

Japanese HWR

Steven B. Poggel
steven.poggel@gmail.com

January 18, 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	10
1.2.1.2	Katakana カタカナ	11
1.2.2	Composition of the Kanji 漢字	11
1.2.2.1	Graphemic Elements	11
1.2.2.2	Radicals	11
1.2.2.3	Readings	11
1.2.3	Structure of the Japanese Writing System	11
1.2.4	Romaji ロマジ	11
1.2.5	Machine Writing of Japanese	11
1.3	Difficulties of Writing Japanese for Learners	12

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

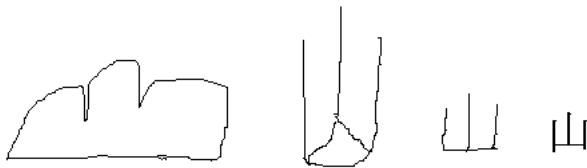


Figure 1.1: Kanji pictograms

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

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

- Kanji are used to write lexical morphemes, i.e. content-bearing morphemes.
- Katakana are used to transcribe foreign words, borrowings and nonstandard areas.
- Hiragana are used to write grammatical morphemes and anything else that is not written in one of the other two scripts, e.g. the spoken syllables of a word that should be written with a Kanji character unknown to the writer of that word.

The actual writing system is mainly based on Kanji and Hiragana, catenated to Kanji-Kana blended writing.

- (1) a. マリア：山田さん、「火の鳥」というアニメをもう見ました
 Maria: Yamada-San, [Firebird] say anime OBJ-PARTICLE already seen
 か。
 QUESTION-PARTICLE
 'Maria: Ms Yamada, say, have you seen the Firebird cartoon yet?'
 Taken from (Katsuki-Pestemer 2006a)
- b. マリア：山田さん、「火の鳥」というアニメをもう見ましたか。

In example (1a), the blending of the different scripts can be seen:

Both the foreign name マリア (Eng. 'Maria') and the borrowing アニメ (*anime*, Jap. short for Eng. 'animation') are written in Katakana. Kanji are used for:

- The Japanese name 山田 (*Yamada*).
- The nouns 火 (Eng. 'fire') and 鳥 (Eng. 'bird').
- The verb stem 見 (Eng. 'see').

The rest is written in Hiragana:

- The politeness ending さん (*san*; Eng. equiv. 'Mr/Ms/Mrs') for addressing a person with their name.
- The genitive particle の (*no*) between 火 (Eng. 'fire') and 鳥 (Eng. 'bird'), to yield 火の鳥 (Eng. 'firebird').
- The interjection とうい (Eng. 'say').
- The object particle を (*wo*).
- The adverb もう (Eng. 'already').
- The past tense conjugation and politeness ending of the verb ました (*mashita*).
- The question particle か (*ka*).

Three different scripts are used next to each other in one sentence, indistinguishable for the untrained eye. Example (1b) shows the sentence as it is printed in (Katsuki-Pestemer 2006a). Without prior knowledge of the different Japanese scripts it is hard to even distinguish the individual word tokens, as blanks are usually absent in Japanese writing. Other than actually knowing the words, which is not the usual case for a beginner of learning Japanese, often the change of script is the only way to recognise a new token. However, Kanji and Hiragana are often used within the same word, too.

Despite those complexities, other features of the Japanese writing system are simpler than in latin-based alphabetic scripts. For example, there is no capitals or lowercase letters. Each character has a reserved space of roughly the same size. In the following sections, the different scripts will be presented in greater detail. Their composition and use will be discussed.

1.2.1 Kana かな

If the Japanese had abolished the Chinese characters after formation of the syllabic scripts and used only those, studying the Japanese script would be a less complex task. (Lange 1922) reports about attempts to remove the Kanji from Japanese and use the Kana or even Latin script, so called ロマジ (*romaji*, a Latin or 'Roman' transcription of Japanese; Eng. lit. 'Roman characters'). However, none of those attempts succeeded and both Kana scripts serve as auxiliary scripts to the predominant Kanji characters.

The Chinese characters have been used in two ways in Japanese. Firstly, in order to express the morphological content of a character, but also in order to use the sound of the character as a syllable. The characters that have been used as syllables were transformed to two separate short-hand notations. One way used cursive writing of the sound kanji, reducing the character graphically, such that its original shape

女	め	Foljanty p. 42 show reduction to hiragana. use svg and arrow
曾	そ	

Figure 1.3: Reduction from Kanji to Hiragana

is became virtually incognisable. This development resulted in the Hiragana script:

The other method of reduction uses only one Grapheme to represent the whole character and lead to the development of the Katakana script:

伊	イ	Foljanty p. 42 show reduction to katakana. use svg and arrow
加	カ	

Figure 1.4: Redcuton from Kanji to Katakana

Using cursive or reduced characters became popular in the 9th century already. Both Hiragana ('smoothend Kana') and Katakana ('fragmented Kana') can easily be distinguished from the more complex Kanji. They represent a different linguistic content than the Kanji, namely rather a syllable than a morpheme. The fact that two parallel scripts came into existence can be explained by their use of different social groups. Hiragana are a product of the literarily active court ladies, while Katakana were developed in the Buddhist seminaries. The current system knows 46 Hiragana and Katakana for identical syllables (Foljanty 1984). See appendix ?? for a complete list of characters.

1.2.1.1 Hiragana ひらがな

The Hiragana are one of the three syllabic scripts. The third one, *Hentaigana* can be neglected, as it is not in active use any more. In principle each character represents a syllable, which can contain either a vowel (like 'u', う), a consonant and a vowel (like 'ta', た) or the nasal consonant ん ('n'). Hiragana are used for any words for which no Kanji exist, like から (*kara*, Eng. 'from') or grammatical particles like the object particle を (*wo*, works like a case marker). Additionally, Hiragana are used, when the Kanji is not known to the writer or reader. For instance, a writer may even passively know a Kanji, but not be able to actively produce it. Say the Kanji in question was 新 (Jap. pron. atara/あたら and shin/シン; Eng. 'new'). Example (2) shows first the spelling with the Kanji character in (2a), where the Hiragana parts are underlined. Then the alternative spelling without the Kanji character in (2b), with blanks in between, in order to visualise the individual characters and their pronunciation. The last part is identical, just the Kanji character has been replaced with an alternative hiragana spelling.

- (2) a. 新しい
atarashii
'new'
- b. あたらし い
a ta ra shi i
'new'

1.2.1.2 Katakana カタカナ**1.2.2 Composition of the Kanji** 漢字

xxx: see (Lange 1922) p.64

1.2.2.1 Graphemic Elements

xxx: see (Foljanty 1984) 2.1.4.2

1.2.2.2 Radicals**1.2.2.3 Readings**

Foljanty 2.1.7

1.2.3 Structure of the Japanese Writing System

Having demonstrated the Hiragana in 1.2.1.1, the Katakana in 1.2.1.2 and the Kanji in section 1.2.2, it is now possible to report about the structure of the writing system as such. It now becomes apparent, why a sentence like the one presented in example (1) is written and spelled the way it is.

The *Okurigana* (送り仮名, Jap. pron. オクリガナ/okurigana; Eng. 'accompanying characters') are Kana that form a word together with a Kanji character (Lange 1922). For instance, the word in example (3) has only one Kanji character: 帰 (Jap. pron. カエ/kae; Eng. 'go-home'), while the other characters are all Kana characters, that are part of the same word and modify the verb.

- り (*ri*): Flexion of the verb used for *tai*
- たく (*taku*): derivation of *tai* = want
- ない (*nai*): Negation ending

(3) 帰りたくない
 kaeritakunai
 go-home-want-not

'I do not want to go home'

1.2.4 Romaji ロマジ

xxx: see (Foljanty 1984) 4

1.2.5 Machine Writing of Japanese

Machine processing of the Japanese scripts has been an issue, ever since humans started to automate their writing.

xxx: see (Lange 1922) p. XII Stichwort Drucklegung xxx: see (Foljanty 1984) 5 xxx: see MS IME dscription (technical report or something?!) xxx: see section ?? for a description of research efforts in order to provide technology for using handwriting as an input method for Japanese. xxx: see (Grassmuck 1997)

1.3 Difficulties of Writing Japanese for Learners

xxx: find places for citations of the following paper (if not already done) (Foljanty 1984) (Lange 1922) (Katsuki-Pestemer 2006a) (Katsuki-Pestemer 2006b) (Haschke and Thomas 2008) (Tsujimura 2007) (Grassmuck 1997)

List of Figures

1.1	Kanji pictograms	7
1.2	Kanji ideograms	7
1.3	Reduction from Kanji to Hiragana	10
1.4	Redcution from Kanji to Katakana	10

List of Tables

1.1	Kanji phonograms	7
-----	----------------------------	---

References

- Chamberlain, B. H. (1982). *The Kojiki: Records of Ancient Matters* (2nd ed.). Boston, USA: Tuttle Publishing.
- Foljanty, D. (1984). Die japanische Schrift (in german). In Institut für deutsche Sprache Mannheim, T. Kaneko, and G. Stickel (Eds.), *Japanische Schrift, Lautstrukturen, Wortbildung*, Volume 1 of *Deutsch und Japanisch im Kontrast*, Chapter 2, pp. 29–63. Heidelberg: Julius Groos Verlag.
- Grassmuck, V. (1997). Die japanische Schrift und ihre Digitalisierung (in german). In W. Nöth and K. Wenz (Eds.), *Reden über Medien*, Volume II of *Reihe Intervalle. Schriften des WZ*. Kassel: Hochschulverlag.
- Guo, Z., K. Liua, X. Lua, H. Maa, K. Lia, S. Yuanb, and X. Wub (2000, October). The Use of AMS Radiocarbon Dating for Xia–Shang–Zhou Chronology. In *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, Volume 172, pp. 724–731. 8th International Conference on Accelerator Mass Spectrometry: Elsevier B.V.
- Haschke, B. and G. Thomas (2008). *Kleines Lexikon Deutscher Wörter Japanischer Herkunft - von Aikido bis Zen (in German)* (1st ed.). Munich: Beck.
- Kano, C., Y. Shimizu, H. Takenaka, and E. Ishii (1990, January). *Basic Kanji Book*, Volume 1. Bonjinsha.
- Katsuki-Pestemer, N. (2006a). *Grundstudium Japanisch 2 (in German)* (2nd ed.), Volume 2 of *Grundstudium Japanisch*. Troisdorf: Bildungsverlag EINS.
- Katsuki-Pestemer, N. (2006b). *Kanji-Arbeitsheft (in German)*, Volume 3 of *Grundstudium Japanisch*. Troisdorf: Bildungsverlag EINS.
- Lange, R. (1922). *Einführung in die Japanische Schrift (in German)* (2nd ed.). Number 15 in *Lehrbücher des Seminars für orientalische Sprachen*. Berlin: Walter de Gruyter.
- Noguchi, M. S. (2009, October 21). Get set for next year's overhaul of official kanji. *The Japan Times Online*. Online. Retrieved 2010-01-14 from <http://search.japantimes.co.jp/cgi-bin/ek20091021mn.html>.
- Tsujimura, N. (2007). *An Introduction to Japanese Linguistics* (2nd ed.). Blackwell Textbooks in Linguistics. Oxford: Blackwell Publishing.