

= Written Chinese =

Written Chinese (Chinese : 中文 ; pinyin : zhōngwén) comprises Chinese characters (汉字 / 汉字 ; pinyin : Hànzì , literally " Han characters ") used to represent the Chinese language . Chinese characters do not constitute an alphabet or a compact syllabary . Rather , the writing system is roughly logosyllabic ; that is , a character generally represents one syllable of spoken Chinese and may be a word on its own or a part of a polysyllabic word . The characters themselves are often composed of parts that may represent physical objects , abstract notions , or pronunciation . Literacy requires the memorization of a great many characters : educated Chinese know about 4 ,000 . The large number of Chinese characters has in part led to the adoption of Western alphabets as an auxiliary means of representing Chinese .

Various current Chinese characters have been traced back to the late Shang Dynasty about 1200 ? 1050 BC , but the process of creating characters is thought to have begun some centuries earlier . After a period of variation and evolution , Chinese characters were standardized under the Qin Dynasty (221 ? 206 BC) . Over the millennia , these characters have evolved into well @-@ developed styles of Chinese calligraphy . Chinese speakers in disparate dialect groups are able to communicate through writing because standard written Chinese is based on a standard spoken language (" Mandarin ") and previously on Classical Chinese . Although most other varieties of Chinese are not written , there are traditions of written Cantonese , written Hokkien , and written Shanghainese , among others .

Some Chinese characters have been adopted as part of the writing systems of other East Asian languages , such as Japanese , Korean , and (formerly) Vietnamese .

= = Structure = =

Written Chinese is not based on an alphabet or a compact syllabary . Instead , Chinese characters are glyphs whose components may depict objects or represent abstract notions . Occasionally a character consists of only one component ; more commonly two or more components are combined to form more complex characters , using a variety of different principles . The best known exposition of Chinese character composition is the *Shuowen Jiezi* , compiled by Xu Shen around 120 AD . Since Xu Shen did not have access to Chinese characters in their earliest forms , his analysis cannot always be taken as authoritative . Nonetheless , no later work has supplanted the *Shuowen Jiezi* in terms of breadth , and it is still relevant to etymological research today .

= = = Derivation of characters = = =

According to the *Shuowen Jiezi* , Chinese characters are developed on six basic principles . (These principles , though popularized by the *Shuowen Jiezi* , were developed earlier ; the oldest known mention of them is in the *Rites of Zhou* , a text from about 150 BC .) The first two principles produce simple characters , known as 文 wén :

?? xiàngxíng : Pictographs , in which the character is a graphical depiction of the object it denotes . Examples : 人 rén " person " , 日 rì " sun " , 木 mù " tree / wood " .

?? zhǐshì : Indicatives , or ideographs , in which the character represents an abstract notion . Examples : 上 shàng " up " , 下 xià " down " , 三 sān " three " .

The remaining four principles produce complex characters historically called 字 zì (although this term is now generally used to refer to all characters , whether simple or complex) . Of these four , two construct characters from simpler parts :

?? / ?? huìyì : Logical aggregates , in which two or more parts are used for their meaning . This yields a composite meaning , which is then applied to the new character . Example : 日 / 木 dìng " east " , which represents a sun rising in the trees .

?? / ?? xíngshēng : Phonetic complexes , in which one part ? often called the radical ? indicates the general semantic category of the character (such as water @-@ related or eye @-@ related) , and the other part is another character , used for its phonetic value . Example : 清 qīng " clear / fair (

weather) " , which is composed of 日 " sun " , and 青 " blue / green " , which is used for its pronunciation .

In contrast to the popular conception of Chinese as a primarily pictographic or ideographic language , the vast majority of Chinese characters (about 95 percent of the characters in the Shuowen Jiezi) are constructed as either logical aggregates or , more often , phonetic complexes . In fact , some phonetic complexes were originally simple pictographs that were later augmented by the addition of a semantic root . An example is 燭 " candle " (now archaic , meaning " lampwick ") , which was originally a pictograph 𩇛 , a character that is now pronounced zhù and means " host " . The character 火 " fire " was added to indicate that the meaning is fire-related .

The last two principles do not produce new written forms ; instead , they transfer new meanings to existing forms :

转 / 转 zhuǎnzhuǎn : Transference , in which a character , often with a simple , concrete meaning takes on an extended , more abstract meaning . Example : 网 / 网 wǎng " net " , which was originally a pictograph depicting a fishing net . Over time , it has taken on an extended meaning , covering any kind of lattice ; for instance , it can be used to refer to a computer network .

借 / 借 jiè : Borrowing , in which a character is used , either intentionally or accidentally , for some entirely different purpose . Example : 哥 " older brother " , which is written with a character originally meaning " song / sing " , now written 哥 . Once , there was no character for " older brother " , so an otherwise unrelated character with the right pronunciation was borrowed for that meaning .

Chinese characters are written to fit into a square , even when composed of two simpler forms written side-by-side or top-to-bottom . In such cases , each form is compressed to fit the entire character into a square .

== Strokes ==

Character components can be further subdivided into strokes . The strokes of Chinese characters fall into eight main categories : horizontal (横) , vertical (竖) , left-to-right falling (撇) , right-to-left falling (捺) , rising (挑) , dot (点) , hook (钩) , and turning (折 , etc .) .

There are eight basic rules of stroke order in writing a Chinese character :

Horizontal strokes are written before vertical ones .

Left-to-right falling strokes are written before right-to-left falling ones .

Characters are written from top to bottom .

Characters are written from left to right .

If a character is framed from above , the frame is written first .

If a character is framed from below , the frame is written last .

Frames are closed last .

In a symmetrical character , the middle is drawn first , then the sides .

These rules do not strictly apply to every situation and are occasionally violated .

== Layout ==

Chinese characters conform to a roughly square frame and are not usually linked to one another , so they can be written in any direction in a square grid . Traditionally , Chinese was written in vertical columns from top to bottom ; the first column being on the right side of the page , and the last column on the left . Text written in Classical Chinese also uses little or no punctuation . In such cases , sentence and phrase breaks are determined by context and rhythm . This layout is still sometimes used in letters and formal documents .

In modern times , the familiar Western layout of horizontal rows from left to right , read from the top of the page to the bottom , has become more popular , especially in the People's Republic of China (mainland China) , where the government mandated left-to-right writing in 1955 . The government of the Republic of China (Taiwan) followed suit in 2004 for official documents . The use of punctuation has also become more common , whether the text is written in columns or rows .

The punctuation marks are clearly influenced by their Western counterparts , although some marks are particular to Asian languages : for example , the double and single quotation marks (“ ” and ‘ ’) ; the hollow period (。) , which is otherwise used just like an ordinary period ; and a special kind of comma called an enumeration comma (、) , which is used to separate items in a list , as opposed to clauses in a sentence .

Signs are a particularly challenging aspect of written Chinese layout , since they can be written either left to right or right to left (the latter can be thought of as the traditional layout with each “ column ” being one character high) , as well as from top to bottom . It is not uncommon to encounter all three orientations on signs on neighboring stores .

= = Evolution = =

Chinese is one of the oldest continually used writing systems still in use . The earliest generally accepted examples of Chinese writing date back to the reign of the Shang Dynasty king Wu Ding (1250 ? 1192 BC) . These were divinatory inscriptions on oracle bones , primarily ox scapulae and turtle shells . Characters were carved on the bones in order to frame a question ; the bones were then heated over a fire and the resulting cracks were interpreted to determine the answer . Such characters are called 甲骨文 “ shell @-@ bone script ” or oracle bone script .

In 2003 , some 11 isolated symbols carved on tortoise shells were found at Jiahu , an archaeological site in the Henan province of China , some bearing a striking resemblance to certain modern characters , such as 目 “ eye ” . Since the Jiahu site dates from about 6600 BC , it predates the earliest confirmed Chinese writing by more than 5 @, @ 000 years . Dr Garman Harbottle , of the Brookhaven National Laboratory in New York , US , who headed a team of archaeologists at the University of Science and Technology of China , in Anhui province , has suggested that these symbols were precursors of Chinese writing , but Professor David Keightley , of the University of California , Berkeley , US whose field of expertise is the origins of Chinese civilization in the Neolithic and early Bronze Ages , employing archaeological and inscriptional evidence , suggests that the time gap is too great for a connection .

From the late Shang Dynasty , Chinese writing evolved into the form found in cast inscriptions on Chinese ritual bronzes made during the Western Zhou Dynasty (c 1066 ? 770 BC) and the Spring and Autumn Period (770 ? 476 BC) , a kind of writing called 金文 “ metal script ” . Jinwen characters are less angular and angularized than the oracle bone script . Later , in the Warring States period (475 ? 221 BC) , the script became still more regular , and settled on a form , called 六国书 / 六国文 “ script of the six states ” , that Xu Shen used as source material in the Shuowen Jiezi . These characters were later embellished and stylized to yield the seal script , which represents the oldest form of Chinese characters still in modern use . They are used principally for signature seals , or chops , which are often used in place of a signature for Chinese documents and artwork . Li Si promulgated the seal script as the standard throughout the empire during the Qin dynasty , then newly unified .

Seal script in turn evolved into the other surviving writing styles ; the first writing style to follow was the clerical script . The development of such a style can be attributed to those of the Qin Dynasty who were seeking to create a convenient form of written characters for daily usage . In general , clerical script characters are “ flat ” in appearance , being wider than the seal script , which tends to be taller than it is wide . Compared with the seal script , clerical script characters are strikingly rectilinear . In running script , a semi @-@ cursive form , the character elements begin to run into each other , although the characters themselves generally remain separate . Running script eventually evolved into grass script , a fully cursive form , in which the characters are often entirely unrecognizable by their canonical forms . Grass script gives the impression of anarchy in its appearance , and there is indeed considerable freedom on the part of the calligrapher , but this freedom is circumscribed by conventional “ abbreviations ” in the forms of the characters . Regular script , a non @-@ cursive form , is the most widely recognized script . In regular script , each stroke of each character is clearly drawn out from the others . Even though both the running and grass scripts appear to be derived as semi @-@ cursive and cursive variants of regular script , it is in fact

the regular script that was the last to develop .

Regular script is considered the archetype for Chinese writing , and forms the basis for most printed forms . In addition , regular script imposes a stroke order , which must be followed in order for the characters to be written correctly . (Strictly speaking , this stroke order applies to the clerical , running , and grass scripts as well , but especially in the running and grass scripts , this order is occasionally deviated from .) Thus , for instance , the character ? mù " wood " must be written starting with the horizontal stroke , drawn from left to right ; next , the vertical stroke , from top to bottom ; next , the left diagonal stroke , from top to bottom ; and lastly the right diagonal stroke , from top to bottom .

= = = Simplified and traditional Chinese = = =

In the 20th century , written Chinese divided into two canonical forms , called simplified Chinese and traditional Chinese . Simplified Chinese was developed in mainland China in order to make the characters faster to write (especially as some characters had as many as a few dozen strokes) and easier to memorize . The People 's Republic of China claims that both goals have been achieved , but some external observers disagree . Little systematic study has been conducted on how simplified Chinese has affected the way Chinese people become literate ; the only studies conducted before it was standardized in mainland China seem to have been statistical ones regarding how many strokes were saved on average in samples of running text .

The simplified forms have also been criticized for being inconsistent . For instance , traditional ? ràng " allow " is simplified to ? , in which the phonetic on the right side is reduced from 17 strokes to just three . (The speech radical on the left has also been simplified .) However , the same phonetic is used in its full form , even in simplified Chinese , in such characters as ? r?ng " soil " and ? nàng " snuffle " ; these forms remained uncontracted because they were relatively uncommon and would therefore represent a negligible stroke reduction . On the other hand , some simplified forms are simply long standing calligraphic abbreviations , as for example ? wàn " ten thousand " , for which the traditional Chinese form is ? .

Simplified Chinese is standard in the People 's Republic of China , Singapore , and Malaysia . Traditional Chinese is retained in Hong Kong , Taiwan , Macau and overseas Chinese communities (except Singapore and Malaysia) . Throughout this article , Chinese text is given in both simplified and traditional forms when they differ , with the traditional forms being given first .

= = Function = =

At the inception of written Chinese , spoken Chinese was monosyllabic ; that is , Chinese words expressing independent concepts (objects , actions , relations , etc .) were usually one syllable . Each written character corresponded to one monosyllabic word . The spoken language has since become polysyllabic , but because modern polysyllabic words are usually composed of older monosyllabic words , Chinese characters have always been used to represent individual Chinese syllables .

For over two thousand years , the prevailing written standard was a vocabulary and syntax rooted in Chinese as spoken around the time of Confucius (about 500 BC) , called Classical Chinese , or ??? wényánwén . Over the centuries , Classical Chinese gradually acquired some of its grammar and character senses from the various dialects . This accretion was generally slow and minor ; however , by the 20th century , Classical Chinese was distinctly different from any contemporary dialect , and had to be learned separately . Once learned , it was a common medium for communication between people speaking different dialects , many of which were mutually unintelligible by the end of the first millennium AD . A Mandarin speaker might say y? , a Cantonese y?t , and a Hokkien chit , but all three will understand the character ? " one " .

Chinese dialects vary by not only pronunciation , but also , to a lesser extent , vocabulary and grammar . Modern written Chinese , which replaced Classical Chinese as the written standard as an indirect result of the May Fourth Movement of 1919 , is not technically bound to any single dialect ;

however , it most nearly represents the vocabulary and syntax of Mandarin , by far the most widespread Chinese dialect in terms of both geographical area and number of speakers . This version of written Chinese is called Vernacular Chinese , or 白話 / 白话 báihuà (literally , " plain speech ") . Despite its ties to the dominant Mandarin dialect , Vernacular Chinese also permits some communication between people of different dialects , limited by the fact that Vernacular Chinese expressions are often ungrammatical or unidiomatic in non @-@ Mandarin dialects . This role may not differ substantially from the role of other *linguae francae* , such as Latin : For those trained in written Chinese , it serves as a common medium ; for those untrained in it , the graphic nature of the characters is in general no aid to common understanding (characters such as " one " notwithstanding) . In this regard , Chinese characters may be considered a large and inefficient phonetic script . However , Ghil 'ad Zuckermann 's exploration of phono @-@ semantic matching in Standard Chinese concludes that the Chinese writing system is multifunctional , conveying both semantic and phonetic content .

The variation in vocabulary among dialects has also led to the informal use of " dialectal characters " , as well as standard characters that are nevertheless considered archaic by today 's standards . Cantonese is unique among non @-@ Mandarin regional languages in having a written colloquial standard , used in Hong Kong and overseas , with a large number of unofficial characters for words particular to this dialect . Written colloquial Cantonese has become quite popular in online chat rooms and instant messaging , although for formal written communications Cantonese speakers still normally use Vernacular Chinese . To a lesser degree Hokkien is used in a similar way in Taiwan and elsewhere , although it lacks the level of standardization seen in Cantonese . However , the Ministry of Education of the Republic of China is currently releasing a standard character set for Hokkien , which is to be taught in schools and promoted amongst the general population .

= = = Other languages = = =

Chinese characters were first introduced into Japanese sometime in the first half of the first millennium AD , probably from Chinese products imported into Japan through Korea . At the time , Japanese had no native written system , and Chinese characters were used for the most part to represent Japanese words with the corresponding meanings , rather than similar pronunciations . A notable exception to this rule was the system of *man'yōgana* , which used a small set of Chinese characters to help indicate pronunciation . The *man'yōgana* later developed into the phonetic syllabaries , hiragana and katakana .

Chinese characters are called *hànzì* in Mandarin , after the Han Dynasty of China ; in Japanese , this was pronounced *kanji* . In modern written Japanese , *kanji* are used for most nouns , verb stems , and adjective stems , while hiragana are used for grammatical elements and miscellaneous words that have no common *kanji* rendition ; katakana are used for transliteration of loanwords from other languages , the names of plants , animals and certain scientific or technical words , onomatopoeia and emphasis . The *Jōyō kanji* , a list of *kanji* for common use standardized by the Japanese government , contains 2 @, @ 136 characters ? about half the number of characters commanded by literate Chinese .

The role of Chinese characters in Korean and Vietnamese is much more limited . At one time , many Chinese characters (called *hanja*) were introduced into Korean for their meaning , just as in Japanese . Today , Korean is written almost exclusively using the Hangul alphabet with a small number of Chinese characters . Each square block character contains Hangul symbols , or letters , that together represent a syllable . Similarly , the use of Chinese and Chinese @-@ styled characters in the Vietnamese *chữ nôm* script has been almost entirely superseded by the Latin @-@ based Vietnamese alphabet . Chinese characters are still actively used in South Korea today , mostly for signs , newspapers , books , and government documents .

Chinese characters are also used within China to write non @-@ Han languages . The largest non @-@ Han group in China , the Zhuang , have for over 1300 years used Chinese characters . Despite both the introduction of an official alphabetic script in 1957 and lack of a corresponding official set of Chinese characters , more Zhuang people can read the Zhuang logograms than the

alphabetic script .

= = Media = =

Over the history of written Chinese , a variety of media have been used for writing . They include :

Bamboo and wooden slips , from at least the thirteenth century BC

Paper , invented no later than the second century BC

Silk , since at least the Han dynasty

Stone , metal , wood , bamboo , plastic and ivory on seals .

Since at least the Han dynasty , such media have been used to create hanging scrolls and handscrolls .

= = Literacy = =

Because the majority of modern Chinese words contain more than one character , there are at least two measuring sticks for Chinese literacy : the number of characters known , and the number of words known . John DeFrancis , in the introduction to his Advanced Chinese Reader , estimates that a typical Chinese college graduate recognizes 4 @, @ 000 to 5 @, @ 000 characters , and 40 @, @ 000 to 60 @, @ 000 words . Jerry Norman , in Chinese , places the number of characters somewhat lower , at 3 @, @ 000 to 4 @, @ 000 . These counts are complicated by the tangled development of Chinese characters . In many cases , a single character came to be written in multiple ways . This development was restrained to an extent by the standardization of the seal script during the Qin dynasty , but soon started again . Although the Shuowen Jiezi lists 10 @, @ 516 characters ? 9 @, @ 353 of them unique (some of which may already have been out of use by the time it was compiled) plus 1 @, @ 163 graphic variants ? the Jiyun of the Northern Song Dynasty , compiled less than a thousand years later in 1039 , contains 53 @, @ 525 characters , most of them graphic variants .

= = = Dictionaries = = =

Written Chinese is not based on an alphabet or syllabary , so Chinese dictionaries , as well as dictionaries that define Chinese characters in other languages , cannot easily be alphabetized or otherwise lexically ordered , as English dictionaries are . The need to arrange Chinese characters in order to permit efficient lookup has given rise to a considerable variety of ways to organize and index the characters .

A traditional mechanism is the method of radicals , which uses a set of character roots . These roots , or radicals , generally but imperfectly align with the parts used to compose characters by means of logical aggregation and phonetic complex . A canonical set of 214 radicals was developed during the rule of the Kangxi Emperor (around the year 1700) ; these are sometimes called the Kangxi radicals . The radicals are ordered first by stroke count (that is , the number of strokes required to write the radical) ; within a given stroke count , the radicals also have a prescribed order .

Every Chinese character falls (sometimes arbitrarily or incorrectly) under the heading of exactly one of these 214 radicals . In many cases , the radicals are themselves characters , which naturally come first under their own heading . All other characters under a given radical are ordered by the stroke count of the character . Usually , however , there are still many characters with a given stroke count under a given radical . At this point , characters are not given in any recognizable order ; the user must locate the character by going through all the characters with that stroke count , typically listed for convenience at the top of the page on which they occur .

Because the method of radicals is applied only to the written character , one need not know how to pronounce a character before looking it up ; the entry , once located , usually gives the pronunciation . However , it is not always easy to identify which of the various roots of a character is the proper radical . Accordingly , dictionaries often include a list of hard to locate characters ,

indexed by total stroke count , near the beginning of the dictionary . Some dictionaries include almost one @-@ seventh of all characters in this list .

Other methods of organization exist , often in an attempt to address the shortcomings of the radical method , but are less common . For instance , it is common for a dictionary ordered principally by the Kangxi radicals to have an auxiliary index by pronunciation , expressed typically in either hanyu pinyin or zhuyin fuhao . This index points to the page in the main dictionary where the desired character can be found . Other methods use only the structure of the characters , such as the four @-@ corner method , in which characters are indexed according to the kinds of strokes located nearest the four corners (hence the name of the method) , or the Cangjie method , in which characters are broken down into a set of 24 basic components . Neither the four @-@ corner method nor the Cangjie method requires the user to identify the proper radical , although many strokes or components have alternate forms , which must be memorized in order to use these methods effectively .

The availability of computerized Chinese dictionaries now makes it possible to look characters up by any of the indexing schemes described , thereby shortening the search process .

= = = Transliteration and romanization = = =

Chinese characters do not reliably indicate their pronunciation , even for one dialect . It is therefore useful to be able to transliterate a dialect of Chinese into the Latin alphabet or the Perso @-@ Arabic script Xiao 'erjing for those who cannot read Chinese characters . However , transliteration was not always considered merely a way to record the sounds of any particular dialect of Chinese ; it was once also considered a potential replacement for the Chinese characters . This was first prominently proposed during the May Fourth Movement , and it gained further support with the victory of the Communists in 1949 . Immediately afterward , the mainland government began two parallel programs relating to written Chinese . One was the development of an alphabetic script for Mandarin , which was spoken by about two @-@ thirds of the Chinese population ; the other was the simplification of the traditional characters ? a process that would eventually lead to simplified Chinese . The latter was not viewed as an impediment to the former ; rather , it would ease the transition toward the exclusive use of an alphabetic (or at least phonetic) script .

By 1958 , however , priority was given officially to simplified Chinese ; a phonetic script , hanyu pinyin , had been developed , but its deployment to the exclusion of simplified characters was pushed off to some distant future date . The association between pinyin and Mandarin , as opposed to other dialects , may have contributed to this deferment . It seems unlikely that pinyin will supplant Chinese characters anytime soon as the sole means of representing Chinese .

Pinyin uses the Latin alphabet , along with a few diacritical marks , to represent the sounds of Mandarin in standard pronunciation . For the most part , pinyin uses vowel and consonant letters as they are used in Romance languages (and also in IPA) . However , although ' b ' and ' p ' , for instance , represent the voice / unvoiced distinction in some languages , such as French , they represent the unaspirated / aspirated distinction in Mandarin ; Mandarin has few voiced consonants . Also , the pinyin spellings for a few consonant sounds are markedly different from their spellings in other languages that use the Latin alphabet ; for instance , pinyin ' q ' and ' x ' sound similar to English ' ch ' and ' sh ' , respectively . Pinyin is not the sole transliteration scheme for Mandarin ? there are also , for instance , the zhuyin fuhao , Wade @-@ Giles , and Gwoyeu Romatzyh systems ? but it is dominant in the Chinese @-@ speaking world . All transliterations in this article use the pinyin system .