

Japanese HWR

Steven B. Poggel
steven.poggel@gmail.com

January 15, 2010

Contents

1	Japanese Script	5
1.1	A Short History of the Japanese Script	5
1.1.1	Historical Development	5
1.1.1.1	History of the Kanji	5
1.1.1.2	Typology of the Kanji	6
1.2	The Modern Japanese Writing System	8
1.2.1	Kana かな	9
1.2.1.1	Hiragana ひらがな	9
1.2.1.2	Katakana カタカナ	9
1.2.2	Composition of the Kanji 漢字	9
1.2.2.1	Graphemic Elements	9
1.2.2.2	Radicals	9
1.2.2.3	Readings	9
1.2.3	Structure of the Japanese Writing System	9
1.2.4	Romaji ロマジ	9
1.2.5	Machine Writing of Japanese	9
1.3	Difficulties of Writing Japanese for Learners	9

Chapter 1

Japanese Script

The Japanese writing system has a long history. It goes back to around 800 A.D. The Japanese script is in fact a writing system, as Japanese is denoted in a combination of three different scripts: *Hiragana*, *Katakana* and *Kanji*. Kanji is a conceptual script, where each character bears the meaning of one or more semantic concepts and represents morphemes. Hiragana and Katakana are both syllabic scripts, and the individual characters do not bear reference to concepts or even words, but merely to phonological units, usually two phonemes.

In this chapter, the development of the script will be reviewed in section 1.1. In section 1.2 the current Japanese writing system will be exemplified, with a focus on the Kanji in section 1.2.2. Hiragana and Katakana will be reviewed in section 1.2.1, which centers around the Kana scripts. Machine processing of the different Japanese scripts and the difficulties that go along will be demonstrated in section 1.2.5. The difficulties of learning to use the Japanese script will be illustrated in section 1.3.

1.1 A Short History of the Japanese Script

The historical development of the Japanese script is tightly connected to the history of the Kanji characters. Kanji, in Japanese 漢字 (Jap. pron. カンジ / kanji; Eng. lit. *Han characters*) refers to the 'characters of the Han', meaning the Han Dynasty (206 B.C.-220 A.D.; simplified Chinese: 汉朝; traditional Chinese: 漢朝) (Foljanty 1984). In Mandarin the same characters are referred to as *Hànzì* (simplified Chinese: 汉字; trad. Chinese: 漢字). Note, that the first character 漢 (Chin. 'han', Jap. 'kan', Eng. 'Han') of both the words *Han dynasty* and *Kanji* is identical in Japanese and traditional Chinese, even though it has a different pronunciation in the Chinese and Japanese language. In traditional Chinese the character with the same meaning (汉) has a different shape. This apparent oddity will be explained in greater detail in section 1.1.1.

1.1.1 Historical Development

1.1.1.1 History of the Kanji

The Kanji script as developed and coined by the Han is in principle still valid today. It is used alone or in combination with phonetic spelling in China, Japan, Taiwan, Hongkong. In Vietnam it was used before it was replaced with the Vietnamese alphabet (Viet.: 'quốc ngữ', Eng. lit. 'national language', Eng. 'national script'), a script based on the Latin alphabet. In South Korea the Han characters were in use until they were replaced with Hangul (Kor. with Han characters 韓國語; Eng. 'Korean') (Foljanty 1984).

The Kanji characters were brought to Japan by Koreans living in Japan around 300-400 A.D. Since the Kanji were used by the Koreans to write Hangul they also used it to write Japanese. There was no other Japanese script before that time. Reports about an original Japanese script called *Jindai Moji* (Jap. 神代文字; Eng. 'scripts of the age of the gods') could not be proven. They are now assumed to be a political and speculative invention by Japanese Nationalists in the early 19th. century (Foljanty 1984). According to (Lange 1922) the *Kogo Shūi* (Jap. 古語拾遺; a historical record of the Inbe clan), which was written around 800 B.C. denies the presence of a Japanese native script before the introduction of the Han characters.

However, the questions seems irrelevant in the sence, that no longer text or document has been found, written in that script.

In the Christian year 712 an ancestral act of writing was performed at Japanese emperor Temmu's court. Hieda no Are, a member of the guild of the *kataribe* or reciters, basically a Japanese Griot, dictates the *Kojiki* (Jap. 古事記; Eng. 'Record of Ancient Matters') to Ō no Yasumaro. Ō no Yasumaro wrote the *Kojiki*, which is not the first written document found in Japan, however it is Japan's oldest attempt to write down spoken Japanese (Grassmuck 1997; Chamberlain 1982).

At the time the Han characters were first used to write Japanese, they were already a developed script, more than 1,000 years old, as they stabilised to their modern form within the Han period¹ (Grassmuck 1997).

The first chinese characters were found on oracle bones from the Shang Dynasty (Chinese 商朝), which ruled over China some 500 to 600 years within the time period between 1600 B.C. and 1046 B.C. (Grassmuck 1997; Guo et al. 2000).

According to the *Kojiki*, a scholar called Wani (Jap. 王仁) from Korea brought two foundational Chinese books to Japan, the *Lunyu* (Simplified Chin. 论语; trad. Chinese: 論語; Eng. 'Analects'), also known as *The Analects of Confucius* and the *Qianziwen* (Chin./Jap. 千字文; Jap. pron. センジブン/senjibun; Eng. 'The Thousand Character Classic'), which is a Chinese poem used as a primer for teaching Chinese characters to children. It contains exactly one thousand unique characters (Grassmuck 1997). The Chinese language comprehends more than 40,000 Hànzì characters lexicographically. Only around 25% of those including about 250 *Kokuji* (Jap. 国字; Eng. 'national characters') are in Japanese dictionaries. Only around 2,000-3,000 of those are part of the common characters (Foljanty 1984).

The Japanese Ministry of Education issued a list of 1,850 standard Kanji in 1946 under the name of *Tōyō kanjihyō* (Jap. 当用漢字表; Eng. 'list of Kanji for general use'). The list of Tōyō Kanji was slightly revised and extended in 1981 and comprised 1,945 Kanji as the Jōyō Kanji (Jap. 常用漢字; Eng. 'often used Kanji') (Foljanty 1984). As of 2010 a revised list of 2,131 characters is in official use (Noguchi 2009).

In China there had been a spelling reform in the 1950s, affecting many of the general use characters, resulting in simplified Chinese. In Japan, the Ministry of Education issued it's own reform when the Tōyō kanji list was introduced. However, the Japanese reform affected a smaller set of characters of only a few hundred and resulted in Shinjitai (Jap. shinjitai: 新字体; Jap. kyūjitai: 新字體; Jap. pron. シンジタイ/shinjitai; Eng. 'new character form'), which replaced the Kyūjitai (Jap. shinjitai: 旧字体; Jap. kyūjitai: 舊字體; Jap. pron. キュウジタイ/kyūjitai; Eng. lit. 'old character forms'). This explains how some characters are still identical in traditional Chinese and Japanese, because they were not affected by any spelling reform, like the aforementioned 漢 (Jap. pron. カン/kan; Chin. pron. 'hàn'), while other characters are different, like the simplified Chinese 'hàn': 汉. Henceforth, and throughout this document, all Japanese characters are in the new character form shinjitai.

1.1.1.2 Typology of the Kanji

In order to study the Kanji and their composition, it is useful to know how they were first indendet and built. Integrated as the integral part into the Japanese writing system, despite the reform and the choice different subsets of what is considered the standard character set, the characters are still mainly composed the way as intendet by the scholars of the Han period.

From the religious writings on the oracle bones mentioned in section 1.1.1 a secular script emerged. In parallel, the process of graphical abstraction advanced and finished around 100-200 A.D., leaving aside the modern reforms of the 20th century. The invention of the paint-brush around 100 B.C. improved and simplified writing, also the writing surfaces in their order of appearance, bone, stone, metall, wood and then paper, brought further simplification and spreading of writing. Paper and paint-brush offered the possibilty to write without hindrance and technical coincidences, therefore it was possible to standartise the characters and improve them from artistic and aesthetic viewpoints.

The Kanji can be classified according to their building principle:

¹Also see timeline in section ??.

xxx put some
real kanji
pictograms
here, after
Kano1990

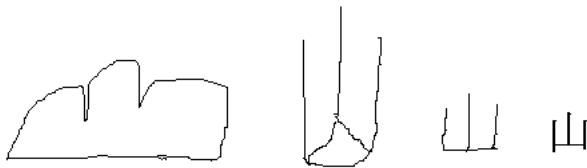


Figure 1.1: Kanji pictograms

xxx put some
real kanji
~~pictograms~~ ideograms
here, after
Kano1990

日 + 月 = 明
sun + moon = bright
both the sun and the moon are 'bright'

人 + 木 = 休
man + tree = rest
a man is 'resting' beside a tree

Figure 1.2: Kanji ideograms

Class radical	Sound radical /ki/	meaning
土 ('earth')	奇	埼 'spit, promotory, cape'
山 ('mountain')	奇	崎 'promontory, cape, spit'
石 ('stone')	奇	碕 'cape, promontory, spit'
王 ('jade')	奇	琦 'gem, precious stone'
糸 ('thread')	奇	綺 'figured cloth, beautiful'
馬 ('horse')	奇	騎 'riding on horses'
宀 ('roof')	奇	寄 'to gather'
金 ('metal')	奇	錡 'cauldron, chisel' (Chinese only)

Table 1.1: Kanji phonograms

1. **Pictograms** are graphically simplified images of real artefacts. The examples in Fig 1.1 after Kano et al. (1990) show the graphical reduction process. Pictograms are only a small minority among the Kanji, their number ranges around 120. Another 100 pictograms appear as a part of more complex characters (Foljanty 1984).
2. **Ideograms** are combinations of two or more pictographical characters. They often bear a more abstract meaning than a simple pictogram. The abstract meaning of the complex character is meant to be associated with the content of the individual parts. The number of ideograms is fairly small, too. Abstract terms like 'top' (Jap. 上, pron. うえ/ue), 'bottom' (Jap. 下, pron. した/shita), 'left' (Jap. 左, pron. ひだり/hidari), 'right' (Jap. 右, pron. みぎ/migi) and numbers like 'one' (Jap. 一, pron. いち/ichi), 'two' (Jap. 二, pron. に/ni), 'three' (Jap. 三, pron. さん/san), 'four' (Jap. 四, pron. し/shi), 'five' (Jap. 五, pron. ご/go) and so forth can be regarded as parts of the ideograms (Foljanty 1984).
3. **Phonograms** are combinations of two Kanji characters. One of those refers to a concept class (for *class characters* or *radicals*, see section 1.2.2.2), while the other character exclusively bears a phonetic value. The content of the second part of a phonogram is not relevant and can be ignored.

In table 1.1 the character 奇 (Jap. pron. き/ki, Eng. 'strange') is used for the purpose of pronunciation only (/ki/), while the radical defines an object class. Object classes can be categories like 'human and human actions', 'metal', 'horse', 'roof / under a roof' etc. The semantic identity within the *Morphogram* is assembled with two reference figures. The pronunciation part is identical for all characters, it serves as a selection criterion within a semantic class selected by the class radical (Foljanty 1984).

As a character type, phonograms are predominant among the Hànzì. Therefore, the phonogram concept, including the radical concept, was transferred to all Chinese characters. Pictograms that are class radicals themselves, are interpreted as characters with an empty sound radical. As the phonograms are historically the last development step of the Han characters, they constitute a different quality in the Chinese script. Phonograms mark the transition between a non-linguistic pictographic script that does not represent linguistic units, but rather images of objects, to a linguistic script. In principle, there is no difference to an alphabetical or syllabic script. Except, morphemes are represented instead of phonemes or syllables. However, one character often denotes more than one morpheme (Foljanty 1984).

In Japanese, basically the same relation between the Kanji characters and morphemes can be observed. However, the correspondence between the morphemes and the syllables (and thus the characters and syllables) is often missing. Since Chinese has a monosyllabic morpheme structure, a one-to-one correspondence between morpheme, character and syllable can be observed. Congruence of character, morpheme and syllable in Japanese can only be found for Chinese borrowings, but not all of them. The original Japanese vocabulary has multisyllabic morphemes, therefore some impreciseness arises in the graphical reproduction of the morpheme structure (Foljanty 1984).

1.2 The Modern Japanese Writing System

xxx: see (Foljanty 1984) 3.1 xxx: see (Lange 1922) p.64 xxx: see (Tsujimura 2007) for morphology stuff
xxx: see (Grassmuck 1997)

xxx: aufbau des schriftsystems generell xxx: Gemischtschreibung xxx: Kurze erwachnung der morphologie. Hiragana an verben zur konjugation. zusammenhang verben / nomen in kanji, xxx: uppercase / lowercase nicht vorhanden. etc. xxx: see <http://japanese.about.com/library/weekly/aa070101a.htm>
: radicals xxx: JB's wwwdict: <http://www.csse.monash.edu.au/~jwb/cgi-bin/wwwjdic.cgi?1R>

The Japanese writing system has a complex structure. The three scripts *Hiragana* (sec 1.2.1.1), *Katakana* (sec 1.2.1.2) and *Kanji* (sec 1.2.2), are combined to a writing system. Each script has its task within the system:

- Hiragana are used to write grammatical morphemes
- Katakana are used to transcribe foreign words or borrowings and nonstandard areas