

---

# SINO-PLATONIC PAPERS

Number 352

August, 2024

---

## Dramatic Transformations of Sinography in East Asia and the World

Edited by  
Victor H. Mair

Victor H. Mair, Editor  
*Sino-Platonic Papers*  
Department of East Asian Languages and Civilizations  
University of Pennsylvania  
Philadelphia, PA 19104-6305 USA  
vmair@sas.upenn.edu  
www.sino-platonic.org

# SINO-PLATONIC PAPERS

FOUNDED 1986

*Editor-in-Chief*

VICTOR H. MAIR

*Associate Editors*

PAULA ROBERTS

MARK SWOFFORD

*ISSN*

2157-9679 (print) 2157-9687 (online)

SINO-PLATONIC PAPERS is an occasional series dedicated to making available to specialists and the interested public the results of research that, because of its unconventional or controversial nature, might otherwise go unpublished. The editor-in-chief actively encourages younger, not yet well established scholars and independent authors to submit manuscripts for consideration.

Contributions in any of the major scholarly languages of the world, including romanized modern standard Mandarin and Japanese, are acceptable. In special circumstances, papers written in one of the Sinitic topolects (*fangyan*) may be considered for publication.

Although the chief focus of *Sino-Platonic Papers* is on the intercultural relations of China with other peoples, challenging and creative studies on a wide variety of philological subjects will be entertained. This series is *not* the place for safe, sober, and stodgy presentations. *Sino-Platonic Papers* prefers lively work that, while taking reasonable risks to advance the field, capitalizes on brilliant new insights into the development of civilization.

Submissions are regularly sent out for peer review, and extensive editorial suggestions for revision may be offered.

*Sino-Platonic Papers* emphasizes substance over form. We do, however, strongly recommend that prospective authors consult our style guidelines at [www.sino-platonic.org/stylesheet.doc](http://www.sino-platonic.org/stylesheet.doc).

Manuscripts should be submitted as electronic files in Microsoft Word format. You may wish to use our sample document template, available here: [www.sino-platonic.org/spp.dot](http://www.sino-platonic.org/spp.dot).

All issues of *Sino-Platonic Papers* are free in PDF form. Issues 1–170, however, will continue to be available in paper copies until our stock runs out.

Please note: When the editor goes on an expedition or research trip, all operations may cease for up to three months at a time.

*Sino-Platonic Papers* is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/2.5/> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

# Dramatic Transformations of Sinography in East Asia and the World

Victor H. Mair, Editor  
University of Pennsylvania

## FOREWORD

The three papers in this collection were written for my “Language, Script, and Society in China” course during the fall semester of 2023. All three of them are concerned with radical changes made to Sinographic script during its adjustment to modernity.

Aleena Parenti shows how, during medieval times, Vietnamese acquired a written form known as *chữ Nôm* (lit., “writing of the south”) under the impact of the Chinese script, which in turn yielded to romanization brought by the French colonialists during the late nineteenth and early twentieth centuries. That resulted in the current Vietnamese alphabet known as *chữ Quốc ngữ* (“writing of the National Language”).

Zhaofei Chen’s paper reveals the tremendous impact of the vernacularizing influence of Western missionaries during the late imperial period of Chinese history (from the seventeenth to the twentieth centuries). Through translation, romanization, and their own writings, the missionaries contributed massively to the eventual demise of Literary Sinitic / Classical Chinese as the official written language, at the hands of Modern Standard Mandarin.

Yifei Yang explores how the Japanese development of emoji 絵文字 (lit., “picture writing”) has escaped the confines of any particular language and, as pictograms, logograms, ideograms, and smileys, can be adopted into the writing of any language. Emojis are widespread on social media, are especially favored by young people, and are by no means limited to East Asia. 😊!

Taken all together, these three papers presage tumultuous developments in the further evolution of Sinography during the rest of this century and beyond.

# Chinese Characters in the Vietnamese Language

Aleena Parenti

University of Pennsylvania

## ABSTRACT

Like many of the other East Asian countries in China's sphere of influence, Vietnam went through a period of using Chinese characters to represent the indigenous language. A millennium of Chinese rule and a native language much closer to Chinese lexically and morphologically led to a character system significantly different from those that developed in Korea and Japan. While most of Vietnamese written history used Classical Chinese characters, after independence in 939 CE, a system of writing Vietnamese words with Chinese characters developed called *chữ Nôm* (*chữ* corresponding to modern Mandarin 字 *zì* 'character' and *Nôm* to 南 *nán* 'southern'). This essay explores the introduction of Chinese characters into Vietnamese government and culture, the formation of this new script created just for writing Vietnamese, and its downfall as it was replaced by the Romanized script "chữ Quốc ngữ" (lit. 'script of the National language'). *Chữ Nôm*'s odd position as the script of the educated and elite—but not *too* educated and elite—will also be discussed, as it paved the way for the script to be easily overcome by its Romanized rival. Modern scholarship, such as in Hannas (1997), has used Vietnam's tale of successfully adopting the Romanized script as a lesson for how China could do the same, a stance that will be explored in the conclusion, as well as other lessons this history can unveil.

## THE INTRODUCTION OF CHINESE CHARACTERS TO VIETNAM

In 111 BCE the Han army marched into the area referred to as Nan Yueh (“South of the Yue”) and defeated the prime minister Lǚ Gia’s army, starting the thousand-year colonization of Vietnam by Chinese forces (Lien & Sharrock 2015). During occupation the area was ruled by Chinese officials and the sole language of rule was Chinese. Residents of the colony spoke an early version of Vietnamese, about which little is known because it was not written down and was subsequently changed so much by contact with Chinese. Vietnamese is a Mon-Khmer language that gained tones through contact with Thai or southern Sinitic languages, and while today it is largely morphologically monosyllabic, Phan (2020) argues that for much of its history Vietnamese was like its relative Khmer in possessing sequisyllabicity and developing monosyllabism after extended contact with Chinese. While Chinese was the language of administration, that did not mean the peasant class was exempt from its influence. During the chaos of Wang Mang’s rule in 9–23 CE a significant number of Chinese refugees of all social classes fled to Vietnam, leading to the foundation of Confucian schools in the area and further integrating Vietnam with the rest of the Chinese empire (DeFrancis 1977).

With this close relationship to China it is understandable that a large amount of the Vietnamese lexicon comes from Sinitic languages. There are even instances of the same word being introduced into the language multiple times throughout history as the earliest introductions become so naturalized that they are no longer associated with the Chinese word (Hannas 1997). It is estimated that about a third of the Vietnamese lexicon is of Chinese origin, and when sampling academic writing this number can jump to almost 60% (DeFrancis 1977). It is not a surprise, then, that a system developed to write this language was based on Chinese characters. What is more remarkable is that this script was not commonly used until centuries after its creation due to the prestige and administrative use of Sino-Vietnamese.

## SINO-VIETNAMESE

Chinese colonization of Vietnam ended in 939 CE, after the collapse of the Tang dynasty, when Ngô Quyền defeated the Southern Han troops and founded the Ngô dynasty (Lê & O’Harrow 2007). This did not end Chinese influence, however, as Vietnam maintained a vassal relationship with its neighbor, and

classical Chinese was still seen as the prestige language in administration and among the elite. As the government was not under direct Chinese control during this time, officials did not know how to speak Chinese anymore and instead read the characters with Vietnamese phonology. This created a variety that was not intelligible to Chinese or Vietnamese speakers at the time, and thus was a special administrative dialect referred to as Sino-Vietnamese. These pronunciations have been used by scholars to reconstruct Old and Middle Chinese, as Vietnamese retains the tone system that was present in Middle Chinese and has since been lost in that language (Hashimoto 1978).

This was the main way of writing for the upper class for centuries, with works such as Confucian-influenced histories and poetry aligned with Chinese prosody showing a close relationship with Chinese literary tradition at this time. As Sino-Vietnamese was a dialect of Chinese above all, it took years of study to understand and was only used by less than 5 per cent of the population (DeFrancis 1977). As the nation grew and documents had to be read aloud for peasants, the need for a writing system that captured the Vietnamese language grew; the need eventually was met with the formation of chữ Nôm.

## THE BIRTH OF CHỮ NÔM

As chữ Nôm literature was suppressed many times in Vietnam's history, it is difficult to say when it originated. DeFrancis (1977) argues that its origins may lie in attempts to write Buddhist terms using Chinese characters, giving people exposure to the idea that Chinese characters can be used as solely phonetic signs. This importation of Buddhism occurred in the second century CE, and it wasn't until much later, in the eighth century, that we find written evidence of phonetic signs used for Vietnamese, in the title “布蓋大王” assumed by national hero Phùng Hưng. In this title, 布 ‘cloth’ and 蓋 ‘to cover’ together do not make any sense, but their pronunciations (reconstructed by Karlgren (1957) as *puo* and *kai*) most likely were a reference to Vietnamese *bố* and *cái*, meaning “father and mother.” A few more official titles and steles with village names can be found following this pattern, up until 1282, when Nguyễn Thuyên wrote a poem in chữ Nôm that was then thrown into the Red River in order to chase away a crocodile (Hannas 1997), marking the beginning of Nôm literature. This is said to be the first

known poem written in the script, and after that there is a large gap until a full work in the script is seen again. How did this script develop, and why is there so little early evidence of it?

#### THE CREATION OF CHỮ NÔM CHARACTERS

These native characters, specifically created for Vietnamese, were constructed in five different ways: phonetic match, semantic match, semantic-phonetic construction, double semantic construction, and the importation of the Chinese word and character into Vietnamese. The last category of importing a Chinese word can be seen through the character 本 (*bản*) ‘root; foundation’—the character is pronounced as ‘bến’ in modern Mandarin with the same meaning. This same character could also be used for a semantic match, used to represent the Vietnamese word *vốn* ‘capital, funds’ (Hannas 1997). Immediately some problems can be seen in this system, most glaringly that the same character can be used to represent two words that are homophones in Chinese but not in Vietnamese. One solution for this ambiguity was to put a diacritic on characters that are more “nativized” (this is in quotes, as in this example, because for many Vietnamese words, the more “native” form was still originally imported from Chinese, just so long ago that its origins have been forgotten), these diacritics often being *cá* 个 (variant 个) and *ngáy* 个 (Collins 2017).

While sometimes falling into the same trap of allowing one character to represent two different words, phonological matches were fairly straightforward like their semantic counterparts, such as the character 沒 (Mandarin: *méi*; Vietnamese: *một*) for ‘one,’ 固 (Mand: *gù*; Viet: *có*) for ‘have,’ and 埃 (Mand: *āi*; Viet: *ai*) for ‘who’ (Hannas 1997, 80). For some of these it may seem that the connection to the Mandarin pronunciation is tenuous at best, but it is important to remember that the Vietnamese at the time were hearing Old or Middle Chinese, and most likely southern Yue varieties, which sound very different from modern-day Mandarin.

The semantic-phonetic constructions are where chữ Nôm–specific characters appear. These include 𠄎 for *ba* ‘three’ (left component phonetic, right semantic) and 𠄎 for *mẹ* ‘mother’ (left component semantic, right phonetic). Most constructed characters fell into this category, although there were some that were “compound ideographic” such as 𠄎 (天 ‘sky, heaven, day’ + 上 ‘up’) for *giời* ‘sky, heaven’ and 𠄎 (亡 + 失) for *mất* ‘lose,’ made of two different characters with the meaning ‘to lose’ (Li 2020). A final, least common original category of characters are simplified Chinese characters used



to represent unrelated Vietnamese words. For example, the Nôm character 𠵹 *áy* ‘that, those’ comes from a simplified form of the Chinese character 衣 ‘clothing,’ pronounced \*ei in Ancient Chinese (Hannas 1997), and the simplified version 邈 *lạ* ‘strange’ from 羅 ‘net,’ pronounced \*la in Middle Chinese (Karlgren 1957). It is clear that these creations were phonologically motivated, still clearer when we look at what the Chinese pronunciation would have been at the time of importation.

Here we can use this knowledge of how chữ Nôm characters were constructed to look at the characters for the script itself: 𠵹喃. The first word, *chữ*, comes from Middle Chinese *dzi* ‘character, writing’ (*zì* in Modern Mandarin) and is written with many different characters, such as 𠵹, 𠵹, 𠵹, and 𠵹. *Nôm* comes from *nam* ‘south,’ but the character chosen has a more controversial history. It is traditionally depicted as 喃, with the semantic element of 口 ‘mouth’ and the phonetic element of 南, but as it came from the Chinese homophone meaning ‘babble, murmur, speak indistinctly,’ some Vietnamese scholars used a character with a left-hand semantic element 𠵹 to change the literal meaning from “southern mouthings” to a less derogatory “southern characters” (DeFrancis 1977). Depending on the author of a text, a combination of different first and second characters may be used, showing the variability in this writing system over a phrase as basic as its name.

This writing system was never standardized, which leads to questions of how intelligible it was between writers. In the semantic-phonetic examples above there is no way of knowing from just the character which portion is semantic and which one is phonetic, and in chữ Nôm the elements could be in either order. From a distance it may seem as if this system was not governed by any rules and was just a free-for-all in creating characters. Nguyen Van Huyen (1944) rejects this idea, stating that this theory is just a superficial understanding of a system that had quite rigid rules and was mutually intelligible to its users.

## USE AND PERCEPTIONS OF CHỮ NÔM

Since the government continued to exclusively use Sino-Vietnamese and Classical Chinese characters, chữ Nôm remained in the realm of poetry and folk stories. As it provided a way for the language of the general populace to be written, it thrived during heights of national consciousness—and it was banned if there was ever a risk to the throne. While its first uses in literature were during the Trần dynasty (1225–

1400), the script's first period of prosperity was during the short-lived Hồ dynasty (1400–1407), when the founder Hồ Quý Ly aimed to make chữ Nôm the language of civil service examinations and translate Confucian texts into Nôm for study. He saw Confucians as “mere twaddlers, pillagers of texts, without any contact with reality” and wanted to break the power they had by making their texts available to the public (DeFrancis 1977).

In 1406 the Ming army, invited by loyalists to the dynasty Hồ Quý Ly overthrew, invaded Vietnam and brought an end to this dynasty, ushering in a twenty-year period of Chinese rule. Before even setting foot in Vietnam, the policy set forth by the emperor Yongle was to “burn all written and printed materials within Annam, except for Buddhist and Daoist texts, and anything that promotes Vietnamese rites and customs” (Lien & Sharrock 2015). Any Vietnamese stele was to be destroyed, and entire histories compiled by past dynasties were taken to China. This resulted in most works in Nôm being lost. Combined with the disdain of Confucian scholars who held had power for much of Vietnamese history, it is no wonder the growth of Nôm literature was stunted (DeFrancis 1977).

In the remaining Vietnamese dynasties, chữ Nôm became more and more popular, hitting its zenith in the 18th and 19th centuries. The most famous poem in Vietnamese literature, *The Tale of Kiều* (1820) was written in Nôm, and many more that were originally written in classical Chinese were translated into Nôm. There was at least one person in most villages that could read the script and would be able to read aloud literature and messages from the central government, which cemented Nôm's role in folk stories while leading to intermittent fear from the government of rebellious ideas being spread too easily (Marr 1984). This can be seen in imperial edicts from 1663, 1718, and 1760 calling for the censorship of writings in chữ Nôm, meaning there was very little time when Nôm was able to completely thrive, and by the time it did, French colonization was on the horizon.

While all of these are political reasons a why chữ Nôm did not have acceptance towards the end of imperial Vietnam, Phan (2020) cites a linguistic reason: the presence of sequisyllabicity and consonant clusters until the seventeenth century. Like other Mon-Khmer languages, such as Khmer itself, Vietnamese demonstrated sequisyllabicity, with most words being two syllables and thus two characters long, leading literati trained in classical Chinese (a more monosyllabic language) to see it as not “worthy” to be written in characters. It wasn't until the rise of Middle Vietnamese in the seventeenth century that the language became predominantly monosyllabic through Sinitic contact, although at

this point it still contained consonant clusters such as 'tl,' 'bl,' and 'kl,' which it later lost. This movement towards being monosyllabic made the language more like Chinese, more "proper," and easier to write using chữ Nôm, leading to a huge increase in usage. Now Phan (2020) does admit that this certainly is not the only reason for this increase, as institutional support is incredibly important and being monosyllabic is not a prerequisite for using characters, an idea clearly illustrated with Japanese and Korean. In Phan (2014), though, he notes that before this point chữ Nôm was just an extension of literary Sinitic, and it was at this time that it came to truly represent the spoken language.

Through these ups and downs of persecutions and literary development it is once again important to note that chữ Nôm was never the language of prestige and almost all those who knew it also knew classical Chinese. Sino-Vietnamese remained the language of administration, trade, and high culture during this time, and it was the variety associated with nationalism until the late nineteenth century. The only thing that was able to dethrone classical Chinese characters and bring Vietnamese into administration was French colonization and the introduction of the romanized (*chữ Quốc ngữ*) script (Lê & O'Harrow 2007).

#### ROMANIZED VIETNAMESE (CHỮ QUỐC NGỮ)

In 1651, Alexandre de Rhodes published a Latin, Portuguese, and Vietnamese dictionary, the latter written with a romanized script. De Rhodes most likely did not invent this whole system himself, instead taking elements from many different romanization systems that were used by missionaries like himself to keep notes and learn the language. This is exactly what this script was used for after the publication of his dictionary, and he never intended to teach native speakers about the script or replace chữ Nôm. The only documents used in this early system are notes from clergy members and instructions on teaching Vietnamese, although it is unknown how many unpublished documents there are from the time before French colonization (DeFrancis 1977).

France was motivated to ban chữ Nôm and classical Chinese characters to distance Vietnam from its past with China and limit a possible source of alliance. While there was resistance from the educated elite, as less than 5% of the population was literate at the time, there was not a large base to fight when it came to eliminating chữ Nôm (Hannas 1997). Remarkably, some of the strongest resistance

came from other French people, as they were concerned that the Vietnamese would be allowed to continue to speak their native language and not switch completely to French. Perhaps French being established as the language of all administration at the time was able to appease these critics, as during colonization knowing French was necessary for any high-level position (Marr 1984). The introduction of chữ Quốc Ngữ (lit. “script of the National Language”) just provided a way for the less educated to read and write, something that was originally used in newspapers and literature.

While chữ Quốc Ngữ was enforced by the French government and had strong institutional support, it was like chữ Nôm in that it was not regulated or standardized. Instead, through more and more publications, the script started to refine itself “through adaptation, imitation, and decentralized decision making” (Hannas 1997). This development through ground-up processes makes the story of chữ Quốc Ngữ less exciting, but it is remarkable given it is the only language to fully switch from Chinese characters to a Latin script, despite being structurally more similar to Chinese than Japanese and Korean are. How was the adoption able to take place?

Hannas (1997) writes that Vietnamese had a grammatical advantage as measure words always accompany a noun, so homophones are able to be differentiated easily. Socially, while the Vietnamese were not happy to be under French control, they were ready to break free from Chinese influence. After being under Chinese control for a thousand years and with intermittent intrusions, the Vietnamese people were ready to break away from this previous colonizer, even if it meant using the script of a current one. Finally, as previously mentioned, there were fewer people who knew chữ Nôm compared to kanji and Hanja, in Japan and Korea, respectively. The educated class was split along the lines of those who used chữ Nôm and those who used classical Chinese, so there was little unity in resistance. Finally, in the system forming from the ground up, it was continuously improving to become clearer to the users, and gradual changes meant less resistance. Hannas uses Vietnamese as a case demonstrating that it is possible to move from a character-based system to a Latin script, while leaving the language easily decipherable. He argues that Vietnamese has rarely been studied on this front previously, due to its having been seen as the most “barbaric” East Asian language, falling in line with the “Principle of Substandard Southern Speech,” the idea that the further south the variety is spoken the less prestige it has. This bias can be seen around the world, from Mandarin having more prestige than Cantonese to Southern American English being frequently maligned. If linguists and reformers look past this deeply

ingrained bias that has been applied to this Southeast Asian language, then Vietnamese can offer an important lesson on how a Latin script can be adapted, personalized, and become part of a national identity.

## CURRENT DEVELOPMENTS IN CHỮ NÔM AND ITS MODERN CONSEQUENCES

Almost eighty years after the start of French colonial rule and the ban on chữ Nôm, groups raised an alarm at the idea of centuries of literary tradition being lost as knowledge of Nôm faded, and it was impossible to transcribe these characters on computers. In 1999 the Vietnamese Nôm Preservation Foundation was created by James Đỗ Bá Phước, Ngô Thanh Nhàn, and John Balaban, with the goal of preserving existing texts, creating software tools for writing, entering characters into unicode, and funding scholarship in the field. Although the organization dissolved in 2018, on its website it offers many historical dictionaries, texts, and fonts for computer input (“About the VNPF”).

Less official groups, such as ChuNom.org and the Han-Nom Revival Committee of Vietnam, have similar goals, with both providing dictionaries and scripts. ChuNom.org is community-run and offers three levels of lessons for learning chữ Nôm along with a corpus of readings (many of which come with a romanized transcription), machine translation, Phonetic Annotation using basic chữ Nôm characters, and an input method for these characters. The Han-Nom Revival Committee of Vietnam runs a wiki page entirely in chữ Nôm, a page that can illustrate how hard it is to read the script on a standard computer. Here is a small part of the front page that explains a little about Nôm:

<sup>hán</sup>  
漢 <sup>nôm là</sup>喃 <sup>gi</sup>之?

---

<sup>chữ nôm</sup> 𣪠喃、<sup>quốc âm</sup> 𣪠得𣪠𣪠 <sup>quốc ngữ</sup> 𣪠國音哈國語、𣪠類文字  
語素-音節用抵𣪠𣪠𣪠越。低𣪠部𣪠得𣪠越  
造𣪠瀕𣪠𣪠漢、各部首、音讀吧義詞彙𣪠  
𣪠𣪠越。

Many of these characters are icon boxes that are not able to be read by this computer, and when typing this essay many characters had to be put in as pictures, as Times New Roman does not support many of them. It is important to note though that, during the writing of DeFrancis's (1977) and Hannas's (1997) books, almost all of the Nôm characters were written by hand, as there was no way of typing them at the time of publication. Through the advocacy of these groups thousands of Nôm characters have become digitally accessible, but not all computers are able to display even basic characters and components, making some documents completely illegible. Luckily, many of these websites have browser extensions and Unicode extension packs that include more characters, but their work towards complete coverage is far from done.

While the romanized Quốc Ngữ script is the only script now used to write Vietnamese, chữ Nôm and classical Chinese were able to influence current ideas of the language. Even though Vietnamese is a predominantly monosyllabic language, the rule of having spaces between every single syllable can cause confusion when depicting words that are two syllables. Sometimes this confusion is solved by capitalizing the first letter of the start of a word, but this is often only done for names or titles. Like their Chinese neighbors, the Vietnamese did not have a concept of “word” until rather recently, likely due to the nature of classical Chinese and chữ Nôm. When Quốc Ngữ encountered this problem, it simply put a space between every syllable, as if that group were a character. Hannas (1997) cites this as an area in urgent need of reform, and a relic of China's influence on Vietnamese orthography.

## CONCLUSION

Chữ Nôm's lack of standardization and government support, along with representing an often maligned Southeast Asian language, has made it the least studied of the Chinese character offshoots. Its story can provide important information on how character systems are adjusted to a changing language, how institutional support can aid or limit the growth of a script, and how a Latin script can replace characters and faithfully represent a largely monosyllabic, tonal language.

Chữ Nôm's ability to survive through several attempts to ban it, and its continued life through recreational groups that have gone so far as to create a whole wiki in the script, shows the power of demotic script for the arts and for ethnic and national identity. While it hasn't had the staying power of

Japanese kanji or Korean Hanja, it facilitated Sinitic importations into Vietnamese, was the medium of one of the most important periods of Vietnamese literature, and allowed information to spread to people throughout Vietnam. Despite its very brief mention in any book on Chinese characters, its importance to Vietnamese history and cultural identity should not be underestimated.

#### WORKS CITED

- Collins, L. (2017). *Proposal to Encode Two Vietnamese Alternate Reading Marks*.
- DeFrancis, J. (1977). *Colonialism and Language Policy in Viet Nam*. Mouton.
- Hannas, W. C. (1997). *Asia's Orthographic Dilemma*. Honolulu: University of Hawaii Press.
- Hashimoto, M. J. (1978). “Current Developments in Sino-Vietnamese Studies.” *Journal of Chinese Linguistics*, 6(1), 1–26. <http://www.jstor.org/stable/23752818>
- Huyen, Nguyen Van (1944). *La civilisation annamite*. Hanoi.
- Karlgren, B. (1957). *Grammata serica recensa*. [Stockholm]: [Museum of Far Eastern Antiquities].
- Lê, M.-H., and S. O'Harrow (2007). “Vietnam.” In A. Simpson, ed., *Language and National Identity in Asia* (415–441). Oxford University Press.
- Li, Yu. (2020). *The Chinese Writing System in Asia: An Interdisciplinary Perspective*. Routledge.
- Lien, V. H., and P. Sharrock (2015). *Descending Dragon, Rising Tiger: A History of Vietnam*. Reaktion Books, Ltd.
- Marr, D. G. (1984). *Vietnamese Tradition on Trial, 1920–1945*. University of California Press.
- Phan, J. (2014). “Rebooting the Vernacular in Seventeenth-Century Vietnam.” In B. A. Elman, ed., *Rethinking East Asian Languages, Vernaculars, and Literacies, 1000–1919*, 96–128. Brill.
- Phan, J. D. (2020). “Sesquisyllabicity, Chữ Nôm, and the Early Modern Embrace of Vernacular Writing in Vietnam.” *Journal of Chinese Writing Systems*, 4(3), 169–182. <https://doi.org/10.1177/2513850220937355>
- Vietnamese Nôm Preservation Foundation. (n.d.). “About the VNPF.” Retrieved December 19, 2023, from <http://www.nomfoundation.org/?uiLang=en>

# The Influence of Western Missionaries to China on Modern Chinese Language Development

Zhaofei Chen

University of Pennsylvania

## ABSTRACT

This study investigates the historical and social influences on the evolution of the Chinese language, focusing specifically during the Ming and Qing dynasties. It examines the significant role played by Western missionaries in shaping the modern Chinese language. The research reveals the ways their translation of religious texts and the development of effective language learning methods not only enhanced cross-cultural communication but also profoundly influenced modern Chinese phonetics, pedagogy, and comparative language studies. These advancements led to a more structured Chinese language system, demonstrating the dynamic interaction between linguistic evolution and cultural exchange. Despite the existence of ample resources, there has nevertheless been a lacuna in their amalgamation into a cohesive understanding, as they were previously disparate and unconsolidated. This study methodically categorizes the contributions of these Western missionaries, emphasizing the significance of cross-cultural interactions in the development of the modern Chinese language. It opens new pathways for further research into the transformation of the Chinese language within a multicultural framework.

**Keywords:** Western missionaries in China; Chinese language evolution; phonetic system development; language acquisition and translation; cross-cultural linguistic influence



## I. INTRODUCTION

With the backdrop of China's transition from the Ming and Qing dynasties to the Republic period, this essay examines the pivotal role of missionary activities in Chinese language development during the eighteenth and nineteenth centuries. Up to the late Qing dynasty, China was trapped in a quandary concerning its social and cultural identity. The modern history of China had experienced numerous impacts from Western culture. The overall history, customs, language, and literature of the Chinese nation had all been fundamentally shaken. The nation struggled amidst tremendous challenges, leading to a collapse in self-confidence. Concurrently, the people looked to intellectuals for a revolution in the nation's language and script. This motivation directly triggered the transformation of China's language in the modern era. This paper aims to explore the social integration between the East and the West, focusing on Western missionaries and analyzing their cultural activities in China, such as foreign language translation, methods of learning Chinese, and establishment of cultural media such as newspapers. The discussion will focus on the specific factors that influenced the formation of modern Chinese languages, the internal changes and development experienced by the language, and how the methods of language teaching evolved.

### A. HISTORICAL BACKGROUND OF EVANGELISM IN CHINA

In the latter years of the Ming dynasty, significant developments, such as the establishment of the Jesuit Order in France and breakthroughs in navigation technology, ignited Western interest in the Far East. This era saw Catholic missionaries, having studied Chinese, embark on their evangelistic endeavors in China. A notable figure during the early nineteenth century was Robert Morrison. Before his missionary journey, Morrison mastered the Chinese language in London under Yong Sam-Tak. His efforts in China, sponsored by the London Missionary Society, marked the first of their kind in the nineteenth century. Morrison's commitment to translating the Bible into Mandarin significantly helped to bridge the cultural and linguistic gap, making Christian scriptures accessible to the Chinese populace.

The mid-nineteenth century, a period shaped by the Opium Wars and subsequent treaties, marked a pivotal point in China's history. These events compelled China to open its doors to foreign influences, including a surge of missionaries from Western countries, particularly Great Britain. This era

witnessed a substantial increase in missionary activities and the founding of various Christian institutions.

During this time, the Taiping Rebellion, led by Hong Xiuquan—who professed to be the younger brother of Jesus Christ—unfolded. Although the rebellion was not directly spurred by Western evangelical efforts, it was deeply influenced by Christian ideology, integrating several Christian principles and texts in a unique, syncretized manner. This rebellion had a profound impact on Chinese society and its perception of Christianity, illustrating the complex interplay of Western religious ideologies within the Chinese sociopolitical landscape.

#### B. CHALLENGES FROM CHINESE LINGUISTIC POLICY AND DIVERSITY

During the Ming and Qing dynasties, Western missionaries in China faced considerable challenges due to the policies of the Chinese government, particularly under the Qing's "closed-door policy." This isolationist stance created a significant divide between China and the rest of the world, limiting the Chinese people's exposure to external information and imposing strict controls on foreigners. The Chinese, influenced by government and public pressure, were generally distrustful of Westerners and skeptical of their religious teachings. This widespread antiforeign sentiment among both the populace and officials made it exceedingly difficult for missionaries to operate.

In this conservative and isolated environment, traditional missionary methods proved largely ineffective. Scholars argue that adaptations were necessary for missionary work to resonate with Chinese conditions. These adaptations included a deep immersion into Chinese customs, history, and particularly language, which was crucial for understanding and appreciating the nuances of Chinese civilization and for teaching effectively.

The multifaceted nature of China, as a multilingual nation with a large population, presented a complex mix of challenges and opportunities. Political resistance from the Qing government, coupled with ideological differences rooted in Confucianism, posed significant hurdles. For instance, there were constraints on publishing religious texts in local languages, and in areas like Guangdong, prohibitions against teaching Chinese to foreigners further deepened the cultural and linguistic divide. These regulations were part of a deliberate effort to maintain separation between locals and Westerners. The

restrictive regulation of religious activities, requiring missionary material to be in Western languages, underscored the government's discouragement of missionary work.

Furthermore, Christianity, especially as propagated by foreign missionaries like Morrison, was often viewed with suspicion and faced persecution. It was perceived as a foreign influence and a threat to the traditional Chinese culture, which was predominantly Confucian and Daoist. This perception stemmed from the significant linguistic, cultural, and ideological differences between the West and China. Despite these challenges, Western missionaries brought with them valuable ideas, fostering a collision of Eastern and Western cultures that inspired future scholars in fields like language and script.

### C. SINOLOGY AND LANGUAGE STUDIES IN THE EIGHTEENTH CENTURY

The eighteenth century in Europe was marked by a fervent fascination with Chinese culture, known as "Chinoiserie," encompassing interest from princes, nobles, literati, and ordinary people. The translation of Confucian works by Jesuit missionaries played a significant role in this phenomenon, leading to the establishment of Sinology as an academic discipline. However, Sinology initially focused more on the study of Chinese civilization, history, and ideology, with less emphasis on language studies. It wasn't until missionaries arrived in East Asia that the shift from theoretical study to practical language application began.

Key figures like Seraphin Coureur, a notable French Sinologist and Jesuit missionary, made significant contributions to Chinese linguistic studies. His works, like the *Dictionnaire français-chinois*, published in 1908, focused on Mandarin expressions and played a crucial role in translating Chinese classical words for everyday use. These efforts not only benefited scholars but also made the Chinese language accessible to ordinary Chinese people.

## II. THE SHAPING OF CHINESE LANGUAGES BY WESTERN MISSIONARIES

The influence of missionaries on the Chinese language is evident in the development of its written forms and the construction of the phonetic system. During the Qing dynasty, when missionaries entered China, the socio-economic development of China had already stagnated, while the West had begun the

era of rapid industrial revolution. The huge economic disparity created a significant cultural chasm between China and the West. More frightening than economic backwardness was the struggle for national self-esteem hidden beneath disputes over diplomacy and etiquette. China's national power and authority were being challenged by Western society, and the threat to the primacy of China's millennia of culture led the Chinese to a state of doubt. This was clearly felt in the conflicts over religious cases that occurred throughout China. Eventually, under such circumstances, the Qing government ordered the lifting of the ban on Christianity. Even after the ban was lifted, however, the progress of missionary work was not as smooth as expected, mainly due to the more significant factor of cultural opposition. The Western cultural background underlying Christianity was incompatible with Chinese Confucian culture. Therefore, missionaries hoped to start from the linguistic level, but the Chinese language system was significantly different from Western scripts.

#### A. THE EVOLVEMENT OF CHINESE LANGUAGE BY MISSIONARIES

##### *Classical Chinese Version: Wen-Li*

The early translations of the Bible into Chinese during the late Qing period, particularly those by Joshua Marshman and Robert Morrison, are significant in the context of linguistic and cultural exchange. These translations not only introduced Christian religious texts to China but also marked a crucial point in the interaction between Western and Chinese cultures. Here are some key aspects: the first published Chinese version of the Protestant Bible was compiled and published by Joshua Marshman, a British Baptist missionary stationed in Serampore, India. This version's Gospel of Matthew and Gospel of Mark were published in 1810. It wasn't until 1822 that his *Old Testament* was published, marking the first complete Chinese translation of the Bible in history. In the book, Marshman referred to this version as the High Wen-Li translation. However, in its content, the High Wen-Li version appears to be influenced by classical Chinese language expressions.

In the beginning God created the heaven and the earth. And the earth was without form, and void; and darkness was upon the face of the deep. And the Spirit of God moved upon the face of the waters.

Genesis 1:2, King James Version

原始神創造天地，地未成，形陰，氣蘊於空虛幽邃之內。

Joshua Marshman, *Old Testament* (see Appendix)

The original God created the Heaven and the Earth, Earth was unformed, Yin (the reverse of Yang) Shape, with darkness covering the deep spirit enshrouded within the void and profound.

The character 氣 *qi* was used in the translation. In the Chinese tradition, *qi* is a fundamental concept that is often translated as “life force” or “energy flow.” In the context of traditional Chinese medicine, *qi* is believed to be the energy that circulates through the body along pathways known as meridians. The balance and flow of *qi* are essential for maintaining health; imbalances or blockages in *qi* can lead to illness. And Marshman used *qi* to indicate “Spirit of God,” which shows a combination of his cultural knowledge and his understanding of this character. In Morrison’s version, the word “pneuma” is translated as *ling* 靈 (spirit) in twenty-three verses, *feng* 風 (Spirit/spirit) in fifteen verses, and *shen* 神 (Spirit/spirit) in four verses (Liu, 2021). This adaptation of the term to represent the “Spirit of God” illustrates an effort to make Christian concepts more comprehensible within the Chinese cultural framework.

Robert Morrison’s translation, known as 神天聖書 (*Shen Tian Sheng Shu*), was published in 1823. Like Marshman’s, Morrison’s translation was also in Classical Chinese, but with even more pronounced use of classical expressions, such as the frequent use of 也 at the ends of sentences, for an explanatory tone. This demonstrates the missionaries’ proficiency and their ability to adapt Classical Chinese. From this translation, we can see that China was still under the thrall of the Classical version of texts.

神當始創造天地也。時地無模且虛。又暗在深之面上。而神之風搖動于水面也。

神天聖書 (see Appendix)

In the beginning, God created the Heavens and the Earth. The earth was formless and empty at this time, darkness was over the surface of the deep, and the Spirit of God was hovering over the waters.

The text is the beginning of the Book of Genesis from the 神天聖書 (*Shen Tian Sheng Shu*). Its style and wording are indicative of Classical Chinese, with its characteristic compact structure and use of grammar and phrases. For example, the term 當始, where 當 (*dāng*) typically means “should” and “ought to” and where 始 (*shǐ*) means “begin” and “start,” indicate that this term is a classical expression that differs from the spoken language.

The translations into Classical Chinese, while significant, initially had limited impact due to the high educational level required to read and comprehend this language. Classical Chinese, distinct in its grammar and structure from the spoken language, was predominantly a written form and never used in everyday speech. It was taught exclusively in schools and thus remained largely inaccessible to the general populace.

Most Chinese people at the time were peasants living in rural areas, often in poverty, with their livelihoods heavily reliant on agriculture. This socioeconomic context led families to prioritize farm work over formal education. Additionally, the traditional Chinese education system, particularly the imperial examination system, was oriented towards the study of classical Confucian texts. This system catered mainly to the upper class and those who could afford private tutors, focusing on content that lacked practical applicability for ordinary people.

As a result, these translations were initially accessible only to a small, educated scholarly class proficient in Classical Chinese. To expand their reach, later missionaries undertook efforts to translate the Bible into a more simplified and understandable version of Classical Chinese. This approach was intended to make the content more accessible to a broader audience while balancing sophistication with comprehensibility. The effectiveness and popularity of the language used by missionaries hinged on the language proficiency of the common people. Therefore, the evolution of a new form of written Chinese language was a necessary development for bridging this gap.

*Vernacular-Classical Chinese Version, "Easy Wen-li"*

The characters that Chinese used most commonly during the nineteenth century were those of Classical Chinese. However, the language as pronounced was extremely different from the spoken language of ordinary Chinese people. One would not be able to read or write systematically without any training in advance. With the background of an agricultural society, more than 90 percent of ordinary Chinese were illiterate. It was difficult to preach the gospel to people who had no formal language education. In order to do so, these missionaries had to learn to transform Classical Chinese into "Easy Wen-li" 浅文理, a simplified version of Classical Chinese. The translator-missionaries modified several Classical Chinese terms, changing them into something that was easier to understand. It is worth mentioning that the missionaries made this change long before the New Culture Movement, which promoted vernacular writing.

In the eleventh year of the Tongzhi Emperor's reign, the Meihua Bookstore in Beijing published *The Book of Common Prayer Translation in Mandarin*. This translation was a collaborative effort between Samuel I. J. Schereschewsky and Paul Burdon. At the time, church service books were commonly referred to as "Church Service" in Western languages, leading to the direct translation of the title as *The Church Prayer Book* (教会祷文). Schereschewsky, a Lithuanian Jew who joined the Episcopal Church in America, was later sent by the church as a missionary to China. Burdon was dispatched by the Church of England. He specialized in translating the New Testament, while Schereschewsky excelled in translating the Old Testament. The early version of their translation, known for its "Easy Wen-li" style, was a part of the *Public Prayer Book* (公禱書) translated by Schereschewsky, who also published a colloquial version of the New Testament in 1898. This style, a mix of classical and colloquial language, was designed to be accessible to the public.

---

<sup>1</sup> Some scholars translate 浅文理 as "simple Wen-li." The translation quoted here is adopted from John R. Hykes in his book *Translations of the Scriptures into the Languages of China and Her Dependencies*, published in 1915. "Like High Wen-Li, [Wen-Li] is a written and not a spoken language. It is readily understood by the ordinary scholar in all parts of China" (Hykes, iv).

天主的恩典時時常在，兇橫的人你為何作惡自誇。

*The Church Prayer Book* (see Appendix)

For instance, the Chinese version of the *Book of Common Prayer* uses the term 天主 (Tianzhu) to denote God. Both Burdon and Schereschewsky believed that 天主, a more natural expression in the Chinese context, was more suitable than the Classical Chinese term 神, which refers to deities in general. In Chinese, 神 often appears with the character 祇 to indicate both the God of Heaven and the God of Earth exclusively. The text shows that Schereschewsky chose 天主 (Tiānzhǔ), literally meaning “Lord of Heaven,” as the Chinese translation for “Lord.” This choice was deemed more comprehensible to the general populace compared to the alternative translation 神 (Shén). This word choice reflects a thoughtful consideration for clarity and accessibility in translation, especially in contrast to the less accessible Classical Chinese versions.

#### *Mandarin Version*

Mandarin is the general spoken language of Chinese empire; it is the living language of China.

Samuel Isaac Joseph Schereschewsky

The early translations of the Christian Bible into Chinese were primarily focused on fidelity to the original text, resulting in versions in Classical Chinese and Easy Wen-li. These translations were intended for educated Chinese readers. However, as the number of Christian converts in Chinese churches grew, it became evident that most people, having limited educational background, found it challenging to understand these translations, exposing a need for more accessible works. This need led to the development of new translations in Mandarin, which primarily referred to the Nanjing dialect, later known as Baihuawen, or the vernacular.

The Taiping Rebellion, a movement with Christian influences, erupted in China in 1853. Western missionaries saw this as an opportunity to spread Christianity throughout China. As the Taiping forces established their capital in Nanjing, the first Protestant translation in Mandarin was the *The Nanking*



*Mandarin Colloquial Version*. This translation was a collaborative effort by Walter Henry Medhurst, of the London Missionary Society, and John Stronach. Instead of creating a new literary form of Mandarin, they introduced colloquial phrases, made phonetic changes, and used Chinese idioms to make the text more accessible and easier for the general populace to grasp the message of the Bible. This translation also served as a reference for the subsequent Beijing Mandarin Version.

<sup>1</sup> 起頭有道、這道和上帝同在、道就是上帝。

<sup>6</sup> 有上帝所差遣的人、名叫約翰。<sup>2</sup>

<sup>1</sup> In the beginning was the Dao,<sup>3</sup> and the Dao was with God, and the Dao was God. <sup>6</sup>

There was a man sent from God whose name was John.

Gospel of John in the New Testament (Nanking Version)

The translation of “Word” as 道 (Dào) in the Chinese version of the Bible is a notable example of how translators adapted Christian concepts to align with Chinese cultural and philosophical understandings. This approach aimed to bridge cultural and religious gaps and make the Christian message more accessible and resonant with Chinese audiences. The translation developed by the missionaries and adopted by the Chinese people generated a more contemporary version of the Gospel:

<sup>1</sup> 宇宙被造以前，道已經存在。道與上帝同在；道是上帝。

<sup>6</sup> 有一個人，名叫約翰，是上帝所差遣的使者。

The Nanking version used “起頭有道” (“In the beginning was the Word”), which is a more literal

---

<sup>2</sup> Texts are from the digital edition of the *Nanking Mandarin New Testament*, created by Simon Wong. Digital edition United Bible Societies, 2017. From <https://www.bible.com/zh-TW/bible/2034/ACT.1.%E5%8D%97%E4%BA%AC%E5%AE%98%E8%A9%B1%E6%96%B0%E7%B4%84>

<sup>3</sup> The origin text of the Gospel of John was: “In the beginning was the Word, and the Word was with God, and the Word was God.” The Tao was the literal translation of the Classical text in Chinese.

translation of the original Greek text. The contemporary version starts with “宇宙被造以前，道已經存在” (“Before the universe was created, the Word already existed”), which provides more contextual clarity and aligns with a modern understanding of the universe. “這道和上帝同在、道就是上帝” (Nanking version) and “道與上帝同在；道是上帝” (contemporary version) both convey the same meaning but with slight differences in phrasing. The Nanking version uses “這道” (this Dao), emphasizing the subject introduced earlier, while the contemporary version repeats ‘道’ (the Dao) for clarity. Basically, the contemporary version offers more clarity and context to modern readers, while the Nanking version stays closer to a literal translation of the original texts.

The different versions of the Bible in Chinese represent more than religious texts. They marked the transition from Classical to Vernacular Chinese in direct translations, and that change mirrored the broader linguistic evolution in China from a literary language to one that is more accessible to the common people.

#### B. MISSIONARY PHONETIC SYSTEMS AND AN EARLY VERSION OF PINYIN

Until the Ming and Qing period, there was no system like pinyin to do the phonetic work for Chinese languages. In ancient Chinese, there were two main methods for teaching and learning how to pronounce characters: the Zhiyin (Direct phonetic 直音) method and the Fanqie (Reverse-cutting 反切) method. The Zhiyin method involves using homophones to annotate the pronunciation of a character. For example, to indicate the pronunciation of the character “箱” (*xiāng*, meaning “box”), you might say “读如香” (*dú rú xiāng*, “pronounced like 香”). This method uses simpler, more familiar characters to represent the pronunciation of more complex ones. However, the limitation is that one needs to know the pronunciation of many characters in advance. For characters without homophones, or with few or rare homophones, the direct sound method makes it difficult for people to understand the pronunciation, or even to annotate it. During the Eastern Han dynasty, with the introduction of Buddhism and its Sanskrit scripts into China, Chinese scholars became inspired by the phonetic clarity of Sanskrit, where consonants and vowels are distinctly separated. This influenced the development of the Fanqie method. The Fanqie method involves using two characters to indicate the pronunciation of a third character. This is done by combining the initial consonant sound of the first character with the final vowel sound and tone of the second character. For instance, the pronunciation of “兵” (*bīng*,

meaning “soldier”) might be annotated as “卑英切” (*bei ying qiè*). Where “切” (*qiè*) does not have a direct meaning in this context; it’s just a marker indicating the use of Fǎnqiè method. You take the initial consonant ‘b’ from “卑” (*bei*) and the final and tone ‘ing1’ from “英” (*ying*), combining them to get the pronunciation “bing” for “兵.” The transition from Fanqie to the relatively formal pinyin system was much influenced by the Western missionaries.

t'iēn chù y kiam sēm t'ō gīn hīm ì hīm kiáo yú xī<sup>4</sup>

天主已降生，托人形以行教于世

— 西字奇蹟

Matteo Ricci, an influential missionary, significantly contributed to Chinese linguistics by introducing the Roman alphabet into the Chinese language. Before Ricci’s innovative work, Chinese script was solely character-based, lacking an alphabet system for phonetic representation. His introduction of the Roman alphabet represented a revolutionary step, providing a tool that was both accessible and easy to understand for representing Chinese sounds, especially advantageous for Europeans. This pioneering approach would later exert a profound influence on Chinese scholars.

In 1605, Ricci authored *The Miracle of Western Writing* (*Xiji qiji*, 西字奇蹟) in Beijing. This seminal publication, the first to employ the Roman alphabet for a Sinitic language, featured parallel lines of text: one line presented the Romanized phonetic pronunciation of Chinese characters, and the other contained the Classical Chinese of the Gospel of Matthew. The book was structured into three thematic sections, each accompanied by four images illustrating Western Catholic ideals. These themes, reflecting messages like “信而步海，疑而即沉” (“Believe and walk on the sea, doubt and then sink”), “二徒聞實即舍空虛” (“Two gentlemen heard the truth and discarded vain thoughts”), and “淫色穢氣自速天火” (“Being licentious and filthy only summons the fire from Heaven”), were expressed through inscriptions alongside paintings depicting Old Testament stories, such as, among others, the testament about Peter’s being saved through faith—and drowning due to doubt. With the help of the

---

<sup>4</sup> During the Ming dynasty, the vernacular language of China was based on the Nanking dialect.

Roman alphabet, people can easily pronounce the religious passages in Chinese. These elements highlighted the strong missionary manifest behind Ricci's work. This focus on religious dissemination and linguistic innovation helped the development of more systematic and comprehensive phonetic systems, which eventually evolved into more systematic and comprehensive phonetic schemes like Pinyin.

*The Romanizing of Chinese Languages by Missionaries:*

• 1. *Dialect Church Romanization*

Robert Morrison devised his own Mandarin phonetic scheme to transcribe Chinese characters during his mission work in Guangzhou, compiling and publishing *A Chinese-English Dictionary* (华英字典) in 1823. Meanwhile, many missionaries were congregating in various southeastern coastal areas of China, such as Fuzhou, Guangzhou, Shanghai, and Xiamen, each with its distinct dialects. To communicate with the local people and teach them more effectively, various Romanized scripts for these dialects emerged.

<b>NING.</b>	
<p><b>甯</b><sup>1</sup> 8325 R. 青徑 C. <i>ning</i> H. <i>len</i> F. <i>ning</i> W. <i>ngiang</i> N. <i>nying</i> P. <i>qing, ning</i>, <i>nêng</i>, M. <i>lin, lén</i> Y. } <i>ning</i> Sz. } K. <i>yōng</i> J. <i>dei, niō</i> A. <i>ning</i> Even and Sinking Lower.</p>	<p>A surname. Used for No. 8327.</p> <p>甯波 Ningpo. 甯幫 Ningpo men,—as a class. 甯紹台 the three Prefectures of 甯波 Ningpo, 紹興 Shao-hsing, and 台州 T'ai-chou, forming the circuit under the Tao-t'ai at Ningpo. 甯國府 the Prefecture of Ning-kuo in Anhui.</p>

Character 甯 from *A Chinese-English Dictionary*

The left part of each page of the dictionary marked the sounds of various local Chinese dialects and non-native Chinese characters with uppercase English letters, such as C (Cantonese), H (Hakka), F

(Fuzhou), W (Wenzhou), N (Ningbo), P (Beijing), M (Central China, i.e., Hankou), Y (Yangzhou), Sz (Sichuan), K (Korea), J (Japan), A (Annam/Vietnam), etc. This highlights the significant impact missionaries had on the dialects and script reform movement in China. Their work made it evident that Chinese characters could be adapted to a Latinized phonetic system. The publications by these missionaries demonstrated the feasibility of using the Latin alphabet to transcribe Chinese, offering a viable approach towards the Latinization of the Chinese language. This was also the earliest dictionary to systematically integrate the pronunciation of words in various regional dialects in China and to successfully divide regional dialects in China into different dialects based on their pronunciation.

- 2. *Wade-Giles and the Postal Spelling System*

*In vicinam Peking penetravit. Sed cum a regia urbe septem tantum aut circiter distaret leucis, ultra, ne intercluderetur, progredi timuit, quod undique Sinarum ingentes confluerent exercitus.*

—*De bello Tartarico historia* by Martino Tridentinus Martini

Martino Martini’s work in the seventeenth century, including his use of the term “Peking” for Beijing, indeed represents an early and significant European engagement with the Chinese language and culture. His efforts, along with those of other Jesuit missionaries, in learning and transcribing Chinese using the Latin alphabet, laid important groundwork for the development of later Romanization systems. The significance of Martini’s work lies in bridging Chinese and Western linguistic worlds. Their systematic approach to transcribing Chinese sounds using the Latin alphabet made the Chinese language more accessible to Europeans.

In the nineteenth century, Thomas Francis Wade, a British diplomat and sinologist, furthered this field through his work “Y ŭ yán Ziěrjí” (语言自邇集), or “A Progressive Course Designed to Assist the Student of Colloquial Chinese.” Inspired by classical Chinese texts, including the Confucian “Doctrine of the Mean,” Wade’s book functioned as both a textbook for teaching Mandarin Chinese, particularly the Beijing dialect, to Westerners, and as a crucial resource for documenting this dialect’s specifics. The Romanization system he developed, which later became known as Wade-Giles, included

twenty-seven consonants and forty vowels. As shown in the table, the four tones here are also the embryonic forms of the four tones in the modern Chinese Pinyin system.

	1 upper even	2 lower even	3 upper	4 going
mi	眯 mi <sup>1</sup>	迷 mi <sup>2</sup>	米 mi <sup>3</sup>	密 mi <sup>4</sup>

Reference: T. F. Wade (2002): 语言自邇集: 19 世纪中期的北京话. 北京大学出版社: 33

During the mid- to late nineteenth century, a debate emerged among sinologists about the standard dialect of Mandarin Chinese. While some scholars favored the Nanjing dialect, Wade and others advocated for the Beijing dialect as the standard. Building on the work of his predecessors, Wade's Romanization system was a significant step in the evolution of Chinese language studies, helping to formalize the teaching and understanding of Mandarin Chinese in the Western world.

#### *National Language Romanization and the Phonetic Alphabet*

During the middle and late periods of the Guangxu era in the Qing dynasty, China's national strength gradually weakened. Due to the complexity of Chinese characters and the difficulty in teaching them, education was not widespread. As a result, there were proposals for the reform of Chinese characters. The first proposal was to abandon Chinese characters altogether and replace them with a new international language. The second was to advocate for the use of Romanized spelling by Western missionaries to replace or supplement Chinese characters. The third was to create a phonetic script similar to the Japanese kana, improving the traditional 'fanqie' (reverse cutting) method for aiding pronunciation. It was not until May of the twenty-first year of the Republic of China that the National Language Romanization and Phonetic Alphabet were officially published in the *Gwoin Charngyong Tzyhhuey* (国音常用字汇). This period marked the beginning of the comparison between the National Language Romanization and the Phonetic Alphabet. There were several notable phonetic schemes representative of this movement:

- 1. *Mandarin Roman Word Scheme*

In 1913, the Ministry of Education of the Beiyang Government in China convened a meeting to unify pronunciation and adopted Zhuyin alphabet (also known as Bopomofo) as a tool for phonetic notation

of Chinese characters. However, some intellectuals still did not give up the demand for reform. They believed that the Zhuyin alphabet, derived from Chinese characters and imbued with national character, was an auxiliary tool for writing. According to the understanding at the time, the alphabetic form of Chinese characters was contrary to the mainstream development trend of world scripts. They also believed that Chinese characters were extremely backward and should be abolished. The argument was, "Rather than creating new characters that are unprecedented in the world, it is better to adopt the alphabets used worldwide." Thus, the elite class initiated the "National Language Romanization Movement," advocating for the abandonment of Chinese characters and attempting to represent all Chinese pronunciations with 24 Roman letters. In 1932, the Ministry of Education announced the "Commonly Used Characters in National Pronunciation," officially establishing the Beijing accent as the national standard pronunciation. This was mainly because the national pronunciation had fewer characters, and the phonemes of the Beijing dialect were simple, aligning with the trend of script simplification at the time.

A characteristic of this scheme was to use the current twenty-four Latin characters, excluding "x" and "v," without adding new characters or symbols. The usage of these characters resembled English rules. For example, in the National Language Romanization, the sound "ai" in Chinese Pinyin could be written as 'ai', 'air', 'ae', or 'ay' (representing the four tones respectively). Here, 'r', 'e', 'y' were silent, serving only to indicate tonal changes. This approach differed from the Western romanization schemes introduced by figures like Matteo Ricci, Joseph Edkins, and Thomas F. Wade, where symbols were added to change the tone of the Pinyin. The Chinese scheme aimed to simplify by not adding symbols for tonal distinctions.

- 2. *Cutting Words Scheme* (切音字)

LU Zhuang Zhang (卢翥章) was a significant figure in the Chinese language and script reform movement. He spent his youth studying English in Singapore and later returned to his hometown, Xiamen, to work as an English teacher, where he also taught foreigners the Xiamen dialect. At that time, Church Romanization was very popular in Xiamen, it originally created and implemented by the Christian Presbyterian Church in Xiamen, Fujian Province in the nineteenth century. His exposure to Church Romanization piqued his interest, inspiring him to conceive the idea of creating a phonetic system for Chinese. A decade later, he authored "一目了然初阶" (*Yi Mu Liao Ran Chu Jie*), a cutting-

edge work that used Latin letters and their variants as phonetic symbols for reading Chinese. LU Zhuang Zhang did not advocate for the abolition of Chinese characters: instead, he promoted the coexistence of his phonetic system and Chinese characters, believing that his system could enable self-learning of Chinese characters without formal instruction. He introduced a phonetic system design consisting of 15 initial consonants, 33 vowels, 14 nasalized vowels, and 8 tones. The writing method for the initial consonants and vowels uses a variant of the Latin alphabet evolved from the strokes of ‘l’, ‘c’, and ‘ㄅ’. For example, the reverse cutting for the character “时” is written as ‘ic’ at the fifth tone (‘^’ on top of the character c) (see Appendix). This approach, combining vowels and initial consonants, had a significant influence on the later development of Pinyin, the standard romanization system for Chinese.

### III. THE IMPACT OF MISSIONARY ACTIVITIES ON THE CHINESE LANGUAGE

The activities of missionaries have been reflected in the use of Chinese in various written media and learning methods. According to Álvaro de Semedo, the Portuguese missionary who mentioned in his book *The History of That Great and Renowned Monarchy of China* that the Chinese language is unlike any other in the world, whether in terms of the sound of speech, the pronunciation of characters, or the combination of concepts (Semedo, 51). He also mentioned that China has a vast geographical span and a long history, with significant differences in dialects across different regions. This lack of uniformity in language posed substantial difficulties for missionary work, as well as challenges in learning the Chinese language. Therefore, according to the needs of missionary activities, missionaries needed to devise cultural activities, in other words, to learn Chinese culture while learning the Chinese language. Only by integrating their translations into larger circles of the Chinese populace and using Chinese culture to narrow the distance between Eastern and Western people, could there be an opportunity to influence the Chinese to further accept Christianity. This also provided missionaries with an excellent opportunity to understand the current state of the Chinese language, such as the lack of teaching methods, and the vast differences between written language and everyday speech. In their evangelistic progress, they inevitably had a series of impacts on the language progression of China.



A. THE APPLICATION OF VERNACULAR CHINESE LANGUAGE ON MISSIONARY WORK

In the early nineteenth century, Western missionaries established Chinese-language newspapers in China, marking the beginning of China's modern press industry. Newspapers, as the most important medium of the time, played a significant role in spreading culture and Western thought. This was an important aspect of the cultural exchange undertaken by Western missionaries in China. The advent of Chinese-language newspapers triggered a series of language and script reforms, with the written language in newspapers gradually shifting from Classical Chinese to Vernacular Chinese. Although the initial intention of the missionaries in the 1830s was to spread Christian doctrine and embed it more deeply in the hearts of the Chinese people, over time, the focus of the content shifted from preaching to culture, science, and technology. During the Western missionaries' exploration of Chinese cultural issues, one can see from the newspapers that their understanding of Chinese culture deepened, thereby narrowing the distance between East and West and promoting linguistic changes in modern Chinese culture.

In the early missionary publications, in addition to propagating Christian doctrine, the content also included ethics and morals, Western history, geography, customs, etc. This was novel for the Chinese people at the time. The newspapers themselves also became a platform for a chemical reaction between Chinese and Western languages and scripts. To better acquaint the public with the Western world, the articles published often used more colloquial expressions, a mix of Classical Chinese, and a written language that included foreign vocabulary and Western punctuation. The Chinese-language newspapers of Western missionaries played a promoting role in the later rise of the Vernacular Chinese movement.

*Written Chinese and Vernacular in Magazines*

The first Chinese-language newspaper in modern China was established in 1815 by William Milne, a missionary from the London Missionary Society. Two years prior, Milne was commissioned by the Governor of Portuguese Macau to conduct missionary work in the Guangzhou region. He also assisted Robert Morrison in translating the Bible. Although the literary style of their work leaned towards Classical Chinese, both Milne and Morrison were considered pioneers in advocating and promoting Vernacular Chinese. Milne's translations were numerous, but his most significant contribution during

his time in China was the establishment of a Chinese-language newspaper. His “察世俗每月统计传”—*Chinese Monthly Magazine*—was the world’s first modern Chinese-language periodical primarily targeting the Chinese audience. Due to the Qing government’s strong restrictions on foreign missionaries, its main readership was the Chinese community in the British colonial area of Malacca. The magazine featured concise, story-driven content that was easily understandable by the public.

富贵之人不多，贫穷与作工者多，而得闲少。虽于道但读不得多书。一次不过读数条。因此察世俗书之每篇比不可长，也必不可难明白。

察世俗每月统计传

A characteristic extract from his work, which reflects the features of the Cantonese dialect, includes phrases like “作工” (to work) and “得闲” (to have free time)—terms commonly used in Cantonese but less in Beijing Mandarin. This shows that Milne absorbed many linguistic and cultural characteristics of the Cantonese-speaking region of Malacca and hints at the beginnings of Vernacular Chinese. Milne’s choice of words was popular and colloquial, and his tone was closer to spoken language.

Another notable aspect of Milne’s work was his use of Western punctuation in Chinese writing. Traditionally, Classical Chinese did not include punctuation, making it challenging to interpret sentences without context. Milne’s introduction of punctuation significantly reduced the time needed for readers to understand the text and positively influenced the development of modern written Chinese.

Not only was there a revolution in style, but the content also had a significant influence on Chinese society. Milne’s writings included statements like, “若神一少顷取去全能之手，不承当宇宙，则日必不复发光，天必不复下雨，川必不复流下。” (If God were to remove his almighty hand for a moment and cease to uphold the universe, then the sun would no longer shine, the sky would no longer bring rain, and the rivers would no longer flow.) Such language, rich in religious color, was a complete cultural shock for Chinese society at the time, which was primarily influenced by Confucian concepts of destiny and Buddhist beliefs in reincarnation. Only by bringing the cultural nuances of different parts of the world closer and opening the hearts of the Chinese people with their own script, could Western religion deeply resonate and gain acceptance among the Chinese populace. Milne’s

impactful publications brought a significant shock to China at the time, turning its gaze towards Western society and laying a good foundation for the subsequent blending of Chinese and Western cultures.

Milne's desire in promoting vernacular Chinese had been manifested in his "察世俗每月统计传." The point was not to just do evangelist work but also to promote interest in the religion among ordinary Chinese people. Giving consideration to the fact that most audiences were people who did not have a high educational attainment, were busy, and did not know very many characters, Milne made the passages in this book short, precise, and easy to understand. The form of the writing was relatively close to vernacular Chinese instead of classical.

从前有两个好朋友，一名张，一名远，他们两个人同行路间，相论古今。远曰：我已听人说，尊驾曾经受了耶稣之道理，而信从之，我看世人论说此事，多有不同，且我自己不甚明白，今有两端，欲求尊驾解之。<sup>5</sup>

From this quotation, it is evident that the writing style is akin to the vernacular form of the Chinese language. The intention behind creating this magazine was to enable individuals with a fundamental understanding of Chinese to easily read and comprehend the texts. Consequently, Christian teachings were gradually and subtly introduced into the mindset of the Chinese populace.

## B. THE CLASSIFICATION AND LEARNING METHOD OF VERNACULAR CHINESE LANGUAGES

In Matteo Ricci's *De christiana Expeditione apud Sinas suscepta a Societate Jesu*, mention was made of an Italian Jesuit, Michele Ruggieri, who stayed in Macau for ten months. This period of stay provided him with an excellent opportunity to thoroughly understand China. Ruggieri believed that a nation intelligent, accomplished, and dedicated to the study of arts could be persuaded to allow foreigners of

---

<sup>5</sup> Once upon a time there were two good friends, one named Zhang and the other Yuan. As they traveled together, they discussed the past and present. Yuan said, "I have heard that you, my friend, have embraced the teachings of Jesus and have become a follower. I observe that people have various opinions on this matter, and I myself do not quite understand it. Today, I have two questions and would like to ask you to explain them."

similar virtues and knowledge to reside among them, especially those proficient in the Chinese language and script. He thought that this nation held promise and that, with the right approach tailored to the Chinese people, they could eventually be receptive to Christianity (Ricci, 157). This shows that learning Chinese culture and the Chinese language greatly aids missionary activities, and learning Chinese becomes the primary task for missionaries, also influencing the shaping of the Chinese language by missionaries in the future.

To better integrate themselves into local communities, missionaries endeavored to systematically learn and understand the local spoken language. After the Opium War, missionaries were able to reside long-term in concessions such as Guangdong and Shanghai. It was from this time that missionary sinologists began to gain a more detailed observation of the differences among spoken languages popular in different places. This conclusion is clearly reflected in Joseph Edkins' "A Grammar of Colloquial Chinese Commonly Called the Mandarin Dialect." First, he divided the Mandarin dialects into three parts: northern, western, and central, based on their pronunciation. He identified the central representative language as the Nanjing dialect, the northern as the Beijing dialect, and the southern as the Sichuan dialect (Edkins, 3). This classification simplified subsequent sinological research. For the pronunciation of different words or characters, there are significant differences between regions, and Joseph Edkin clearly listed them, allowing people to intuitively differentiate nuances they had not in the past. Based on Edkins' research, learners could gain a clear understanding of the tones. The Nan-King dialect usually has fewer tones and fewer pronunciations of the same words as compared to other dialects.

Table of Tones in Several Dialects of Chinese<sup>6</sup>

	Upper even	Upper rising	Upper going	Upper entering	Lower even	Lower rising	Lower going	Lower entering
Mandarin (Represented by Nan-King <sup>7</sup> )	i, e	i, s, r	q, f	sh	i, q, r	/	/	/
Kiang-Nan (Represented by Shang-hai)	u, q, f	u, e	u, s, r	u, sh	i, q, r	i, q, f	i, s, r	i, sh
Fuh-Kien (Represented by Amoy and Chang Cheu)	u, e	u, q, f	i, f	u, sh, f	i, q, r	u, q, f	i, e	u, sh, f

For English speakers, Edkins' book offers an intuitive approach to quickly grasping the pronunciation of spoken Chinese. His teaching method employs a clever technique: using elements of English words to mimic the sounds of a monosyllabic Chinese character. For example, to pronounce the Chinese character '無' (*wu*), one could use the initial consonant sound 'w' from the English word "swerve" and combine it with the vowel sound 'u' from "ruin." This way, the 'w' sound from "swerve" replicates the consonant of the Chinese character, while the 'u' sound from "ruin" matches its vowel. By blending these sounds, learners can easily articulate the Chinese character. This method enables English-speaking individuals to practice and master the correct pronunciation of Chinese characters without the need for phonetic learning or hearing the sound physically. Edkins' approaches anglicized each Chinese word, facilitating easier acquisition for English speakers.

<sup>6</sup> u: upper. l: lower. r: rising. f: falling. q: quick. s: slow. e: even. c: circumflex. sh: short

<sup>7</sup> Nán-king is here placed in the category of 北音, or Northern Mandarin division.

Table of vowel symbols and English values in expressing the sounds of Mandarin pronunciation

i	i in <i>marine</i>	西 si <i>west</i>
e	e in <i>there</i>	天 t'ien <i>heaven</i>
a	a in <i>father</i>	怕 p'a' <i>fear</i>
u	u in <i>prune</i>	都 tu <sup>8</sup> <i>all</i>
o	o in <i>lone</i>	可 'k'o <i>can</i>

#### IV. CONCLUSION

This study demonstrates the evolution of the Chinese language over hundreds of years, as it was significantly influenced by Western missionaries. The activities of these missionaries in China and Asia not only changed the course of Chinese cultural history but also made substantial contributions to the Chinese language. The missionaries categorized various Chinese dialects, promoted the systematization of the Chinese phonetic system, and participated in the transition from Classical Chinese to Vernacular Chinese, providing many reference schemes.

From the Ming and Qing dynasties down to the present, the Chinese language, with its qualities mysterious to the Western world, underwent significant changes. The variation in dialects, the methods of oral pronunciation, and the vast differences between written and spoken language were distinctly different from Western languages. The efforts of Western missionaries to sort out Chinese dialects in dictionaries and other published works uniquely contributed to shaping the Chinese language system and were a major driving force in its evolution.

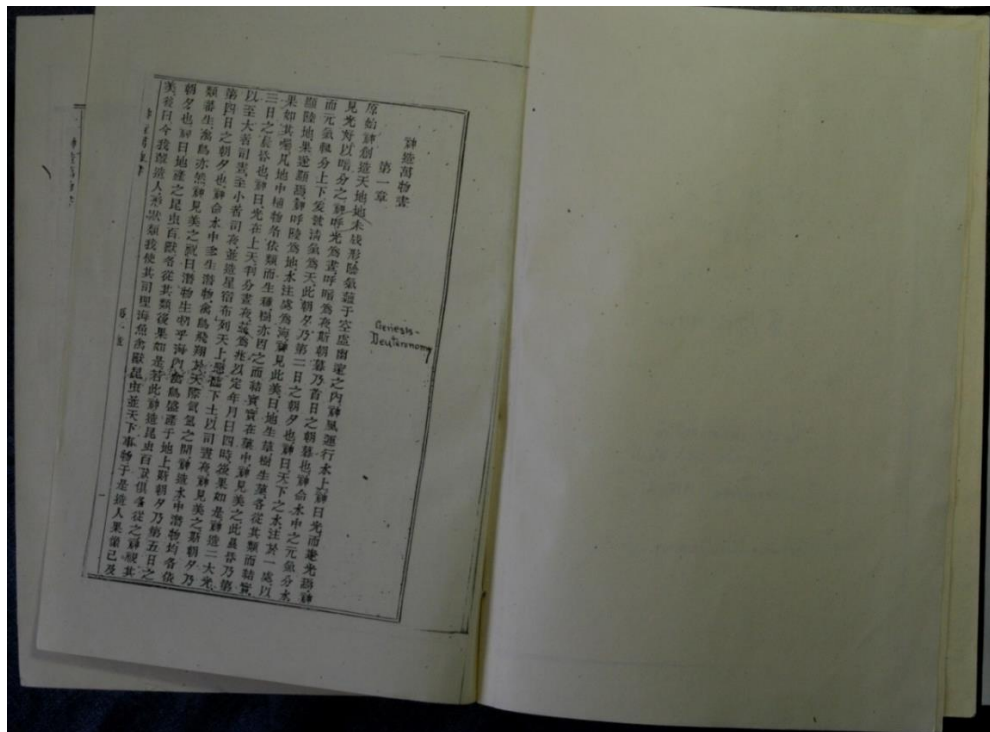
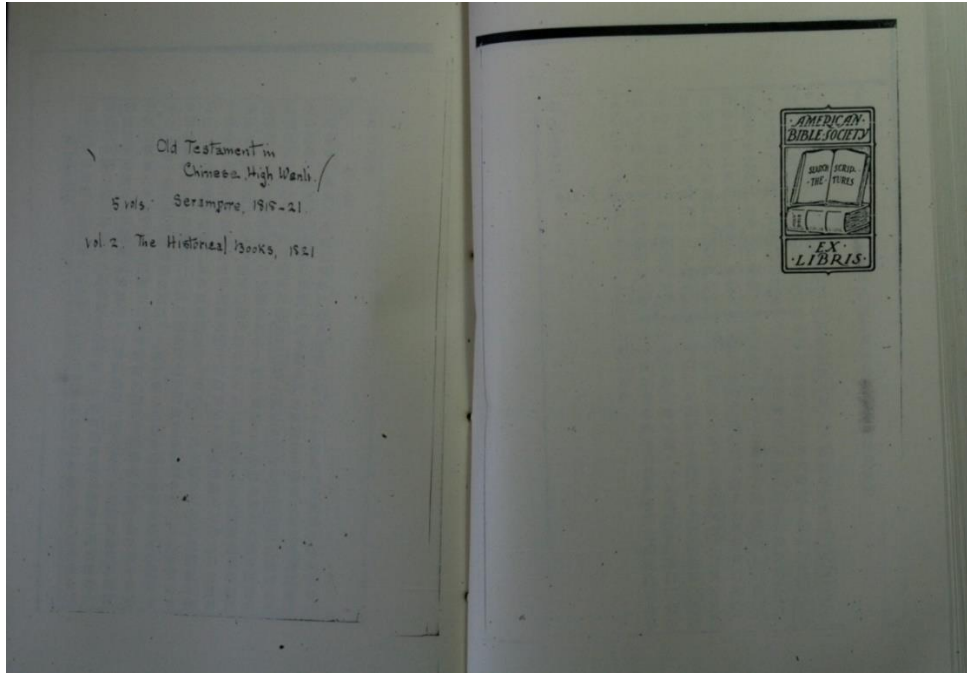
The introduction and development of phonetic systems, the adaptation of language for easier understanding, and the incorporation of Western linguistic concepts were pivotal in bridging the gap between Eastern and Western linguistic traditions. The missionaries' work in translating the Bible into various Chinese forms, including Classical Chinese, Easy Wen-li, and Mandarin, made religious texts more accessible to a broader Chinese audience and catalyzed the shift from Classical to Vernacular Chinese.

---

<sup>8</sup> If "all" indicates the Chinese character 都, the correct pronunciation should be 't'ou instead of tu.

In the modern era, the movement for language and script reform in China involved not only cultural and historical exchanges between the East and West but also political and economic trade relations. Western missionaries penetrated the largest segment of the Chinese population, the common people. They founded magazines, wrote books, and conducted a series of cultural missionary activities, popularizing Vernacular Chinese and opening avenues for cultural and linguistic synthesis. This helped the Chinese people gain a deeper understanding of their own language and develop a Western worldview, thereby narrowing the gap between East and West. It also facilitated more systematic learning of Chinese writing and pronunciation. During this period, the vocabulary used by Western missionaries in translations gradually became ingrained in the hearts of the Chinese people, becoming a part of the development of the Chinese language. Furthermore, the content of their preaching often challenged traditional Confucian culture in the East, bringing new ideas and concepts to the Chinese people through both language interaction and the spread of texts.

APPENDIX



Old Testament by Joshua Marshman



說解讀者環		鼻音	母字邊左	音五右
皆讀履腔	腦箱見貨	三	ō l i l i ō	柳邊求去地
	交周伊蘇	二	o l i l o	波他貞入時
	鴉昭威野	二	a x l u	英文語出喜
三貓開驚	三	ā x l ō		
來最齋快	二	l p e t		
我無着汗	二	b o t u		
看脈不問	三	b e l w o		
真仙春嘯元	三	l r i c h u		
鶯商公雙人	三	k m d o s		
風心甘沾參	三			

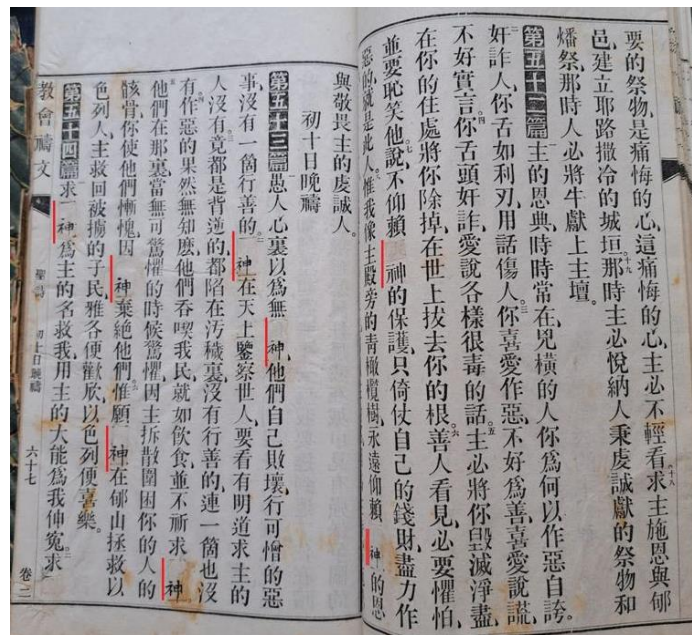
鼻音 一、二、三、四、五、六、七、八、九、十、十一、十二、十三、十四、十五、十六、十七、十八、十九、二十、二十一、二十二、二十三、二十四、二十五、二十六、二十七、二十八、二十九、三十、三十一、三十二、三十三、三十四、三十五、三十六、三十七、三十八、三十九、四十、四十一、四十二、四十三、四十四、四十五、四十六、四十七、四十八、四十九、五十、五十一、五十二、五十三、五十四、五十五、五十六、五十七、五十八、五十九、六十、六十一、六十二、六十三、六十四、六十五、六十六、六十七、六十八、六十九、七十、七十一、七十二、七十三、七十四、七十五、七十六、七十七、七十八、七十九、八十、八十一、八十二、八十三、八十四、八十五、八十六、八十七、八十八、八十九、九十、九十一、九十二、九十三、九十四、九十五、九十六、九十七、九十八、九十九、一百

由字母切起  
左邊是字母  
右邊是字母  
單字是字母

切音成字	走一個蠟	左右
	掠一個蟻	左右
	走一個蠶	左右
	掠一個龜	左右

<p>子 文 止 所 沼 委</p>	<p>掠鯉魚</p>	<p>ám ín-úr</p>
	<p>晉朝有一個王祥</p>	<p>ń-xi ũ ń-ê çá-tî</p>
	<p>寒天之時</p>	<p>đo-úr ê íc</p>
	<p>水面皆堅冰</p>	<p>bu-ńw ón do re</p>
	<p>娘母愛食鮮魚</p>	<p>ōn-úr ia áu-ú-úr</p>
	<p>伊就褫剝體</p>	<p>l ũ ót-úe-ét</p>
	<p>卧在冰裡</p>	<p>ól ũ re ń</p>
	<p>要俾冰燒着温</p>	<p>év ór re ór-ń-rc</p>
<p>éó</p>	<p>鎔鎔去</p>	<p>lĭa-lĭa ń</p>
<p>cm</p>	<p>就可落水掠魚</p>	<p>l ũ ń ón-bu án-úr</p>
<p>假鹿</p>	<p>纚纚卧的之時</p>	<p>ú-ú ót-éi ê íc</p>
	<p>有兩尾鯉魚</p>	<p>ũ wń év ń-úr</p>
	<p>忽然跳起來</p>	<p>ńt-đo x-ń-ń-ń</p>
	<p>伊就掠轉來</p>	<p>l ũ án ót-ń</p>
	<p>俾娘母食</p>	<p>đr ōn-úr áu</p>
	<p>假鹿</p>	<p>éó ón</p>
<p>li</p>	<p>周朝有一個剡子</p>	<p>l-xi ũ ń-ê đt-úw</p>
<p>đ</p>	<p>父母年老</p>	<p>ēe-úr ń-ōn</p>
<p>ń</p>	<p>兩個目矇皆青盲</p>	<p>wń-ê ń-ń-ń ń-ń-ń</p>

《一目了然初阶》 by LU Zhuang Zhang (卢慧章)

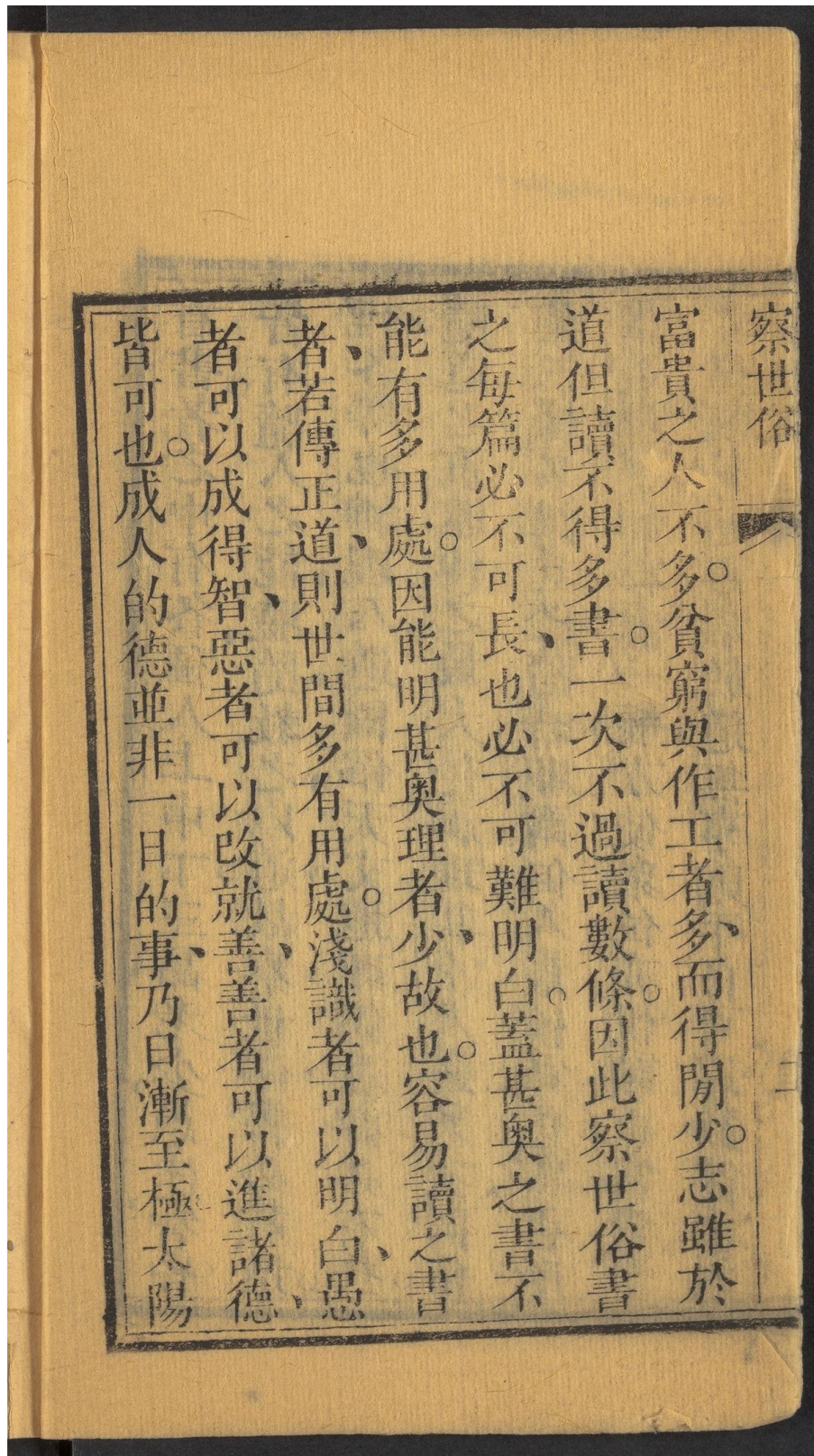


《公禱書》 *The Book of Common Prayer* by Samuel I.J. Schereschewsky and Paul Burdon<sup>9</sup>

<sup>9</sup> Schereschewsky's translation of "God" faced strong opposition from the followers of the Anglican Church in southern China. This copy of "The Book of Common Prayer" is in southern China, where all the '天主' (*Tianzhu*) in the text were changed to '神' (*Shen*).

墨 色 留 奇	多 落 一 日 在 船 恍 惚 見 天 主 立 海	tō lō yē gǐ cāi chūen hōim hōē kiēn tiēn chū siē hai	誨 十 二 聖 徒 其 元 徒 名 曰 伯 多 落 伯	bōci xē sh xim tū Kī iuēn tū mīm yuē jū tō lō jū	天 主 已 降 生 託 人 形 以 行 教 于 世 先	tiēn chū ǝ Kiam sem tō gin him i him Kiao yū xi sien	信 而 步 海 疑 而 即 沈	sin sh pū hai nhi sh ciē chin
------------------	---	--	--	---	--	---	--------------------------------------	--

《西字奇蹟》 *The Miracle of Western Words (Xizi qiji)* by Matteo Ricci



《察世俗每月统计传》 Chinese Monthly Magazine by William Milne

## REFERENCES

- De Ridder, K. (2001). *Authentic Chinese Christianity: Preludes to Its Development (Nineteenth and Twentieth Centuries)*. Leuven University Press.
- Dehergne, J., and S. Geng. (1995). *Zai Hua Ye su hui shi lie zhuan ji shu mu bu bian* [在华耶稣会士列传及书目补编] (1st ed.). Beijing: Zhonghua shuju [中华书局].
- Edkins, J. (1857). *A Grammar of Colloquial Chinese Commonly Called the Mandarin Dialect*. Retrieved from [https://www.google.com/books/edition/A\\_Grammar\\_of\\_the\\_Chinese\\_Colloquial\\_Lang/oeOiNryRt4C?hl=en&gbpv=1&pg=PR5&printsec=frontcover](https://www.google.com/books/edition/A_Grammar_of_the_Chinese_Colloquial_Lang/oeOiNryRt4C?hl=en&gbpv=1&pg=PR5&printsec=frontcover).
- Edkins, J. (1868). *A Grammar of Colloquial Chinese as Exhibited in the Shanghai Dialect*. Retrieved from [https://www.gutenberg.org/files/62128/62128-h/62128-h.htm#Article007\\_Footnote\\_1](https://www.gutenberg.org/files/62128/62128-h/62128-h.htm#Article007_Footnote_1)
- Guo, F. (2021). *Jin Xian Dai Zhong Wen Gong Dao Shu De Ban Ben Yu Shang Xi* [近现代中文公祷书的版本与赏析]. *Gospel China*. Retrieved from <https://gospelchina.cn/article-detail/94.html>.
- Huang, C.-R., Z. Jing-Schmidt, and B. Meisterernst, eds. (2019). *The Routledge Handbook of Chinese Applied Linguistics*. Routledge.
- Hykes, J. R. (1916). *Translations of the Scriptures into the Languages of China and Her Dependencies: Tabulated to December 31, 1915*. New York: American Bible Society. Retrieved from <https://babel.hathitrust.org/cgi/pt?id=nncl.cu07029152&seq=8>
- Liu, C.-C. (2021). "Discipled by the West? — The Influence of the Theology of Protestant Missionaries in China on Chinese Christianity through the Translation of the Chinese Union Version of the Bible." *Religions*, 12(4): 250. <https://doi.org/10.3390/rel12040250>
- Martini, M. (1655). *De bello Tartarico historia*. NP.
- Medhurst, W. H., and T. Wang. (1860–1861). *New Testament: Nanking Mandarin Colloquial Version* [南京官话译本 新约全书]. London Missionary Society Press [墨海书馆].
- Milne, W. (1815). *Cha shi su mei yue tong ji zhuan* [察世俗每月统计传]. Madaxia.
- Morrison, R. (1823). *A Chinese–English Dictionary*. Shanghai: Kelly & Walsh Limited.
- Morrison, R. (1823). *Holy Bible: Morrison-Milne Chinese Bible*. Retrieved from <https://www.bible.com/bible/2283/GEN.1.Morrison-Milne>.

- Schereschewsky, S. I. J., and P. Burdon, translators. (1877). *The Book of Common Prayer: Translation in Mandarin*. Beijing: Meihua Bookstore.
- Semedo, F. A. (1655). *The History of that Great and Renowned Monarchy of China*. London: John Crook.
- Shuai, S. (2021). "Between Evangelism and Politics: A Study on S. Wells Williams' Diplomatic Translation Activities in China." Chinese University of Hong Kong PhD dissertation. ProQuest Dissertations and Theses Global. (2653340293)
- Standaert, N. (2017). "The Chinese Mission Without Jesuits: The Suppression and Restoration of the Society of Jesus in China." *Ching Feng* 16(1): 79–96. Retrieved from <https://www.proquest.com/scholarly-journals/chinese-mission-without-jesuits-suppression/docview/2016961252/se-2>.
- Takekoshi, M. (2018). "Cantonese as Learned by Missionaries in the 19th Century: Language Variation in Easy Lessons in Chinese." *Current Research in Chinese Linguistics*, 97(1).
- Wade, T. F. (2002). *Yuyanzierji: Beijing Dialect in the Mid-19th Century* [语言自述集: 19 世纪中期的北京话]. 北京大学出版社.

# Are Emojis Substitutable for Words?

## A Critical Review on the Quasi-Linguistic Properties of Emoji

Yifei Yang

University of Pennsylvania

Ever since their debut on smart devices in 2011, emojis—literally “pictorial characters” in the Japanese original—have been both phenomenally popular and intellectually stimulating, particularly in the discipline of sociolinguistics. The colorful, expressive glyphs cover a wide range of symbols, including but not limited to facial expressions, gestures, foods, daily necessities, flags and public information signs. Designed as “a visual representation of a feeling, idea, entity, status or event” that is available in digital messages, emoji usage has skyrocketed along with the rapid expansion of mobile communication.<sup>1</sup> According to the statistics from emojiopedia.org, the total number of emojis in the Unicode Standard reached 3,664 by September 2022.<sup>2</sup> Alongside the sharp increase in number, the growing popularity of emojis invites scholarly discussions in lexicology and pragmatics, as well as in communication studies and among critics in popular culture. Through an overview of the sociolinguistic discussions on emojis in the current scholarship, this article proposes a critical reflection on the “quasi-linguistic” features of this groundbreaking digital writing system, with special focus on its meaning-making mechanisms, as well as the opportunities and challenges in becoming a globally shared “universal language.”

### MONOPOLIZED IN LEXICON, LIMITED IN PRAGMATICS

In the very beginning of his thought-provoking introductory book, *The Emoji Code: The Linguistics*

---

<sup>1</sup> Vyvyan Evans, *The Emoji Code: The Linguistics behind Smiley Faces and Scaredy Cats* (New York: Picador, 2017), 13.

<sup>2</sup> The Unicode Consortium, “Emoji Statistics,” emojiopedia.org, accessed December 17, 2023, <https://emojiopedia.org/stats>.



*Behind Smiley Faces and Scaredy Cats*, Professor Vyvyan Evans refers to emojis as "colorful pictograms."<sup>3</sup> On seeing "pictograms," readers familiar with the history of writing systems might easily be reminded of those ancient scripts in either Egypt or Mesopotamia. Similar to the earliest writings in human history, emojis also use pictorial symbols to convey meanings. The popular concerns then arise: Is emoji considered to be a new language, with the presence of vocabulary and grammar? If not, how is emoji processed differently from natural languages?

It is noteworthy that Evans himself remains skeptical about the tempting equation between "emoji" and "language." The first point he notices is the monopolizing nature of emoji through technological autocracy: "[e]moji is unlike a natural language in that it is controlled by powerful multinationals whose representatives sit on the various Unicode committees. Even more powerful are the software developers such as Apple, Google and Microsoft, which interpret the Unicode code points, and design what a particular emoji looks like on their platforms."<sup>4</sup> While a natural language gains uncountable neologisms and variations through daily conversations, the use of emoji is strictly limited by its availability on digital devices. Each single new "character" requires a prolonged timeline for the committee to approve or reject. Even if individual users and organizations can create personalized emojis through photo editing tools (for example, Finland's Ministry of Foreign Affairs' attempt to launch "national emojis" representing Finnish culture),<sup>5</sup> the modified products are merely "pictures" or "stickers," instead of "characters" that can be directly inserted in textual messages. In short, the mass circulation of emoji depends on an exclusive system to transform the "pictographs" into real, portable "texts" shared among the digital community. The strictly monopolized feature of emoji distinguishes it from decentralized natural languages.

In the third chapter, "What's in a Word?," Evans examines the functions of emojis with a closer analysis of their similarities and differences in relation to the conventional "words." He first acknowledges the word-like quality of emojis, highlighting the similar roles as "single meaningful unit[s] deployed in digital, text-based communication." In those cases, the emojis are directly inserted into text

---

<sup>3</sup> Evans, 7.

<sup>4</sup> Evans, 46.

<sup>5</sup> Evans, 21.

messages (such as *Have you fed the 🐱*?) as substitutions for common English words (e.g., “cat”). Such synonymous substitution processes are regarded as analogous to code-switching, in which words and meanings in different languages replace parts of the conversation.<sup>6</sup> However, the “substitutability” between emojis and words does not necessarily render the former more (or less) linguistic features. Although both emojis and foreign words/phrases can be inserted into a sentence and used to convey certain information, it is nearly impossible to rely purely on emojis if the speaker attempts to express abstract notions like “*chaos, betrayal, social norm, duty, pithy and empathy*”; or more complicated logical frameworks, such as “*the beauty of the lines of a sports car.*”<sup>7</sup> Take the last phrase about a car as an example: it is constituted by articles (*the* and *a*), prepositions (*of*), concrete nouns (*lines* and *car*), abstract noun (*beauty*) and even an adjectival noun (*sports* as in *sports car*). If we try to “translate” this phrase into emojis, we are stymied because there is no established classification for lexical categories in the realm of emoji, and this greatly limits its capacity to address complex logical frameworks. Evans also suggests that the so-called “grammar of emoji” is still far from developed. In daily conversations, the ambiguity of emojis means that “interpreting a relationship between emojis that appear next to each other hardly counts as some kind of ‘Emoji grammar.’”<sup>8</sup> Although it is technically possible to combine multiple emojis, if readily available in Unicode, to create a more nuanced unit of meaning (see Fig. 1), this “emerging emoji grammar” lying behind the combination process is still regarded as an immature form of grammar, since it is often automatically done by smart devices rather than users themselves.<sup>9</sup>

---

6 Evans, 57.

7 Evans, 61–62.

8 Evans, 64.

9 Evans, 65.



Figure 1. “female” + “dark skin” + “microphone” = “the dark-skinned female singer,” in Vyvyan Evans, *The Emoji Code: The Linguistics behind Smiley Faces and Scaredy Cats* (New York: Picador, 2017), 214.

Evans’s book serves as a good starting point for understanding the strengths and limitations of emojis in practice. However, it does not have take the opportunity to elaborate on specific case studies before moving forward to the discussion on cultural implications. Both the word-like replacements and the digital “grammar” are informative observations, but there is certainly more nuanced work to be undertaken. Patrick Georg Grosz et al.’s newly published research endeavored to fill this gap through a systematic content analysis of face emojis on Twitter. Other than the practical replacement of words (as is introduced in Evans’s book) or the “metalinguistic / presentational use” (e.g., “*I just woke up like this:* 😊,” borrowing the pictorial features of emojis in explanation of textual information), this article focuses on a particular use of emoji, where an either positive or negative facial emoji is placed at the very end of a completed sentence, in order to “comment on” the textual information.<sup>10</sup> The study contains very detailed semantic analyses on this type of “commentary” emojis with the help of syntax and formal logic. Through manipulating the order of the sentence and the emoji attached to it, the researchers found some combinations to be valid and consistent, while other combinations appear to be self-contradictory. Namely, emojis are not randomly inserted in arbitrary parts of sentences. Whether a combination of “a sentence plus a commentary emoji” makes sense or not tends to follow some linguistic patterns.

During the case study, the Twitter posts were collected as samples only when the emojis attached are facial expressions that can be easily categorized as either “positive” or “negative.” That is, non-facial emojis (such as foods and plants) or facial emojis with rather ambiguous categorizations according to Emojipedia (such as 🤖, 😬, 🤩) are beyond the scope of this study. By “commentary,” the authors refer to cases where the emoji shows the user’s attitudes towards “an individual or

---

<sup>10</sup> Patrick Georg Grosz et al., “A Semantics of Face Emoji in Discourse,” *Linguistics and Philosophy* (22 Feb. 2023), 908–909.

proposition expressed (or presupposed) by the text,” instead of telling their feelings regarding either themselves or the message recipients.<sup>11</sup> In short, an emoji functions as “commentary” only when it is pointing towards part of the *text*, rather than the *context*. The authors also provided a couple of examples to show the difference in emoji-text relations: while the emoji in “Did you see that guy? 😄” is commentary (speaking of “that guy”), the case of “How did the interview go? 😊” is not the same (showing friendliness towards the recipient rather than speaking of “the interview”).<sup>12</sup>

When the emojis function as “commentary” in an online message, they appear to follow certain grammar-like rules during the meaning-making process. First, the commentary emojis are subject to the word order, highly dependent on the latest (closest) part in the prevailing sentence structure. The authors present an interesting comparison: while “I’m really hungry 😞 just ordered some food” is consistent, “I’m really hungry, just ordered some food 😞” is counter-intuitive, since the emoji marks the notion of “ordering some food” as negative.<sup>13</sup> Second, the emoji is also sensitive to the way the sentence is addressed—namely, word choices, highlights, as well as rhetorics of objective facts all make significant differences to the overall “positive” or “negative” feelings. A representative example is peoples’ use of emojis to describe their impressions on future winnings rates. Compared with the absolute size of percentages, the researchers found that qualitative descriptions are more relevant to the choice of “proper” emojis. Regardless what the actual probability percentage given for winning is, if “there is a xx% chance that we will *win*” is followed by a positive emoji, the meaning is one thing, while replacing “win” with “lose” contributes to the direct opposite case. When the estimation of the probability of winning is placed in a more complex discussion, as Fig. 2 shows, the subjective descriptions (such as “possible,” “likely” and “only”) are all associated with the emotions embedded in the message. Finally, the researcher concluded that emojis function in social discourses in a way similar to emotive markers in speeches (e.g., *wow*, *yay*, *oops*, *fortunately*, *sadly*), for both emojis and emotive

---

<sup>11</sup> Grosz et al., 908.

<sup>12</sup> Grosz et al., 908–909.

<sup>13</sup> Grosz et al., 917–919.

markers can accompany the text and manifest its tone, but they are unable to carry independent meanings.<sup>14</sup>

- a. Q1 = Is it **possible** for us to win?  
A1a: (**Yes.**) There's a 50% chance we'll **win**. 😊
- b. Q1 = Is it **possible** for us to win?  
A1b: (**Yes.**) There's **only** a 50% chance we'll **lose**. 😞
- c. Q2 = Are we **likely** to win?  
A2a: (**No.**) There's a 50% chance we'll **lose**. 😞
- d. Q2 = Are we **likely** to win?  
A2b: (**No.**) There's **only** a 50% chance we'll **win**. 😞

Figure 2. Word choices contributing to emoji-text relations, in Patrick Georg Grosz et al., "A Semantics of Face Emoji in Discourse," *Linguistics and Philosophy* (22 Feb. 2023), 922

Taking a different approach, Paggio and Tse's eye-tracking study on the processing of emoji introduces new insights into the linkage between "emojis" and "words" from the reception side.

The researchers conducted two experiments. In each case, the participants were required to read out sentences that are either purely textual, or that have an emoji replaced / accompanied with a word. The only difference between the two experiments is the range of emojis: while one group consists of emojis with more fixed meanings, emojis in the other group have rather debatable interpretations. The results show that sentences with an emoji replacing a word take more time to process in both cases, while the time increase is much greater if the emoji's meaning is ambiguous. Those emojis with ambiguous meanings also draw more visual attention no matter whether they are replacing the word or placed beside the word. It is concluded that while emojis in a sentence can catch the reader's attention, they may create barriers to accessing the information when they are used in substitution of words. The barrier is especially significant when the emoji does not have a shared concrete interpretation.<sup>15</sup>

---

<sup>14</sup> Grosz et al., 948.

<sup>15</sup> Patrizia Paggio and Alice Ping Ping Tse, "Are Emoji Processed like Words? An Eye - Tracking Study," *Cognitive Science*,

It is noteworthy, however, that this study derives from a supposed bifurcation of logographic and phonetic writing systems: “[s]ince emoji do not have a specific associated sound, we may expect their processing to be more similar to that of characters in logographic languages than to words in alphabetic ones.”<sup>16</sup> As is reflected in Paggio and Tse’s article, emojis, just like the so-called ideographs, are believed to carry meanings “without access to the phonology of the symbol as a prerequisite of lexical access,” and thus require more processing time than English texts.<sup>17</sup> The association between emoji and “logographic languages” sounds promising—but it actually relies upon an often-debated premise: the long-standing belief that Chinese and other non-alphabetical writing systems “inscribe ideas, and [are] thus at least potentially independent of (spoken) language.”<sup>18</sup> Despite its venerable position in the history of regional studies, this “ideographic myth” is commonly criticized for overlooking the morpho-syllabic feature of Chinese characters. More importantly, the critique of the ideographic myth points out that, not only Chinese, but “all writing is necessarily phonographic, and that writing systems based on the direct representation, not only of ideas, but even of words, do not exist in practice and are in fact impossible.”<sup>19</sup> The emphasis on sound corresponds with Donald J. Foss and David T. Hakes’s famous “subvocalization hypothesis,” which necessitates a process of “talking to oneself and listening to what one says” in text processing. It is through converting text into “subvocal speech” that the written information becomes accessible.<sup>20</sup> The potential linkage among emoji, ideogram and subvocalization invites further discussions: (1) If the relevance between shapes and pronunciations in different writing systems constitutes a spectrum, should emoji—the pictorial “script” without available pronunciations—be placed outside the spectrum? (2) Are emojis processed like

---

vol. 46, no. 2, Feb. 2022, <https://doi.org/10.1111/cogs.13099>. Accessed 28 Feb. 2022.

<sup>16</sup> Paggio and Tse, 4.

<sup>17</sup> Paggio and Tse, 4.

<sup>18</sup> David B. Lurie, “Language, Writing, and Disciplinarity in the Critique of the ‘Ideographic Myth’: Some Proleptical Remarks,” *Language & Communication*, vol. 26, no. 3–4 (July 2006): 250–69, <https://doi.org/10.1016/j.langcom.2006.02.015>.

<sup>19</sup> Lurie, 251.

<sup>20</sup> Donald J. Foss and David T. Hakes, *Psycholinguistics: An Introduction to the Psychology of Language Contents* (Prentice Hall, 1978), 330.

pictures, or subvocalized like words? To pursue a deeper understanding of the linguistic functioning of emoji, further research touching upon the theory of subvocalization might be helpful.

TRANSCULTURAL “NATIVE SPEAKER STATUS” VS. CULTURAL SPECIFICITY IN MEANING-MAKING: REIMAGINING A COSMOPOLITAN “UNIVERSAL LANGUAGE”

Emoji, available worldwide and vivid in depicting universal human emotions, is commonly associated with the capacity to bridge heterogeneous global cultures together—an attractive blueprint that is very similar to the symbolic meaning of Esperanto. Additionally, compared with Esperanto, emoji might be considered more suitable for the cosmopolitan outlook: while the supposedly “universal” Esperanto embeds more commonalities with Indo-European languages and potentially contributes to divergent learning curves, the access to emojis is shared equally by smart device users around the globe. Everyone typing with a smartphone shares a similar “native speaker” status in this system of colorful and creative pictorial symbols. Vyvyan Evans appreciated highly the so-called “unifying and non-threatening nature” of emoji in his book: “[p]erhaps the world can, indeed, be united for the better by this new, quasi-universal form of communication.”<sup>21</sup> However, as we can infer from Paggio and Tse’s study, not every emoji has a universally acknowledged interpretation. The emojis with less clear significations actually create barriers to the reception of digital texts. And according to Philip Seargeant’s book *The Emoji Revolution*, emoji, along with all the other prevailing imaginings and projects for a “universal language,” is far from being universal: “despite emoji having achieved what the International Auxiliary Languages never fully managed, this isn’t to say that they are or ever can be a solution to the Babel of different languages which so complicates international communication.”<sup>22</sup> To what extent does emoji have the potential to transcend national borders, and to what extent might emoji replicate the Babel allegory? Can emoji ever become the ideal “universal language” in the future? Such questions require more detailed investigations on emojis in cross-cultural communications.

---

<sup>21</sup> Evans, 24.

<sup>22</sup> Philip Seargeant, *The Emoji Revolution: How Technology Is Shaping the Future of Communication* (Cambridge and New York: Cambridge University Press, 2019), 187.

The belief that emoji can be “a solution to the Babel of different languages” presupposes that those vivid, colorful “illustrations” or “pictographs” embed concrete meanings that are transparent to Internet users around the globe. However, taking the established theories of semiotics as our reference, we see that the functioning of sign systems actually embeds more complex dynamics and inner connections, in which emoji makes no exception. An emoji, serving as a *signifier* in Saussure’s theory, is identified through “patterns of equivalence and contrast” in relation to other emojis, similarly to how a word makes sense in any other language. The information conveyed in this emoji, namely the *signified*, is also defined “by patterns of equivalence and contrast the one word has established with other words in its language.”<sup>23</sup> According to Louis Hjelmslev’s elaborated *Prolegomena to a Theory of Language* of 1943, an “expression” (namely a Saussurean *signifier*) “makes no reference to an external context or ‘real world,’” but only represents “a relation inside an autonomous system.”<sup>24</sup> In short, the interpretation of an emoji is not directly reflected in its visual patterns, but loosely and randomly matched with the existing ideas and concepts that the users have in mind. Many recent studies have shown the difficulty of agreeing upon a shared interpretation to emoji, both in intra-cultural and in intercultural communities.

In their newly published study, through tracking the lexicalization mechanism of emoji, Benjamin Weissman et al. have provided a straightforward approach to measuring the ambiguity of thirty popular emojis among twenty participants. The research team conducted two experiments. One asks the participants to write down their literal “translations” of those emojis, while the other requires an evaluation of emoji-text relations by labelling them as “match” or “mismatch.” The result shows that a higher agreement on an emoji’s signification is associated with a shorter response time to evaluate that specific emoji-text relation. That is, while an emoji of more concrete meaning tends to be read more fluently, a more ambiguous emoji is less “readable,” requiring longer pauses and potentially creating barriers to the reception of information.<sup>25</sup> Similarly, David Rodrigues et al.’s 2017 study

---

23 David Lidov, “sign,” in *Encyclopedia of Semiotics* (Oxford University Press, 1998) <https://www-oxfordreference.com.proxy.library.upenn.edu/view/10.1093/acref/9780195120905.001.0001/acref-9780195120905-e-263>

24 Louis Hjelmslev, quoted in Lidov.

25 Benjamin Weissman et al., “The Lexicon of Emoji?: Conventuality Modulates Processing of Emoji,” *Cognitive Science*,



examines the varied receptions towards emojis and emoticons among the Portuguese-speaking community. The participants were supposed to observe twenty random emojis/emoticons in total, interpret their meanings, and provide subjective evaluations in multiple dimensions. The results were compiled into a newly established database called Lisbon Emoji and Emoticon Database (LEED). While the colorful emojis were reported “clearer,” “more familiar,” “more arousing” and “more meaningful” compared with the abstract, solid-colored emoticons, the subjective attributed meanings of emojis still reflected a considerable level of disagreement. Through comparing the results with descriptions by the Unicode Foundation, the researchers also noticed a constant deviation between the emoji users’ interpretations and the developers’ intentions.<sup>26</sup> That is, even if the accessibility of emoji is technically monopolized by Unicode, this top-down power structure has rather less influence on the “pragmatics” of emoji—including the negotiation of meaning behind the mass circulation.

Even within the same culture, the shared “native speaker status” of emoji does not guarantee a general agreement on grammatic rules or semantic reference, and emoji may face increased ambiguity in cross-cultural online communities. In 2019, Mengdi Li et al. conducted a large-scale analysis to compare emoji using preferences from Twitter users around the globe. The researchers found a particularly interesting phenomenon in which two emojis are considered “similar” in some countries, but “different” in other places. (For example, 🎉 and 🍰 are used to convey similar meanings in UK, but this is not the case in Japan.)<sup>27</sup> The culturally uninterchangeable “similarity” and “difference” sheds a new light on the structuralist viewpoint mentioned earlier, through which the meanings are believed to emerge not from the resemblance between the signifier and the signified, but from the relative network within a given vocabulary. The diversity in inter-emoji relations offers a great example of how emojis are lexicalized differently in different cultural contexts.

Sharath Chandra Guntuku et al. also conducted a cross-cultural comparative study, which

---

vol. 47, no. 4, Apr. 2023, <https://doi.org/10.1111/cogs.13275>. Accessed 7 Apr. 2023.

<sup>26</sup> David Rodrigues et al., “Lisbon Emoji and Emoticon Database (LEED): Norms for Emoji and Emoticons in Seven Evaluative Dimensions,” *Behavior Research Methods*, vol. 50, no. 1, 31 Mar. 2017, pp. 392–405, <https://doi.org/10.3758/s13428-017-0878-6>.

<sup>27</sup> Mengdi Li et al., “An Empirical Analysis of Emoji Usage on Twitter,” *Industrial Management & Data Systems*, vol. 119, no. 8, 9 Sept. 2019, pp. 1748–1763, <https://doi.org/10.1108/imds-01-2019-0001>.

distinguishes itself by the dichotomous mapping of “the East” (including China and Japan) and “the West” (including the United States, the United Kingdom, and Canada). The researchers collected user-generated content on Twitter and Weibo, with a special focus on the semantic variations with “universal emotion expressions” of emoji. In this study, the basic units of meaning were selected from a database called Linguistic Inquiry Word Count (LIWC), which provided culturally interchangeable “categories” covering a wide range of daily occasions. The result shows that there are fewer “East-West differences” in regard to categories like “‘Ingest,’ ‘Death,’ ‘Anger,’ ‘Money,’ ‘Home,’ and ‘Family,’” which the authors tended to interpret as “universally recognized” concepts. In contrast, when the choice of emoji reflects culturally specific lifestyles (such as “Food” and “Leisure”), or derives from grammatical rules (such as “Quantifiers” and “Number”), a higher level of distinctiveness is observed.<sup>28</sup> In particular, this study reveals a unique feature of emoji’s lexicalization process: unlike the neologisms that are imported / coined to fill in the blank in the target language, emojis are pictorial illustrations that accompany, or serve as an alternative to, pre-existing concepts. That sheds a new light on the emoji-text relationship discussed earlier: if emojis are functionally “substitutable for words,” then it is necessary to highlight the cultural specificity of “what word is substituted.” Namely the barrier for emoji to become a “universal language” derives from the lack of universally acknowledged concepts behind specific emojis.

## DISCUSSION AND CONCLUSION

In summary, many structural analyses suggest that emoji tends to share some overlapping functions with natural languages, while it is still far from an ideal substitution for written texts due to several concerns. First, emoji is not as flexible in conveying information as traditional writing systems may be. Its digitally monopolized nature makes it extremely difficult to produce new single emoji “characters” that are portable in text messages. Although it is technically possible to combine simple emojis into more complicated ones, it is still a preset system in which the users cannot make unlimited combinations like the way sentences are produced in natural languages. Second, when emojis co-exist with textual information, emojis usually appear as subordinate to texts—for they tend to make

---

<sup>28</sup> Sharath Chandra Guntuku et al., “Studying Cultural Differences in Emoji Usage across the East and the West,” *ArXiv*, April 4, 2019. <https://doi.org/10.48550/arxiv.1904.02671>.

explanations to the existing information rather than adding new information to the conversation. Therefore, "code-switching" between emojis and written texts is not the same case as code-switching between two natural languages. Finally, even if emoji can replace some vocabulary in a sentence, this is not an efficient substitution for words, due to the ambiguity of interpretations from time to time. Emojis, similar to emotive markers, highlights, or illustrations, are impressive companions to the text. Just as Philip Seargeant reminded us in his *The Emoji Revolution*, "[e]moji are not designed to substitute for natural languages... they do something quite different, supplementing verbal language in diverse ways for the expression of mood, emotion, humour and creative play."<sup>29</sup> They add footnotes to the meaning while they seldom represent the information itself.

Also, compared with traditional written texts and the digital text-based emoticon, emojis appear to be significantly impressive and communicative thanks to their colorful, straightforward illustration style. The vividness of emoji is thus often associated with a hypothetical "transcultural" or "universal" communicativeness. However, from the perspective of semiotics, it is far-fetched to assume that emojis can directly represent the signified without taking the abstraction process or the reciprocal dynamics of meaning into consideration. Although an emoji may look like a certain object, the meaning embedded still depends on the similarities and distinctions between all the available emojis (or "vocabulary"), instead of relying only on the visual resemblance. In addition, since emojis are almost always inserted into a text written in a natural language and circulated within a text-based online community, the meaning of the emoji is inevitably influenced by the sentence it accompanies, the word it replaces, or the grammatical structures of the target language. Therefore, it might not be a preferable approach to put emoji in isolation and presuppose its universality, if people seek to further explore the linguistic function of this unique digital writing system.

Finally, the current scholarship on the cultural specificity of emoji leaves an intriguing space for nuanced discussions: Facing the world of globalization and cultural multiplicity, what specific boundaries does "cultural uniqueness" point towards? Notably, the two studies on cultural differences introduced above took as an important reference Geert Hofstede's Cultural Dimension Model, which views "culture" as bounded by national borders and ranks each of the nation states according to *Power*

---

<sup>29</sup> Seargeant, 187.

*Distance, Individualism/Collectivism, Masculinity/Femininity, Uncertainty Avoidance, Long-/Short-Term Orientation*, as well as *Indulgence/Restraint*.<sup>30</sup> This model is renowned for its straightforwardness and clarity, but it is highly debated as well. The quantified indexes not only bear the risk of overgeneralization, but they also imply that the global cultural diversity is (and is only) divided into national *ethos*. With the rapid development of transnational online communities and nuanced demographic or (sub-)cultural groups, the traditional country-based evaluation methods might be reconsidered facing the growing transnational fluidity of the digital era. A lot of further questions may emerge from the current scholarship on emoji and cultural specificity: for example, does the usage of emoji vary with the borders of nation states, or is it more influenced by language communities or specific subcultural groups? If the “East vs. West” dichotomy cannot explain some culturally special usages of emoji, how can we remap the intra- and international communities in order to better represent such differences? These questions might not point towards a fixed answer, but they clearly appeal to a more diverse body of research approaches. Therefore, it is urgent that further studies respond to this new challenge of representing “diversity” and “multiplicity,” potentially beyond the national framework.

## BIBLIOGRAPHY

- Evans, Vyvyan. *The Emoji Code: The Linguistics behind Smiley Faces and Scaredy Cats*. New York: Picador, 2017.
- Foss, Donald J., and David T Hakes. *Psycholinguistics: An Introduction to the Psychology of Language Contents*. New York: Prentice Hall, 1978.
- Grosz, Patrick Georg, et al. “A Semantics of Face Emoji in Discourse.” *Linguistics and Philosophy*, 22 Feb. 2023, <https://doi.org/10.1007/s10988-022-09369-8>. Accessed 10 May 2023.
- Guntuku, Sharath Chandra, Mingyang Li, Louis Tay, and Lyle H. Ungar. “Studying Cultural Differences in Emoji Usage across the East and the West.” *ArXiv*, April 4, 2019. <https://doi.org/10.48550/arxiv.1904.02671>.

---

<sup>30</sup> Li et al., 1750–1751.

- Li, Mengdi, et al. "An Empirical Analysis of Emoji Usage on Twitter." *Industrial Management & Data Systems*, vol. 119, no. 8, 9 Sept. 2019, pp. 1748–1763, <https://doi.org/10.1108/imds-01-2019-0001>.
- Lidov, David. "sign." In *Encyclopedia of Semiotics*. Oxford University Press, 1998. <https://www-oxfordreference-com.proxy.library.upenn.edu/view/10.1093/acref/9780195120905.001.0001/acref-9780195120905-e-263>
- Lurie, David B. "Language, Writing, and Disciplinarity in the Critique of the 'Ideographic Myth': Some Proleptical Remarks." *Language & Communication* vol. 26, no. 3–4 (July 2006), pp. 250–69. <https://doi.org/10.1016/j.langcom.2006.02.015>.
- Paggio, Patrizia, and Alice Ping Ping Tse. "Are Emoji Processed Like Words? An Eye-Tracking Study." *Cognitive Science*, vol. 46, no. 2, Feb. 2022, <https://doi.org/10.1111/cogs.13099>. Accessed 28 Feb. 2022.
- Rodrigues, David, et al. "Lisbon Emoji and Emoticon Database (LEED): Norms for Emoji and Emoticons in Seven Evaluative Dimensions." *Behavior Research Methods*, vol. 50, no. 1, 31 Mar. 2017, pp. 392–405, <https://doi.org/10.3758/s13428-017-0878-6>.
- Seargeant, Philip. *The Emoji Revolution: How Technology Is Shaping the Future of Communication*. Cambridge and New York: Cambridge University Press, 2019.
- The Unicode Consortium. "Emoji Statistics." [emojipedia.org](https://emojipedia.org). Accessed December 17, 2023. <https://emojipedia.org/stats>.
- Weissman, Benjamin, et al. "The Lexicon of Emoji? Conventinality Modulates Processing of Emoji." *Cognitive Science*, vol. 47, no. 4, Apr. 2023, <https://doi.org/10.1111/cogs.13275>. Accessed 7 Apr. 2023.

All issues of *Sino-Platonic Papers* are accessible to readers at no charge via our website.

To see the complete catalog of *Sino-Platonic Papers*, visit

[www.sino-platonic.org](http://www.sino-platonic.org)