



Chinese Characters

Leon Z. Lee

A Quick Social, Political and Linguistic Survey

Introduction

An old American proverb states "The only history worth knowing is the one you do not already know." This outlook is quite evident as trade globalization rapidly unites people, cultures and languages between the East and West in gaining insights to form strategic business relationships. From afar, Chinese characters may seem like an insurmountable communication barrier, but once introduced to the socio-linguistic and political intricacies of the Chinese language, one will find both uniqueness and commonalities in comparison to one's own language and history. This survey will introduce the political, ethnic and cultural lineage of Chinese characters in East Asia and its impact on the modern online information realm.

Pan-Asian Legacy

Chinese characters are used for daily print communication in China, Taiwan, Japan and Singapore as exemplified by Japan's [Asahi Simbun](#) and China's [China Web](#). South Korea uses the characters sparingly for certain nouns, but consistently when writing personal names, as seen on the Korean news site [Digital Chosun](#). Vietnam uses the characters as part of its literary heritage, while the Chinese, Japanese and Korean diasporas in Southeast Asia, Australia, the Americas and Europe regularly employ the writings for daily communication or business events.

The rationale for the diverse use of Chinese characters includes: 1) Chinese characters were one of the first written languages introduced in East Asia circa 5,000 years ago, 2) Being character-based, their pronunciation has been able to evolve throughout the centuries, while their written component remains relatively stable, 3) Used in Classical Chinese prose, the characters became the standard written communication medium for the government and educated classes (similar to the prestige that Latin has in European cultures).

Socio-linguists have noted that the standardized orthography of the characters during the past two millennia has provided a pan-cultural foundation whereby multi-ethnic groups in East Asia can simultaneously assert regional cultural differences, yet identify with a common literary lineage. In alphabetic languages, a word can morph with its pronunciation. Take the word *Lion*, for example. The translation using *AltaVista's Babelfish* produces multiple results such as *Löwe* (German), *León* (Spanish), *Leão* (Portuguese) and *Leone* (Italian). Contrasted with Chinese characters, the form of this word retains the same orthography despite pronunciation differences.

正義常存

Figure 1. "Justice is Persistence" is verbally rendered in different Asian languages, but its semantic remains the same.

For example, the classical Chinese expression "Justice is Persistence" in Figure 1 is verbally rendered in Mandarin Chinese as "Zheng Yi Chang Cun," Japanese "Sei Shi Jou Zon," Korean "Chong Iu Sang Chon" and Vietnamese "Chih Nghia Thuong Ton." Regardless of the different pronunciations, its written form remains unaltered. (Additional information on this subject can be found at http://en.wikipedia.org/wiki/Classical_Chinese).



Figure 2. Red octagon sign instructs traffic to "stop" whether it is written in English "STOP," Portuguese "PARE," French "ARRÊT" or Chinese character "TING." Chinese characters operate in a similar ideographic manner in terms of their orthographic, semantic and pronunciation features.



This concept is similar to a driver finding a red octagon traffic sign and instinctively understanding the instruction to stop, whether the word is written in French "Arret," Portuguese "Pare," Chinese character "Ting" or English "Stop," as explained in *Stop Sign* at http://en.wikipedia.org/wiki/Stop_sign. Although Classical Chinese was replaced as the main communication medium in the early 20th century, its vestiges can still be seen in modern Asian languages, such as in the national anthem of the Republic of China (Taiwan) and in certain proclamations made by Japan's Emperor. It receives different titles depending on the locale:

Chinese <i>Wen Yen</i>	(http://www.omniglot.com/writing/chinese.htm)
Japanese <i>Kanbun</i>	(http://en.wikipedia.org/wiki/Kanbun)
Korean <i>Hanja</i>	(http://en.wikipedia.org/wiki/Hanja)
Vietnamese <i>Chu Nom</i>	(http://en.wikipedia.org/wiki/Chu_nom)

Orthographic Standards

In American and European societies, [orthography](#) is commonly associated with font typefaces (ex. Verdana, Arial, Times Roman) and impacts the alphabet's cosmetic rendition (ex. serif vs. non-serif). In Asian societies, the written format of a Chinese character has a direct sociopolitical connotation since its official composition is sanctioned by the government. Failure to appreciate this intricacy could inadvertently fuel disputes and controversies.

Although character variants have existed for thousands of years (derived from brush calligraphy and regional flair), the onset of the 20th century brought language reforms to streamline communication and public education. The result is that three official orthographic styles are in use today: 1) Traditional, 2) Simplified and 3) Reduction. The three standards are not completely exclusive and share many common characters, as shown in Figure 3.

	Traditional	Simplified	Reduction
Center	中	中	中
People	民	民	民
Vehicle	車	车	車
Door	門	门	門
Country	國	国	国
Dragon	龍	龙	竜
Gateway	關	关	関

Figure 3. Traditional – Simplified – Reduction character sets are not completely exclusive and share many common characters (outlined in red)

For example, characters for *Center* and *People* are written the same across all three formats. Words for *Vehicle* and *Door* are shared by Traditional and Reduction sets, while *Nation* is shared between Simplified and Reduction sets. The words for *Dragon* and *Gateway* are glyphs of each other.

Traditional characters (locally called *Fan-ti-zi*) are used by Taiwan, Hong Kong and the majority of the ethnic-Chinese populations in Southeast Asia, Australia, the Americas and Europe. The written components follow the official formats established in China some 2,000 years ago very closely, thereby projecting an aura of literary heritage. More information on the 12,000 characters used for daily communication can be found at http://en.wikipedia.org/wiki/Traditional_character.

Simplified characters (locally called *Jian-ti-zi*) are used by China and Singapore. They originated from China's language reform in the late 1950s, designed to accelerate national literacy and public education. Complex characters were identified by the government and simplified via stroke reduction or outright replacement. Due to China's increasing economic power and emigration, these characters are gradually gaining acceptance abroad in the 21st



century. More information on the Simplified Chinese Character set, which uses some 4,000 characters, can be found at http://en.wikipedia.org/wiki/Simplified_character.

Reduction characters (locally called *Joyo Kanji*) are used in Japan. Although these are also called Simplified characters, some professionals draw a distinct line between Japan and China's implementation. Japan was able to simultaneously reduce the stroke pattern, yet retain the character's fundamental nuance. This may seem like an esoteric exercise, but written languages in general often require passionate opinions. The system uses about 2,000 characters. More information on *Kanji* and *Japanese Kanji* can be found at <http://en.wikipedia.org/wiki/Kanji> and http://www.omniglot.com/writing/japanese_kanji.htm, respectively.

Korea also uses Traditional Chinese characters (locally called *Hanja*), although mainly for certain nouns like names of people and places. About 2,000 characters are used in this system and you can find more information at <http://www.omniglot.com/writing/korean.htm>.

Vietnam uses Chinese characters sparingly and only for *Chu Noh* and *Chu Nom* Classical Chinese renditions. Since their language was Romanized in the early 20th century, Latin alphabets (locally called *Chu Quoc-Ngu*) are used for daily written communication. More information on Vietnamese can be found at <http://www.omniglot.com/writing/vietnamese.htm>.

Political Legacy

Language debates often have social, cultural, ethnic and political overtones. This is also true in the discussion of Traditional and Simplified Chinese characters. To the novice, they may seem like different languages, but a closer examination quickly brings understanding and convergence. The two character groups are basically glyphs of each other. A sentence written in Traditional characters can be closely related to its Simplified counterpart, as shown in Figure 4.



Figure 4. Traditional and Simplified Chinese do share common characters, as in the phrase "Brazil is a very beautiful country." Character glyphs outlined in red.



The phrase "Brazil is a Very Beautiful Country" is translated into Chinese as *Baxi Shi Ge Hen Meili Di Quojia* (literally meaning "Brazil is a Very Beautiful of Country"). Note that the sentence only contains three glyph differences, with the remaining characters sharing the same format. In addition, with no direct correlation between Chinese and Brazilian Portuguese, the country name *Brasil* was phonetically sinicized as *Baxi* which literally means *Earnestly West*. This transcription process is almost a science unto itself since the Chinese characters must promote phonetic approximation while also projecting good imagery via character semantics.

The division between Traditional and Simplified character sets originated from the 20th-century [Chinese Civil War](#) between the Chinese Nationalist and Communist forces. In 1911, Nationalists established the [Republic of China](#) and continued the centuries old usage of Traditional Chinese characters. However, after a bitter 25-year civil war, the Communist secured Mainland China in 1949, established the [People's Republic of China](#) and initiated Simplified Chinese reform in the late 1950s. Cold War geopolitics caused the Nationalists to flee to Taiwan, where they established a rival government and have been there ever since. Thus, the two character sets assumed socio-political symbolism well into the 21st century via cultural affiliations and governmental loyalties.

On one hand, Traditional Chinese supporters can be considered "purists" in supporting the same orthographic style rendered some 2,000 years earlier. They often decry Simplified characters as lacking true Chinese heritage and legitimacy. On the other hand, Simplified Chinese supporters can be considered "modernists" in that expanding public literacy via language reform is of urgent necessity. Without this investment, China could not have achieved its modern economic prowess.

The open secret is that both China and Taiwan speak the same official language, Mandarin Chinese. People in China can read and understand Traditional Chinese characters since both character sets are taught in public education systems. People in Taiwan can generally comprehend Simplified Chinese characters by extrapolating their derivative form or by contextual comparison, as illustrated in Figure 4.

In more poignant terms, the conflict between Traditional and Simplified Chinese is actually a political dispute that masquerades itself as though it were a sociolinguistic dispute. This predicament is similar to Iberian versus Brazilian Portuguese. As long as the reader is willing to invest the effort, then mutual intelligibility can be attained at some level. However, if the reader is resistant to such interaction, then he or she can conveniently hide behind the vale of unintelligibility, as alluded to in the Ccaps Newsletter article [Nice Country; I'll Take It](#).

This dilemma also illustrates the preeminence of *Localization Generalists* in simultaneously formulating the interdependence between language, sociology and international relationships upon the worldwide localization industry. Simply offering raw technology solutions and competitive price points are no longer sufficient drivers. One must establish a strategic



relationship with the clientele in order to forestall market commoditization. The road to this success is to integrate all nuances of the local customer base into the business model.

Romanization Differences

Romanization is using Latin alphabets to phonetically transliterate certain languages such as Chinese, Japanese, Arabic, Hindi and Cyrillic-based Slavic. However, many Romanization standards exist and are customized for specific languages. With each standard possessing its own interpretive rules, there may be inherent social and political overtones (ex. preference of dialect, dominance of political entity, selection of a formal language). Understanding which standard is appropriate for the designated audience is crucial to seamless communication. You can find more information on this subject at <http://en.wikipedia.org/wiki/Romanization>.

Japan employs the [Hepburn Romanization](#) method, which matched Latin alphabets to corresponding Japanese phonetic scripts. For example, the Sino-Japanese word for *Up* is rendered as *Jyou* instead of *Joo*. This system also eliminated duplicate spellings from rival systems. For example, the Japanese surname *Itou* had previous duplicate spellings of *Itō*, *Itoo*, *Itoe* and *Itoh*.

China uses [Hanyu Pinyin](#), which is based on the official [Mandarin-Chinese](#) pronunciation and is part of the United Nations Mandarin Phonetic Symbols standard.

Taiwan has officially used [Tongyong Pinyin](#) since 2000. Although it is also based on the official Mandarin-Chinese pronunciation and shares about 80% compatibility with China's Pinyin standard, its adoption has been slow and sporadic at best. The majority of the Taiwanese population still relies on the early 20th-century [Wade-Giles Standard](#).

The China versus Taiwan romanizations lead to an interesting dilemma in that different spellings actually refer to the same pronunciation. For example in Figure 5, the nation *China* is written as *Zhong Guo* in China's Hanyu Pinyin, *Zhong Gwo* in Taiwan's Tongyong Pinyin and *Chung Kuo* in Taiwan's Wade-Giles, even though they are all pronounced in the exact same manner. This situation is similar to German versus Dutch spellings, such as the case with words like ship (schiff : schip), apple (apfel : appel) and stone (stein : steen).

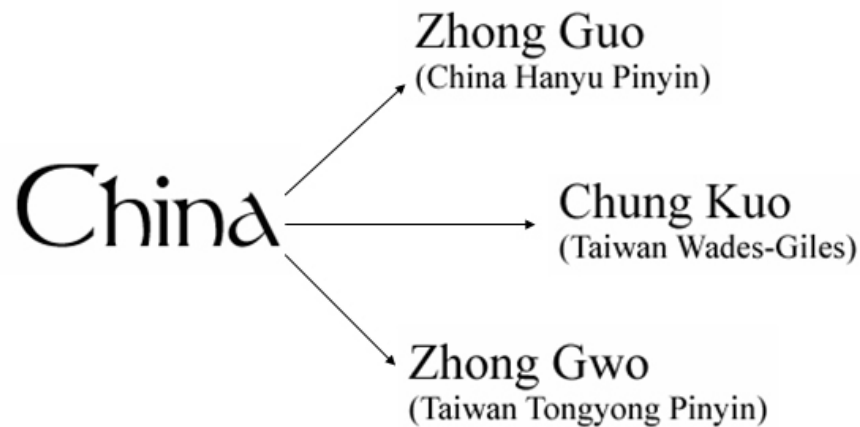


Figure 5. Chinese romanization is peculiar in that different spellings can still possess the same pronunciation. Take the case of China: even though romanized as Zhong Guo, Chung Kuo and Zhong Gwo, all words are pronounced in the exact same manner.

Korea originally used the [McCune-Reischauer system](#) before adopting the new [South Korea Romanization standard](#) in the year 2000. Although it was intended to streamline consonant representations and eliminate redundant diacritic marks, some proponents felt the original McCune-Reischauer system possessed too much Japanese influence since it was developed in 1937 when Korea was subjugated by Japan's empire.

Take for example names of places based on the old and new spelling methods, such as *Pusan* : *Busan*, *Ichon* : *Icheon*, *Cheju* : *Jeju* (for more information on Korean Romanization, please visit http://en.wikipedia.org/wiki/Korean_romanization). Nevertheless, some words remained the same in both systems, such as the capital *Seoul*.

Character Encoding Standards

Character encoding for the Internet and software industry is the convention of storing language information within a computing environment, such as the [ASCII standard](#) for Latin alphabetic languages. However, Asian languages that use Chinese characters have faced many different challenges in their encoding schemes over the past 20 years. A principal consideration is whether a specific glyph variant is stored as part of its character encoding value.

Alphabetic Encoding Method

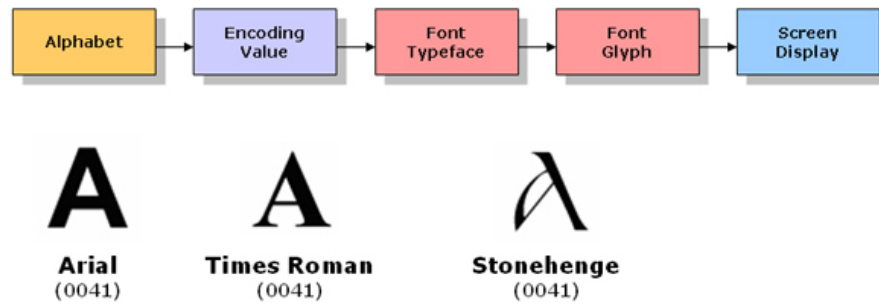


Figure 6. Alphabetic encoding does not store glyph renditions. Hence, its encoding value is not associated with a specific font typeface.

For alphabetic encoding, the alphabet is stored separately from its glyph variant, as illustrated in Figure 6. In this way, the alphabet can assume multiple font typeface displays while retaining the same encoded value. In the example above, the alphabet "A" retains its [Unicode](#) value "0041" even though its display is altered among *Arial*, *Times Roman* and *Stonehenge* font selections.

Character Encoding Method

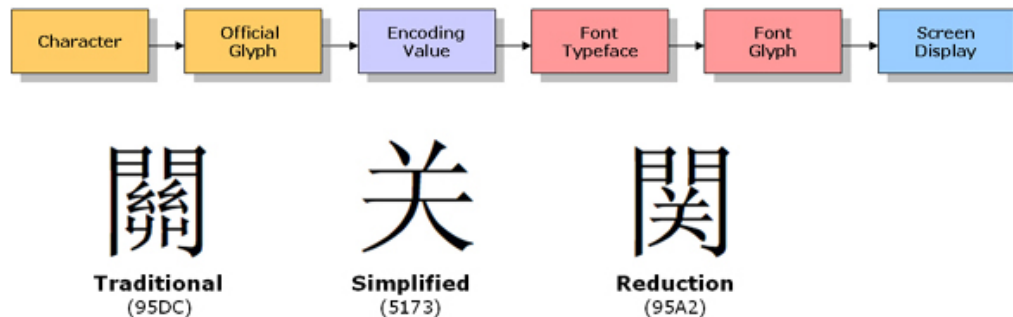


Figure 7. Character encoding does store glyph renditions based on national standards. However, the character's semantic remains the same, as in the above example for "Gateway."

For character encoding, the glyph variant is stored as part of its encoding value, as outlined in Figure 7. The character assumes a specific glyph display assigned by individual national



governments. In the example above, the character "Gateway" retains the same semantic, but assumes multiple glyph renditions and Unicode values : 1) Traditional – 95DC, 2) Simplified – 5173, 3) Reduction – 95A2.

To manage the proliferation of Chinese characters and conserve their finite encoding storage values, the [Unicode consortium](#) established the [Han Unification](#) initiative to consolidate duplicate Chinese character glyphs based on international consensus. This consensus is necessary since language falls under the jurisdiction of nation-states and not the privy of international organizations, as highlighted by the [2000 Verisign-China controversy](#).



Figure 8. The Chinese character for "Han" literally means "Chinese," and is rendered via different glyphs among the three character sets. With Han Unification, the Traditional and Reduction style characters share the same Unicode encoding value.

Based on this convention, Chinese characters with minor glyph deviations are assigned the same Unicode value, while others with more divergent appearances are attributed unique encoding values. For as objective as these evaluations may be, a certain degree of geopolitical influence is also present in these assignments.

For example, the Chinese character "Han" shown in Figure 8 is rendered via different glyphs among Traditional, Reduction and Simplified character sets. However, its Traditional and Reduction characters share the same Unicode encoding value "6F22," while its Simplified version is assigned the unique value of "6C49."

Some claim the Simplified glyph is too obscure for common recognition, while others assert that China's nationalistic and economic overtones compelled the international forum to adopt this convention. More information on these independencies is located at [Han unification](#) as well as Ken Lunde's book entitled "CJKV Information Processing," ISBN : 1-56592-224-7, 1999.



Conclusion

Chinese Character usage is a dichotomy of East Asian sociopolitical and linguistic interactions. The versatile orthography of these characters has allowed regional cultures and ethnicities to share a common literary heritage during the past two millennium. However, language reforms initiated by nation-states over the past five decades have also created divergent character glyph representations and Romanization standards. Unifying these differences will be the venue and challenge of modern Internet information societies.

Leon Z. Lee is the International Business Manager for Dell Inc. During his 15-year corporate tenure, he has worked for multiple transnationals including Nortel, IBM and Dell. His specializations include global branding, web globalization, localized marketing and knowledge management. His interests include scaled-military modeling, Japanese Kenjutsu swordsmanship and sociology.