

August 2005

*global branding corporate communication brand transcreation v*

# *Chinese Characters*

## **Chinese Characters**

A Quick Social, Political and Linguistic Survey

**Lehrmäch**

Published for



[www.ccaps.net](http://www.ccaps.net)



# *Chinese Characters : A Quick Social, Political and Linguistic Survey*

## **I. 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 converges people, cultural and language between 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 its socio-linguistic and political intricacies, one will find both uniqueness and commonalities in comparison to one's own language and history. This survey will introduce the political, ethnic, cultural lineage of Chinese characters in East Asia and its impact on the modern Internet information realm.

## **II. PAN-ASIAN LEGACY**

Chinese characters are used for daily print communication in China, Taiwan, Japan, Singapore, as exemplified by the Japan's *Asahi Simbun* (<http://www.asahi.com>) and China's *China Web* (<http://www.china.com.cn>). South Korea used it sparingly for certain nouns, but consistently when writing their personal names, as seen on the Korean news site : *Digital Chosun* (<http://www.chosun.com>). Vietnam uses the characters as part of their literary heritage. While the Chinese, Japanese, Korean diasporas in Southeast Asia, Australia, Americas and Europe regularly employ the writings for daily communication or business events.

Rationale for its diverse usage includes : 1) Chinese characters were one of the first written language introduced in East Asia circa 5,000 years ago, 2) Being character-based, its

pronunciation can evolve throughout the centuries, yet its written component remains relatively stable, 3) Used in Classical Chinese prose, it became the standard written communication medium for government and educated classes ( akin the prestige of Latin in European cultures ).

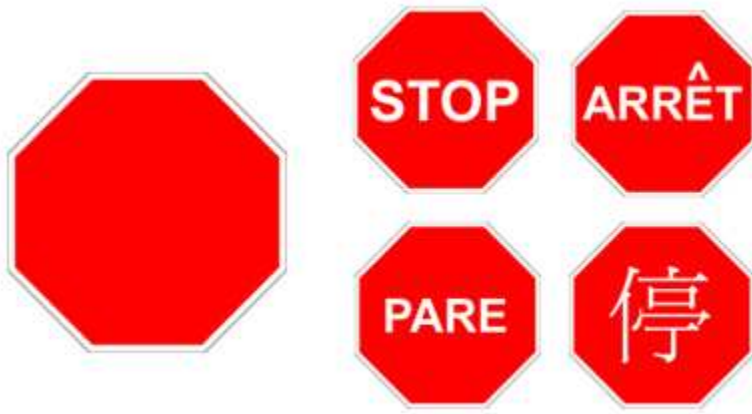
Socio-linguists have noted that the characters' standardized orthography in the past two millennium provided a pan-cultural foundation whereby multi-ethnic groups within East Asia can simultaneously assert regional cultural differences, yet identify with a common literary lineage. In alphabetic languages, a word can morph with its pronunciation. For example, the word *Lion* translated through *AltaVista's Babelfish* (<http://babelfish.altavista.com/tr>) produces multiple results such as German *Löwe*, Spanish *León*, Portuguese *Leão* and Italian *Leone*. Contrasted with Chinese characters, its form retains the same orthography despite pronunciation differences.

For example, the classical Chinese prose "Justice is Persistent" in Figure 1 is verbally rendered in Chinese-Mandarin as "Zheng Yi Chang Cun", Japanese "Sei Shi Jou Zon", Korean "Chong Iu Sang Chon" and Vietnamese "Chih Nghia Thuong Ton". Regardless of divergent pronunciations, its written form is unaltered. Additional information at Classical Chinese ([http://en.wikipedia.org/wiki/Classical\\_Chinese](http://en.wikipedia.org/wiki/Classical_Chinese)).

This concept is similar to a driver encountering a red octagon traffic sign and instinctively understands the instruction to stop, whether it

正義常存

**Figure 1.** Classical Chinese prose "Justice is Persistence" is verbally rendered in different Asian languages, but its semantic



**Figure 2.** Red octagon sign instructs traffic to “Stop” whether it is written in English “STOP”, Portuguese “PARE”, French “ARRET” or Chinese character “TING”.

Chinese characters operate in similar ideographic manner among its orthography, semantic and pronunciation features.

is written in French “Arret”, Portuguese “Pare”, Chinese character “Ting” or English “Stop”, as explained in Stop Sign ([http://en.wikipedia.org/wiki/Stop\\_sign](http://en.wikipedia.org/wiki/Stop_sign)). Although Classical Chinese was replaced as the main communication medium in the early 20<sup>th</sup>-century, its vestiges can still be seen in modern Asian languages, such as Republic of China’s (Taiwan) national anthem and certain proclamations by Japan’s Emperor. Depending on the locale, it is called by different titles, such as :

Chinese *Wen Yen*  
(<http://www.omniglot.com/writing/chinese.htm>)

Japanese *Kanbun*  
(<http://en.wikipedia.org/wiki/Kanbun>)

Korean *Hanja*  
(<http://en.wikipedia.org/wiki/Hanja>)

Vietnamese *Chu Nom*  
([http://en.wikipedia.org/wiki/Chu\\_nom](http://en.wikipedia.org/wiki/Chu_nom))

### III. ORTHOGRAPHIC STANDARDS

To American and European societies, orthography (<http://en.wikipedia.org/wiki/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, a Chinese character’s written format possesses direct socio-political connotation since its official composition is sanctioned by the government. Failure to appreciate this intricacy could inadvertently ferment disputes and controversies.

Although character variants have existed for thousands of years (derived from brush

calligraphy and regional flair), onset of the 20<sup>th</sup>-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, 3) Reduction. The three standards are not completely exclusive and do share many common characters, as shown in Figure 3.

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. However, words for *Dragon* and *Gateway* are glyphs of each other.

Traditional characters (locally called *Fan-ti-zi*) are used by Taiwan, Hong Kong and majority of the ethnic-Chinese populations in Southeast Asia, Australia, Americas and Europe. The written components follow very closely to the official formats established in China about 2,000 years ago, thereby projecting an aura of literary heritage. About 12,000 characters are used for daily communication, more information at *Traditional Chinese Character* ([http://en.wikipedia.org/wiki/Traditional\\_character](http://en.wikipedia.org/wiki/Traditional_character)) and *Chinese Script* (<http://www.omniglot.com/writing/chinese.htm>).

*Simplified characters* (locally called *Jian-ti-zi*) are used by China and Singapore. It originated from China’s language reform in the late-1950s 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 aboard in the 21<sup>st</sup>-century. About

4,000 characters are used, more information at *Simplified Chinese Character* ([http://en.wikipedia.org/wiki/Simplified\\_character](http://en.wikipedia.org/wiki/Simplified_character)).

Reduction characters (locally called *Joyo Kanji*) are used in Japan. Although it is also called Simplified characters, some professionals draw a distinct 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 solicit passionate opinions. About 2,000 characters are used, more information at *Kanji* (<http://en.wikipedia.org/wiki/Kanji>) and *Japanese Kanji* ([http://www.omniglot.com/writing/japanese\\_kanji.htm](http://www.omniglot.com/writing/japanese_kanji.htm)).

Korea also uses Traditional Chinese characters (locally called *Hanja*), but mainly for certain nouns like personal or place names. About 2,000 characters are used, more information at *Hanja* (<http://en.wikipedia.org/wiki/Hanja>) and *Korean* (<http://www.omniglot.com/writing/korean.htm>).

Vietnam uses Chinese characters sparingly and only for *Chu Noh* and *Chu Nom* Classical Chinese renditions ([http://en.wikipedia.org/wiki/Chu\\_nom](http://en.wikipedia.org/wiki/Chu_nom)). Since their language has been Romanized in the early 20<sup>th</sup>-century, Latin alphabets (locally called *Chu Quoc-Ngu*) are used for daily written communication, more information at *Vietnamese* (<http://www.omniglot.com/writing/vietnamese.htm>).

#### IV. POLITICAL LEGACY

Language debates often instill social, cultural, ethnic and political overtones. This is also true between Traditional and Simplified Chinese characters. To the novice, they may seem like different languages, but closer examinations quickly bring understanding and convergence. The two character groups are basically glyphs of each other. A sentence written in Traditional characters can possess a high degree of correlation to its Simplified counterpart, as shown in Figure 4.

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 the sentence only contains three glyph differences, with the remaining characters sharing the same format.

	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 do share many common characters (outlined in red)

In addition, with no direct correlation between Chinese and Brazilian Portuguese, the country name *Brasil* was phonetically sinicized into Chinese as *Baxi* which literally means *Earnestly West*. This transcription process is almost a science unto itself since the Chinese characters must promote phonetic approximation, yet also project good imagery via character semantics.

The schism between Traditional and Simplified character sets originates from the 20<sup>th</sup>-century Chinese civil war ([http://en.wikipedia.org/wiki/Chinese\\_Civil\\_War](http://en.wikipedia.org/wiki/Chinese_Civil_War)) between the Chinese Nationalist and Communist forces. In 1911, Nationalists established the *Republic of China* ([http://en.wikipedia.org/wiki/Republic\\_of\\_China](http://en.wikipedia.org/wiki/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* ([http://en.wikipedia.org/wiki/People%27s\\_Republic\\_of\\_China](http://en.wikipedia.org/wiki/People%27s_Republic_of_China)), and initiated Simplified Chinese reform in late-1950s. Due to Cold War geopolitics, the Nationalist fled to Taiwan, established a rival government, and has been there ever since. Thus, the two character sets assumed socio-political symbolism well into the 21<sup>st</sup>-century via cultural affiliations and governmental loyalties.

On one end, Traditional Chinese supporters can be considered “purists” in supporting the same orthographic style rendered some 2,000 years previously. They often decry Simplified characters as lacking true Chinese heritage and legitimacy. On the other end, 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 of 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 its derivative form or by contextual comparison, as illustrated 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.

In more poignant terms, the conflict between Traditional and Simplified Chinese is actually a political dispute that masquerades itself as though it was a socio-linguistic dispute. This predicament is similar to Iberian versus Brazilian Portuguese. So 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 one can conveniently hide behind the vale of unintelligibility, as alluded to in the CCAPS.net article “Nice Country; I’ll Take It” comparing Iberian and Brazilian Portuguese mutual intelligibility.

This dilemma also illustrates the preeminence of Localization Generalists in simultaneously formulating the interdependence of language,

sociology and international relationships upon the worldwide localization industry. Simply offering raw technology solution 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 avenue to this success is to integrate all nuances of the local customer base into the business model.

## V. 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, more information at *Romanization* (<http://en.wikipedia.org/wiki/Romanization>).

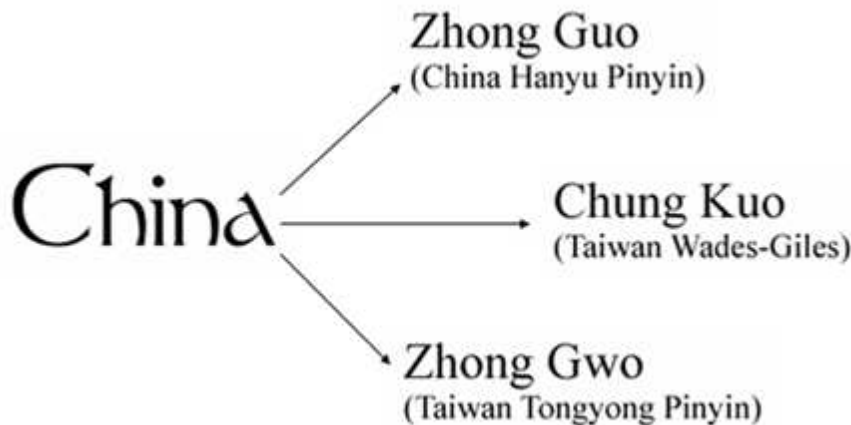
Japan employs the *Hepburn Romanization* (<http://en.wikipedia.org/wiki/Hepburn>) method which matched Latin alphabets to corresponding Japanese phonetic scripts. For example, the Sino-Japanese word for *Up* is rendered as *Jyou* instead of *Joo*. This system also eliminated duplicate spellings from rival systems. For example, the Japanese surname *Ito* had previous duplicate spellings of *Itô*, *Ito*, *Itoe*, *Itoh*.

China uses *Hanyu Pinyin* (<http://>

[en.wikipedia.org/wiki/Pinyin](http://en.wikipedia.org/wiki/Pinyin)) which is based on the official Mandarin-Chinese ([http://en.wikipedia.org/wiki/Mandarin\\_Chinese](http://en.wikipedia.org/wiki/Mandarin_Chinese)) pronunciation and is part of the United Nations Mandarin Phonetic Symbols standard.

Taiwan officially uses *Tongyong Pinyin* ([http://en.wikipedia.org/wiki/Tongyong\\_pinyin](http://en.wikipedia.org/wiki/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. Majority of people in Taiwan still rely on the early 20<sup>th</sup>-century *Wades-Giles* (<http://en.wikipedia.org/wiki/Wade-Giles>) standard.

The China versus Taiwan romanizations lead to an interesting dilemma by which 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 words for *Ship* ( *schiff* : *schip* ), *Apple* ( *apfel* : *appel* ), and *Stone* ( *stein* : *steen* ).



**Figure 5.** Chinese Romanization peculiarity in that different spellings can still possess the same pronunciation.

As in the case of China being romanized as *Zhong Guo*, *Chung Kuo*, *Zhong Gwo*.

Yet all are pronounced in the exact same manner.

Korea originally used the McCune-Reischauer (<http://en.wikipedia.org/wiki/McCune-Reischauer>) system, before adopting the new *South Korea Romanization* ([http://en.wikipedia.org/wiki/Revised\\_Romanization\\_of\\_Korean](http://en.wikipedia.org/wiki/Revised_Romanization_of_Korean)) standard in 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.

For example, comparing place names between the old and new spelling methods, these include *Pusan* : *Busan*, *Ichon* : *Icheon*, *Cheju* : *Jeju*, more information at *Korean Romanization* ([http://en.wikipedia.org/wiki/Korean\\_romanization](http://en.wikipedia.org/wiki/Korean_romanization)). Nevertheless, some words remained the same in both systems, such as the capitol *Seoul*.

## VI. CHARACTER ENCODING STANDARDS

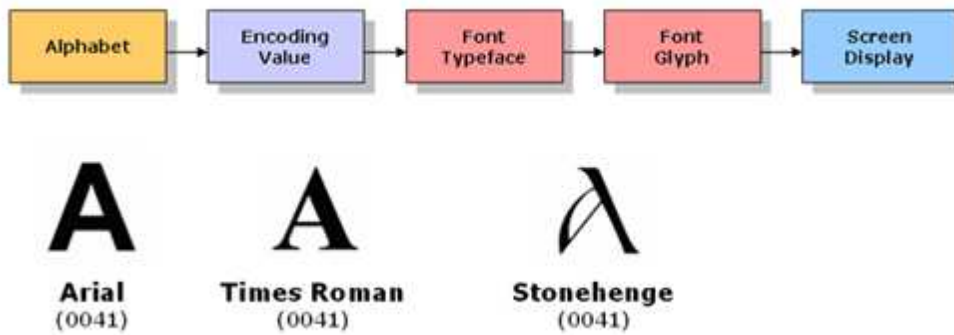
Character encoding for the Internet and software industry is the convention of storing language information within a computing environment, such as the ASCII standard (<http://en.wikipedia.org/wiki/Ascii>) for Latin alphabetic languages. However, for Asian languages using Chinese characters, their encoding schemes have encountered many challenges in the past 20 years. Principal consideration is whether a specific glyph variant is stored as part of its character encoding value.

For alphabetic encoding, the alphabet is stored separately from its glyph variant, as illustrated

in Figure 6. This enables the alphabet to assume multiple font typeface displays while retaining the same encoded value. In the example above, the alphabet "A" retains its Unicode (<http://en.wikipedia.org/wiki/Unicode>) value "0041" even though its display is altered among Arial, Times Roman, and Stonehenge font selections.

For character encoding, the glyph variant is stored as part of its encoding value, as outlined

### Alphabetic Encoding Method



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

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.

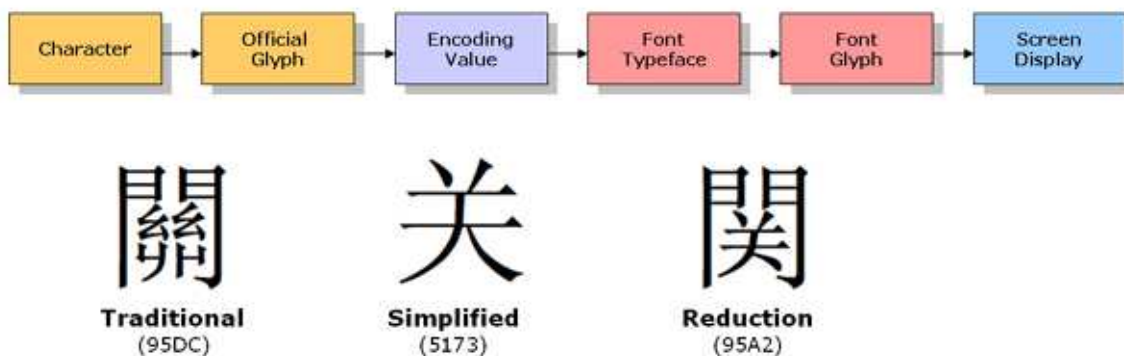
To manage the proliferation of Chinese characters and conserve its finite encoding storage values, the Unicode consortium (<http://www.unicode.org/>) established the *Han Unification* (<http://www.unicode.org/charts/unihan.html>) initiative to consolidate duplicate Chinese character glyphs based on international 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* (<http://www.newsfactor.com/perl/story/6310.html>).

Based on this convention, Chinese characters possessing minor glyph deviations are assigned the same Unicode value, while others with more divergence appearances are relegated unique encoding values. Objective as these evaluations may be, a certain degree of geo-political influence are also present in these assignment.

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", yet its Simplified version is assigned the unique value "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

### Character Encoding Method



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



**Figure 8.** Chinese character for "Han" literally means "Chinese", 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.

independencies is located at Han unification ([http://en.wikipedia.org/wiki/Han\\_unification](http://en.wikipedia.org/wiki/Han_unification)) and the book : *CJKV Information Processing*, ISBN : 1-56592-224-7, 1999.

## VII. CONCLUSION

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

**ABOUT THE AUTHOR**

Leon Z. Lee is Lehrmach's Senior Strategist for Brand Management, his 15-year corporate tenure includes Nortel, IBM, University of Texas at Austin, and Dell. He can be reached at [lee@lehrmach.com](mailto:lee@lehrmach.com)

**ABOUT THE COMPANY**

Lehrmach is a management consultancy with concentrations in transnational marketing, brand transcreation, corporate communication, and online globalization. For additional information, contact [lehrmach@lehrmach.com](mailto:lehrmach@lehrmach.com)

# Lehrmach

Lehrmach Branding  
Lehrmach Communication  
Lehrmach Strategy

[www.lehrmach.com](http://www.lehrmach.com)  
[lehrmach@lehrmach.com](mailto:lehrmach@lehrmach.com)

**Round Rock**  
3002 briar oak lane  
rock round, texas usa 78681-2294  
512 / 431.3320