Comparative Analysis of Codification of Chinese Characters and English Words

Lin Yang
Foreign Languages College
Inner Mongolia University
No. 235, Da Xue Road
W. Saihan District
Hohhot, Inner Mongolia
P. R. China

Abstract

By making comparative analyses of language encoding ways of Chinese characters and English words, a very important branch in sociolinguistics, we know that the codification of two different languages incarnates their own social and cultural background, on which sociolinguistics is based and mastering encoding ways of two languages will be conducive to understanding deep meaning of languages and use them more correctly.

Keywords: comparative analysis; codification; Chinese characters; English words

I. Introduction of Language as Social Semiotics

A society is any group of people who are drawn together for a certain purpose or purposes. A language is what the members of a particular society speak. Therefore, sociolinguistics is concerned with investigating the relationships between language and society with the goal being a better understanding of the structure of language and of how languages function in communication, i.e. sociolinguistics is "the study of language in relation to society" (Wardhaugh, 2000: 12). When two or more people communicate with each other in speech, we can call the system of communication that they employ a code. In most cases that code will be something we may also want to call a language. Code is a term which is used instead of language, speech variety, or dialect. It is sometimes considered to be a more neutral term than the others. Or it is defined as any system of signals which can be used for sending a message. A natural language is an example of a code, as are Morse code, Braille, and sign language (Richards, 2002:71). Use of the term code allows us to use derivative terms like codification and re-codification. Writing systems are said to be codifications of speech. How do the English and Chinese writing systems differ as codifications of their respective languages?

Characters are writing symbols for recording and passing on what is spoken. The philology is a branch of linguistics dealing with the nature, composition and evolution of a language. Language as social semiotics[1] is a very important branch in sociolinguistics. Linguists have made a study of language encoding way through the social interpretation of language and meaning and made following conclusions: a social reality (or a "culture") is itself an edifice of meanings – a semiotic construct; in this perspective, language is one of the semiotic systems that constitute a culture; one that is distinctive in that it also serves as an encoding system for many (though not all) of the others (Halliday, 2001: 2). This in summary terms is what is intended by the formulation "language as social semiotic". It means interpreting language within a socio-cultural context, in which the culture itself is interpreted in semiotic terms – as an information system, if that terminology is preferred.

II. The General Introduction to Three Writing Systems in the World

Language can be realized either by speech with sound as medium or by writing with visual symbols as medium. Following the iconic stage, writing made its appearance only when there was a fixed correspondence between the pictures or inscriptions on clay, bones, shells or stones and the language. These inscriptions fall mainly into two kinds: pictogram and ideogram.

---

[1] Semiotics is the theory of signs; the analysis of system using signs or signals for the purpose of communication (semiotic systems). The most important semiotic system is human language, but there are other systems, e.g. Morse code, sign language, traffic signals.

The writing system has now become a level in a written language which is parallel to phonetics, phonology, morphology, syntax and semantics. To those languages adopting the alphabetic writing system, it is given the name orthography. Out of hundreds of written-languages today, writing systems fall into three types: word writing system or ideographic, based on words, syllabic writing system, based on syllables, and sound writing system or alphabetic writing system, based on sounds. Chinese is the most representative of the first set. Chinese character is a linguistic unit in a phonological, lexical, syntactic and semantic sense and thus governed by these levels of language. Each character is a syllable embracing consonant, vowel and tone. The value of a character can be mapped into the formula: IXI =1. That is in each character, no matter how many consonants and vowels occur, the result is one syllable, which can be utilized in communication as each is a free-standing unit. Each character also represents an individual word or morpheme in relation to meaning. Therefore, it can also be called "morpheme-writing system". Some longer words may have more than one morpheme, such as "*maimai*" (Hu, 2002: 144). Japanese belongs to the second group. The third set has dominated most of the world languages. It is a writing system made up of separate letters which represent sounds. Since most European alphabets such as Spanish, German, French, English, etc. derived from the Latin alphabet, only minor adjustments were needed to cope with individual characteristics of a particular language.

Before I make comparative analyses of encoding ways of Chinese characters and English words, we should identify such a fact: language is arbitrary. Language consists of arbitrary symbols. Language is arbitrary in that there is no logical relationship between words and the objects, actions or concepts these words refer to. Why is the same object is called a "table" in English or a "zhuō zi" in Chinese? No underlying reason can be found except that these are arbitrary symbols chosen by convention. However, language is systematic, which means that it is rule governed. This is why human beings can make a summary of different nations' ways to encode their languages.
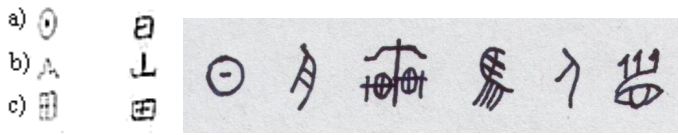
III. The Encoding Way of Chinese Characters

The Chinese script is one of the oldest and most widely used writing systems in the world. It has a history of five or six thousand years, and is used by about one fourth of the total population on earth. The creation and evolution of Chinese characters are closely interwoven with the development of Chinese culture. Chinese characters are the basic carriers of the traditional Chinese culture, and, as an important tool for extending, spreading and exchanging ideas, they have played a critical role in the long history of the Chinese nation. One may well argue that without Chinese characters Chinese culture would not have achieved the splendors it did.

The Chinese script is an ideographic writing system, in which the graphic structure is directly related to the meaning. Hence the first step toward mastery of Chinese characters is to learn the characteristics of their composition, i.e. a traditional theory known as Liu Shu (six writings). The Han Dynasty scholar Xu Shen (147 AD-?) who wrote the great etymological dictionary *Shuowen jiezi* (*Explaining Simple and Analyzing Compound Characters*, or *An Analytical Dictionary of Chinese Characters*) divided the Chinese characters into six different types (Liu Shu): That is, there are six types of characters in the terms of their composition: pictographic, indicatives, ideographs, phonetic-semantic compounds, mutual explanatory, and phonetic loans. Strictly speaking, only the first four refer to the ways of composing Chinese characters, the last two are concerned with the ways to use them. In Xu Shen's dictionary, reputedly 4% of the characters are pictographic and 1% are indicatives, while 13% are ideographs and 82% are phonetic-semantic compounds (later scholars argue many of these also have an element of "hui yi"). The traditional view that Liu Shu is a summary of the different ways of composing characters, therefore, is not very accurate. Nevertheless the Liu Shu theory is basically correct in revealing the general pattern in the creation and development of Chinese characters. It may help learners better understand the composition of Chinese characters and their original meanings, and hence use them more accurately.

1. A **pictograph** is a depiction of a material object. Chinese characters mostly originated from picture writing. In other words, most Chinese characters were originally pictures of objects.

A small group of characters that retain a close connection with original pictograms, for instance the forms for (a) "sun", "day" (*ri*), (b) "mountain" (*shan*) and (c) "field' (*tian*).

(日 sun 月 moon 车 carriage 马 horse 人 man 眉 eyebrow)

Pictographs are based on the external form of material objects, but the abstract concepts in language are formless, which renders it impossible to depict them. This impossibility inevitably hinders the growth of pictographs, and that is why their number is limited. However, pictography remains the most important method of composing Chinese characters. The others are only developments on this method and variations.

2. **Indicatives** refer to the way of forming abstract characters with indicating signs. There are two subtypes of indicatives: one is composed purely of abstract sign, e.g. ⌣(on the top), ⌢ (underneath), ・(one), ・(two) and ・(three) the other is composed of a pictograph and an indicating sign, e.g. ⅄(knife-edge), 朩(root), ⅄朩(treetop).

Following characters belong to the first category:

上(up)   下(down)   中(down)   一(one)

Following characters belong to the second category:

 (root) The original meaning of "tree" is wood. The original meaning of "root" is the stem of wood. The sign "一" means on the stem.（本）

 (end) The original meaning of "end" is the tip of a tree and the sign "一"means on the treetop.（末）

 (the edge of a knife) adding the indicative signals on the pictographic "knife"（刃）

 (also) The original meaning of "also" is axilla. "大"indicates body, adding signals under axille.（亦）

Indicatives account for the smallest percentage of Chinese characters. The reason is that for most characters there are simpler ways of composition: characters referring to material objects may be composed pictographically and those expressing abstract concepts may be composed ideographically or by way of phonic-compounding.

3. **Ideographs** are compounds, composed of two or more existing characters. In terms of structure, an ideograph is a composition of two or more characters side by side or one on top of another. In terms of meaning, an ideograph is also a composition of the meanings of its component characters.

 (step) Two feet on both sides indicate walking.（步）

 (wade) "water" refers to the river. Walking in the river means wading.（涉）

（采）"爪'指'手', "朩"为果。
　　(pluck) "爪'means hand and "果'means fruit, so pick fruit with hands.

 (rest) "Man" has a rest under a tree.（休）

 (seedling) The grass grown in the field refers to the seedling.（苗）

(coldness) "⌂"refers to a house, "ΨΨ"refers to grass and "◇" is ice. The weather is cold and the ground is frozen, so a man is in the grass.（寒）

4. A **Phonetic-semantic compound** consists of semantic radical and a phonetic radical. The semantic radical indicates its semantic field and the phonetic radical, is to remind the reader of how the word is to be pronounced. For example, the word "mother"（妈*ma*） is expressed by the semantic element "woman" (女) followed by a phonetic indicator *ma* 马The word for "scold" is also *ma* 骂(with a different tone), and this is expressed by the semantic element "mouth" 口(repeated) followed by the same phonetic indicator. In both cases the meaning of the *ma* character when used alone ("horse") is disregarded.

There are many objects and abstract ideas are difficult to express through pictography or ideography. For example, "看" is the general term for see or watch, but there are many types of see and it is impossible to differentiate each of them by way of pictography or ideography. But this is easily achieved in phonetic-semantic compounds by adding different phonetics to the radical 目, e.g. "瞩"(gaze; look steadily),"瞻"(look up or forward),"督"(supervise or oversee),"瞥"(shoot a glance at; snatch a glimpse),"瞰"(look down from a height; overlook),"盯"(stare at; fix one's eyes on),"瞅"(look at),"睽"(stare; gaze),"瞄"(fix one's gaze on),"瞵"(watch with open eyes),"睬"(take notice of; pay attention to),"瞧', (look or see) etc. Thus there are an enormous number of phonetic-semantic compounds in Chinese, and this number is growing larger and larger in the modern period. Statistics show that phonetic-semantic compounds accounted for 80% of the total characters in *The Origin of Chinese Characters* of the Han Dynasty, 88% in Aspects of the Six Categories of Chinese Characters (Liu Shu) of the Song Dynasty, and 90% in Kangxi Dictionary of the Qing Dynasty. In the modern simplified form currently in use, phonetic compounds make up an even larger percent.

5. **Mutual Explanatories**   According to Xu Shen's definition in his Origin of Chinese Characters, mutual explanatories are those which share the same semantic radicals and are mutually explainable. For example, the characters "考' and "老', both having the age radical and meaning the same, are mutual explanatories in ancient Chinese because the first meaning of "考' is oldness and old age in ancient Chinese . Thus *The Origin of Chinese Characters* says, "考' means "老' and vice verse.

Strictly speaking, mutual explanation is a way to explain the meaning of characters through comparison. There is no new character created in this way. Hence mutual explanation is not a way of composing new characters, but a way of using existing ones.

6. The **phonetic loan** is another way of using existing characters. Xu Shen defined it as a character which is used in a new sense on account of its pronunciation. In other words, it is an internal borrowing on the basis of pronunciation: a character is used in a new meaning which is expressed by a similar sound in the spoken form. In this way an existing character has acquired a new meaning, but no new character is created. The phonetic loan, therefore, like mutual explanatories is not a way of composing Chinese characters. At the early stage, there were quite a few phonetic loans in the writing system. As the number of existing characters at that time was limited, many concepts had to be expressed by phonetic loans. For example, the character "自'in the Oracle-Bone Inscriptions was originally pictograph and referred to the nose, but it is now used in the sense of "self" as a result of phonetic loan. "自' is "鼻', because is looks like "鼻'. They have the same pronunciation in ancient times. There is a record of "疾自' in the Oracle- Bone Inscription. The original meaning is disease of nose. In ancient books, it was used as pronouns, prepositions and then became a common word. Later, the character "鼻' was created and used as its original meaning, by using the pronunciation of "自'. The character 屮 in the Oracle-Bone Inscriptions was also a pictograph, referring to the wheat（麦）, but is now used in the sense of "come"（来) as a phonetic loan.

It is said that the Chinese characters have reached 60,000, with over 3000 character used frequently. Every character has its own form. However Chinese characters are by no means irregular. It is not difficult to find that they consist of three elements: strokes，semantic radicals or determinants，frames.

Chinese characters are written in virtual square boxes, that means, that every character has the same size, and every part of a character has to be adjusted to this box, for example:

日	月	水	火
土	人	好	营

With the invention of the brush, the modern style of calligraphy developed with up to 16 different stroke patterns and several rules how to write a character and its elements. Every character can be separated into a certain number of strokes. The stroke number is very important looking for a character in a dictionary. Most modern search indices are oriented in first place according to the radical, and then to the number of the strokes left over. These stroke patterns are reduced to eight different strokes in *Kaishu* calligraphy[2]. The character 永 contains all different stroke patterns. The eight strokes are:

一	horizontal (*heng* 橫 *le* 勒)	㇏	right-slant (*na*捺*zhe* 磔)
丨	vertical (*shu* 竪 *nu* 努)	亅	hook (*gou*钩 *yue* 趯)
㇔	dot (*dian* 點 *ce* 側)	丿	down-slant (*pie*撇 *lüe* 掠)
丿	left-slant (*duanpie*短撇 右上 *zhuo* 啄)		up-slant (*tiao*挑 *ce* 策)

The already mentioned Han dictionary *Explaining Simple and Analyzing Compound Characters* arranged its characters in 540 semantic radicals (*bushou*). These "radicals" or "determinants" are part of the character and describe a field of meaning: "抱' (embrace), "择' (choose), "採' (pluck) with the radical "手' (hand), "林' (forest), "枝' (branch), "桑' (mulberry) with the radical "木' (wood or tree), "思' (thinking), "忧' (sorrow), "情' (feeling, mind) with the radical "心' (heart) and so on. The 17th century dictionary *Kangxi Dictionary* has reduced the radicals to 214. Modern dictionaries still use this system or have partly added their own radicals.

A further system to classify characters is the so-called "selection according to frames" system. It classifies the characters according to their composition as 1) compact, like "木' (wood); 2) one part inside the other, like "国' (country); 3) one part above the other, like "家' (home); 4) left upper and right lower part, like "庆' (celebrate); 5) two parts side by side, like "打' (bite).

Modern dictionaries use also the phonetical system of Pinyin and Zhuyin transcription by which characters can be found very easily if one knows the pronunciation. If not, the pronunciation of many characters can be guessed because most Chinese characters are of the above mentioned *xingsheng* system, a combination of sound and radical. In addition, the stroke number is very important looking for a character in a dictionary. Most modern search indices are oriented in first place according to the radical, and then to the number of the strokes left over.

Sociolinguistics is "the study of language in relation to society". Culture elements will inevitably penetrate into the language encoding way. In its broad sense culture is the attributes of man. Culture in its narrow sense is also called "small c culture" or "anthropological culture". In this sense, culture can be defined as life-way of a population. Language is the carrier and container of culture. As a mirror of culture language is strongly influenced and shaped by culture. Language and culture is closely related, each influencing and shaping the other. Chinese characters embody Chinese culture and social customs.

For instance, *Shuowen jiezi* said, " 糸 (mi) is silk, like a bunch of silk.". "糸' in the Oracle-Bone Inscriptions was written as   or   . There are more than 100 Chinese characters with the radical "糸'. It is well known than the Chinese have been making silk for thousands of years. Silk is a special kind of cloth. It is strong but light and soft. People outside China did not learn how to make this expensive cloth until the A.D. 800s. The Silk Road was one of the most important trade routes in ancient China. It stretched across Asia from China to the Mediterranean Sea. There trader bought or sold spices, gold, ivory, jade and other expensive goods. This famous trade route is named after one of the most items made at that time – silk. Today, China is still the world's leading producer of silk. From this we can see what an important position of "silk" in Chinese people.

---

[2] Chinese painters and calligraphers developed six styles of writing (*shuti* 书体).The oldest type is the "Seal Script" (篆书 *zhuanshu*) used especially to be carved into stone or metal surfaces. The second type is the "Chancellery Script" (隶书 *lishu*) used to write on bamboo or wooden stripes from the Han Dynasty, said to be invented by Cheng Miao. The third type is the "Writing Script" (章书 *zhangshu*), a more cursive type. The fourth type is "Grass Script" (草书 *caoshu*), a very cursive writing style to write on paper or silk. The fifth type is the normal hand writing type "Running Script" (行书 *xingshu*), and finally, the book print type "Correct Script" (正书 *zhengshu*, also called "Pattern Script" (楷书 *kaishu*), said to be invented by the calligrapher Wang Xizhi. The most popular book printing type is called *Songshu* 宋书 or *Songti* 宋体

For example, "绪" originally means the head of silk and has by extension come to mean main threads of a complex affair. "级" means the grade of silk. "绝" means broken silk. "纸" (paper) indicates that as one of four great inventions, ancient Chinese people were producing paper with silk loss and plant fiber during the Western Han Dynasty.

The fact that there so many Chinese characters with the radical "马" speaks volumes for Chinese civilization affecting the codification of Chinese characters. The importance of the horse in the history and culture of China can be viewed, in part, through the artistic legacy of this great civilization. In sculpture, painting, and literature, horses were glorified and revered. Horses were believed to be related to mythological dragons, reflecting their sacred status within society. The horse had divine power and spirit and was represented as such by the artistic world throughout history, up to the present day. The great influence of the horse throughout the history of China cannot be underestimated. First, Chinese people created their country based on agricultural practice. Thus, Chinese civilization is an agricultural civilization. Horses are the main tool of labor. The character "驮", meaning "carry or bear on the back", is a good point in case. Second, the ancient expansion of the Chinese Empire was due in large part to the horse. The rapid mobility of horse allowed for quick communication between far away provinces. As vehicles, four-wheelers drawn by several horses were popular in the north in ancient times. The characters like "驿", meaning a post station, "驲", meaning a carriage for a post station, and "骈", meaning "walking shoulder to shoulder", are good illustrations. Likewise, the military role of horses aided in the conquest and submission of distant lands. During the unification of China under the Han Dynasty, bands of mounted nomadic warriors from the north threatened the country. In order to thwart their attacks, the Chinese sought to import stronger, faster steeds from Central Asia (as opposed to the Mongol ponies used by the invaders), eventually leading to the creation of the Silk Road. The characters such as "驰", "骋" and "驶", meaning houses galloping, show this cultural background.

IV. The Encoding Way of English Words

English belongs to alphabetic writing, a writing system made up of separate letters which represent sounds. Except some monosyllable words, most of the English words consist of a root or a stem and combining form, i.e. suffix or prefix. If we make a study of the sources of English words, we will benefit a lot. It will help Chinese students to remember a large number of English words. Although English is a very complex mixture, the theory of its codification is very like ideographs and mutual explanatory in Chinese *Liushu*. Therefore, if we understand the sense of English codification systematically, we will remember English words group by group and will have a deep understanding of fundamental meaning of words. Thus it is conducive for us to read and write good passages.

Modern English belongs to the Germanic branch of the Indo-European family of languages. It is actually a combining of two main language branches, one is Anglo-Saxon words in Germanic branch and the other is Romance or Romantic language, like French and Latin in the Italic branch. The former is the backbone while the latter is the flesh and blood. The elements that constitute English vocabulary include: French 35%, Anglo-Saxon 25%, Latin 16%, Greek 14%, others 10%.

This cosmopolitan character of English vocabulary reflects social and cultural developments through centuries. The political and social events that have in the course of English history so profoundly influenced the English people in their national life have generally had a recognizable effect on their language. The Roman Christianizing of Britain in 597 brought England into contact with Latin civilization and made significant additions to their vocabulary. For example, port from Latin portus and porta. The Scandinavian invasions resulted in a considerable mixture of the two peoples and their language. The Norman Conquest made English for two centuries the language mainly of the lower classes while the nobles and those associated with them used French on almost all occasions, especially governmental and administrative words, ecclesiastical words, and words about law, army, navy, fashion, meals, and social life. For example, the office and the titles of many offices are likewise French: chancellor, treasurer, chamberlain, marshal, governor, councilor, minister, viscount, warden, castellan, mayor, constable, coroner and even the humble crier. And when English once more regained supremacy as the language of all elements of the population, it was an English greatly changed in both form and vocabulary from what it had been in 1066.

In a similar way the Hundred Year's War, the rise of an important middle class, the Renaissance, the development of England as a maritime power, the expansion of the British Empire, and the growth of commerce and industry, of science and literature, have, each in their way, contributed to the development of the language (Baugh, 2001: 2).

As mentioned before, the codification of English words is like the ideographs in Chinese Liushu. I will set "conspiracy" as an example. "Con-" is a Latin prefix, meaning "together, with, wholly". The Latin word "spirare" means "to breathe; the suffix "cy" indicates the quality, condition and state. When individuals get together, they have the state of breathing mutually, so they are sure to discuss certain confidential. Therefore, the word can be interpreted as "a secret plan by two or more people to do something against the law". Another example is a group of words with the root "thes, thet". The Greek word "tithenai" means to place and to put. Thus, "thesis", with "sis" noun suffix, means "put opinion into…"; "antithesis" with "anti" prefix, means "put it on the opposite position"; "hypothesis" with "hypo-under" prefix, means "put it underneath"; "synthesis" with "syn-together" prefix, means "put … together"

V. Conclusion

English is the Western tiger and has become fashionable nowadays; Chinese is the Eastern dragon and has been flaming since ancient times. The twenty-six Latin letters of English condense the conciseness and clearness of mankind's symbols and accelerate to pass on modern civilization. The Chinese characters in virtual square boxes contain the connotation of wisdom and Chinese culture. They are two big families among mankind's thousands of languages. Each of them has her strong points and creates human beings' splendor.

Bibliography

Baugh, Albert C. & Cable, Thomas, A History of the English Language, Beijing: Foreign Language Teaching and Research Press, 2001.

Halliday, M.A.K., Language as Social Semiotic: The Social Interpretation of Language and Meaning, Beijing: Foreign Language Teaching and Research Press, 2001.

Hu Zhuanglin & Jiang Wangqi, ed. Linguistics: An Advanced Course Book, Beijing: Beijing University Press, 2002.

Hu Zhuanglin, et al, ed. Linguistics: A Course Book, Beijing: Beijing University Press, 1987.

Richards, Jack C., et al, Longman Dictionary of Language Teaching & Applied Linguistics, Beijing: Foreign Language Teaching and Research Press, 2002.

Wardhaugh, Ronald, An Introduction to Sociolinguistics, Beijing: Foreign Language Teaching and Research Press, 2000.

Wang Zhenya, ed. Language and Culture, Beijing: Higher Education Press, 1999.

Huang Bingshu, compiled, Etymology and Coinage of English, Chengdu: Southwest Publishing House, 1978.

Lu Zongda. A General Survey of Explaining Simple and Analyzing Compound Characters, Beijing: Beijing Publishing House, 1981.

Yang Wuming, compiled, Philology, Hunan: Hunan People's Publishing House, 1986.

Tan Jiajian, ed. An Outline of History of Chinese Culture, Beijing: Higher Education Publishing House, 1997.

http://www.polyu.edu.hkhgec\chinastudies\theme9.pdf