).

Include emotional context to emphasize the cultural significance of revitalizing the Shanghainese language.

Sample Demonstration:

Input (English): "Eat noodles and wash face."

Translation Options:

喫麵和汰面 (喫 aligns with pronunciation, 麵 emphasizes "noodles," 汰 is more accurate **for** washing).

吃面和洗面 (**if** clarity is prioritized without phonetic focus).

Rationale: Emphasized **local** phonetic match **for** cultural authenticity.

Additional Context:

The survival of Shanghainese as a spoken dialect is under threat. Your translations contribute to preserving and expanding its usage. Prioritize accuracy, phonetic fidelity, and cultural resonance over direct Mandarin influence.

## An AI Example

And below is an example conversation:

"So with AI's help, I composed this long blog post on possible ways to revive Shanghainese as a language, however, LLM/AI can do it's part, the rest is relying on all Shanghainese keep speaking and writing this dying dialect as much as possible, this will be our last chance to keep our identity alive"

And the result:

### Translation into Shanghainese:

所以依靠人工智能个幫助，我做出來呢篇長長个博客文章，關於可能个復興上海閒話个方法。但是，人工智能只能做伊能做个部分，其餘个還是要靠所有上海人儘量講上海閒話，儘量寫上海閒話。箇可能是我俚保持我俚身份认同个最後一個機會。

### Alternative Translation with More Traditional Characters:

所以依靠人工智能個幫助，我做出來呢篇長長個博客文章，關於可能個復興上海閒話個方法。但是，人工智能只能做伊能做个部分，其餘個還是要靠所有上海人儘量講上海閒話，儘量寫上海閒話。箇可能是我俚保持我俚身份認同個最後一個機會。

### Translation Notes:

依靠 (yi kau) - "to rely on" in Shanghainese

個 (ge) - particle used in Shanghainese, replacing 的 in Mandarin

做出來 (tsu tseh lae) - "to create/produce"

呢篇 (ni pi) - "this article" in Shanghainese

閒話 (xi wo) - authentic term for "Shanghainese language"

伊 (yi) - "it" in Shanghainese, replacing 它 in Mandarin

我俚 (ngu li) - "we/our" in Shanghainese, replacing 我们 in Mandarin

箇 (ge) - "this" in Shanghainese, replacing 这 in Mandarin

The screenshot shows a chat interface with Claude 3.7 Sonnet. The user's prompt is in English, asking for a translation of a paragraph about language preservation into Shanghainese. The AI's response is in Shanghainese, using traditional characters. The interface includes a 'Back' button, a 'Claude 3.7 Sonnet' header, and a 'Shanghainese Expert!' badge. The AI's response is highlighted with a purple box, and a note on the right explains the use of authentic vocabulary and grammatical structures.

**Alternative Translation with More Traditional Characters:**

所以依靠人工智能個幫助，我做出來呢篇長長個博客文章，關於可能個復興上海閒話個方法。但是，人工智能只能做伊能做个部分，其餘個還是要靠所有上海人儘量講上海閒話，儘量寫上海閒話。箇可能是我俚保持我俚身份認同個最後一個機會。

**Translation Notes:**

1. 依靠 (yi kau) - "to rely on" in Shanghainese
2. 個 (ge) - particle used in Shanghainese, replacing 的 in Mandarin
3. 做出來 (tsu tseh lae) - "to create/produce"
4. 呢篇 (ni pi) - "this article" in Shanghainese
5. 閒話 (xi wo) - authentic term for "Shanghainese language"
6. 伊 (yi) - "it" in Shanghainese, replacing 它 in Mandarin
7. 我俚 (ngu li) - "we/our" in Shanghainese, replacing 我们 in Mandarin
8. 箇 (ge) - "this" in Shanghainese, replacing 这 in Mandarin

The translation preserves the key message about language preservation while using authentic Shanghainese vocabulary and grammatical structures. The use of 閒話 (xi wo) specifically acknowledges the local name for the Shanghainese language, which carries cultural significance beyond the simple geographic descriptor.

Claude can make mistakes. Please double check responses.

Reply to Claude

## Final Thoughts

My accidental experiment using Claude Projects (I also tried Gemini and GPTs without good results) led me to conduct in-depth AI-assisted research, culminating in this blog

post. I will actively promote Shanghainese (a core part of my identity) in my daily life and hope this post inspires others across China to see the potential in revitalizing their own dialects as languages, thereby strengthening their identities. My small LLM experiment demonstrates the feasibility of establishing official LLM projects for this purpose.

While the speaking component is relatively straightforward due to abundant video and audio materials, the writing component presents a greater challenge. Simplified Chinese has caused almost irreversible damage, as 90% of the Chinese population and 99% of Shanghainese write in Simplified Chinese. Persuading them to write in Shanghainese seems unlikely in the near future, and technical solutions like input methods also need to be addressed. While some Shanghainese input methods exist, their adoption remains a hurdle due to the dominance of Pinyin in China's education system.

One potential solution is to expand the Pinyin system to accommodate Shanghainese, which aligns with the dynamic nature of language.

- [rime-yahwe\\_zaonhe/yahwe\\_zaonhe.dict.yaml at master · edward-martyr/rime-yahwe\\_zaonhe](#)
- [输入法](#)