# 6    Writing and sources

## 6.1    Writing

From the late Nara period a gradual process of simplification and abbreviation of the shapes of *man'yōgana* took place, slowly at first but gaining speed through the early part of the Heian period, and eventually resulting in the development of *hiragana* and *katakana* scripts. As the reduced shapes were dissociated from the *kanji* from which they derived, they unambiguously became phonograms and it became possible to write Japanese phonographically in a simple and straightforward manner. From the beginning of the tenth century until the end of the eleventh century we have a large body of text which was written almost entirely phonographically in *hiragana*, with only a few logographically used *kanji*. This style of writing is a continuation of the tradition of extensive phonographic writing which developed in the context of writing or recording poetry in the OJ period. Most of the EMJ *hiragana* texts were written by women who traditionally were excluded from the world of learning and Classical Chinese. The first line of the *Tosa nikki* from 935 where its author, Ki no Tsurayuki, somewhat tongue-in-cheek pretends to be a woman trying her hand at writing diaries, is famous: *wotoko mo su naru nikki to ipu mono wo wonna mo site mimu tote suru nari* 'I as a woman will also try keeping this thing called diary, which men are said to keep.' It also shows, however, that it is a stereotypical oversimplification to view prose writing in *hiragana*, or in Japanese, during the Heian period as the exclusive domain of women. Also both men and women wrote poetry in *hiragana*.

### 6.1.1    Kanji-kana majiribun

From the middle of the Heian period the style of writing known as *kanji-kana majiribun* ('mixed *kanji* and *kana* writing') started being used more widely. *Kanji-kana majiribun* has a heavy preponderance of *kanji*, used logographically for most content and some grammatical words, and with *kana* (almost always *katakana*), used phonographically for some grammatical words and endings. The earliest extant text identified as being written in *kanji-kana*

*majiribun* is the *Tōdaiji fujumonkō* from the early ninth century, but it was not until after the middle of the EMJ period that *kanji-kana majiribun* became more widespread and was used in particular in the *setsuwa* literature (6.2.1). From the LMJ period it became widely used and largely replaced *hiragana* writing in prose. Thus the writing of extensive text passages in *hiragana* (or *katakana*) did not survive the early LMJ period and, regrettably perhaps, the elegant and economic tradition of simple *hiragana* writing, which is a fully sufficient means of representing Japanese, was lost. It also is worth noting that the establishment and spread of *kanji-kana majiribun* from the late LMJ period coincides with the establishment of the written norm of Classical Japanese. The way Japanese is written today is a direct descendant not of the *hiragana* writing of the EMJ period, but of the *kanji-kana majiribun* of EMJ and LMJ, although the proportions of logographic and phonographic writing differ between then and now, with more words spelled out phonographically today and with *hiragana* replacing *katakana*.

All logographic writing of Japanese derives from reversing *kanbun-kundoku* (the rendition in Japanese of a Chinese text) from reading to writing (see 9.1.2), but the link is particularly close for *kanji-kana majiribun* which arose as a transfer of the techniques of *kunten* (annotations to a Chinese text) to the writing of Japanese. Especially the practice of writing grammatical elements of a Japanese rendition in *kana* glosses next to Chinese text served as the model for the *kanji-kana majiribun* style of writing. In 9.1.1 there is a short example of an annotated Chinese text and it is conspicuous that its appearance is quite similar to Japanese texts written in *kanji-kana majiribun*. Also the Japanese language used in *kanji-kana majiribun* was heavily influenced in style, vocabulary and grammar by the Japanese language used in *kanbun-kundoku*, the so-called *kuntengo* (9.1.6).

### 6.1.2    Kana (hiragana, katakana, hentaigana)

The *hiragana* and *katakana* developed as reduced shapes of *man'yōgana*. Table 6.1 shows the standard sets of *kana* and their *kanji* origins in the usual 'fifty-sound table' (*gojūonzu* 五十音図, see further 6.1.4) grid arrangement, in ten columns by five rows plus an extra slot for the moraic nasal. The columns (*gyō* 行) are referred to by the head *kana*, e.g. *a-gyō* ア行 the '*a*-column' or *ka-gyō* カ行 the '*ka*-column', and the rows (*dan* 段 (or *retsu* 列) by the single vowel *kana*, e.g., *a-dan* ア段 the '*a*-row' or *i-dan* イ段 the '*i*-row'. Of these, the following *kana* from the *wa*-column were abolished in the writing of modern Japanese in the script reform of 1946 and are today only used in the writing and presentation of Classical Japanese: ゐ/ヰ, ゑ/ヱ, を/ヲ (except that the latter pair is used for the accusative particle *o*). As may be seen, in some cases equivalent *hiragana* and *katakana* originate in the same

Table 6.1 Kanji origins of kana

| | a | ka | sa | ta | na | ha | ma | ya | ra | wa | N |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **row a** | a<br>あ 安<br>ア 阿 | ka<br>か 加<br>カ 加 | sa<br>さ 佐<br>サ 散 | ta<br>た 太<br>タ 多 | na<br>な 奈<br>ナ 奈 | ha<br>は 波<br>ハ 八 | ma<br>ま 末<br>マ 末 | ya<br>や 也<br>ヤ 也 | ra<br>ら 良<br>ラ 良 | wa<br>わ 和<br>ワ 和 | N<br>ん 无<br>ン 尓 |
| **row i** | i<br>い 以<br>イ 伊 | ki<br>き 幾<br>キ 幾 | si<br>し 之<br>シ 之 | ti<br>ち 知<br>チ 千 | ni<br>に 仁<br>ニ 仁 | hi<br>ひ 比<br>ヒ 比 | mi<br>み 美<br>ミ 三 | | ri<br>り 利<br>リ 利 | wi<br>ゐ 為<br>ヰ 井 | |
| **row u** | u<br>う 宇<br>ウ 宇 | ku<br>く 久<br>ク 久 | su<br>す 寸<br>ス 須 | tu<br>つ 川<br>ツ 州 | nu<br>ぬ 奴<br>ヌ 奴 | hu<br>ふ 不<br>フ 不 | mu<br>む 武<br>ム 牟 | yu<br>ゆ 由<br>ユ 由 | ru<br>る 留<br>ル 留 | | |
| **row e** | e<br>え 衣<br>エ 江 | ke<br>け 計<br>ケ 介 | se<br>せ 世<br>セ 世 | te<br>て 天<br>テ 天 | ne<br>ね 祢<br>ネ 祢 | he<br>へ 部<br>ヘ 部 | me<br>め 女<br>メ 女 | | re<br>れ 礼<br>レ 礼 | we<br>ゑ 恵<br>ヱ 恵 | |
| **row o** | o<br>お 於<br>オ 於 | ko<br>こ 己<br>コ 己 | so<br>そ 曾<br>ソ 曾 | to<br>と 止<br>ト 止 | no<br>の 乃<br>ノ 乃 | ho<br>ほ 保<br>ホ 保 | mo<br>も 毛<br>モ 毛 | yo<br>よ 与<br>ヨ 与 | ro<br>ろ 呂<br>ロ 呂 | wo<br>を 遠<br>ヲ 乎 | |

*kanji* (e.g. *ka, na, ma, ya, ra, wa*), in others not (e.g. *a, sa, ta, ha*). Overall, the reduction took place in two ways: as a general rule, *hiragana* resulted from a *cursive writing* of full *man'yōgana*, whereas *katakana* originate in *parts* of *man'yōgana*. It must be emphasized, however, that *man'yōgana* in their full shape continued to be used through the MJ period as phonograms. It is not the case that *hiragana* or *katakana* simply replaced *man'yōgana*: they coexisted for a long time.

Published editions of texts today use the standard sets of *kana*, but that inventory of graphs was only settled upon in the Elementary School Order (*Shōgakkōrei*) of 1900. Through the MJ period a large number of competing and variant, but phonographically equivalent, shapes of *kana* was used; these are today known as *hentaigana*. Table 6.2 below gives an impression of the variability among early *kana* shapes.

Legend ascribes the invention of *hiragana* to the priest Kūkai (774–835), but it is clear that *kana* were not invented by any individual or group. Many *man'yōgana* are graphically complex and cumbersome to write, and it is only natural that they would tend to be simplified and abbreviated for ease of writing. With the discovery of more primary material from the OJ period, mostly in the form of *mokkan* (wooden tablets, see 1.2.2), examples of graphically reduced *man'yōgana* are attested in increasing numbers also from the OJ period. It is important to note that many of the reduced shapes which became *hiragana* in Japan are found in Chinese cursive writing styles, and that many of the reduced shapes which became *katakana* are found in Korean *kugyŏl* (cf. 9.1.1). This suggests that in fact the *hiragana* and *katakana* letter shapes were not the result of independent developments in Japan, but followed continental models.

It is only possible to follow the gradual emergence of the *hiragana* and *katakana* letter shapes through authentic, contemporary writing, not later copies of manuscripts. In addition to *mokkan* and fragments of various sorts, many of which seem to be incidental writing, or even writing practice, the most important authentic, primary text materials from the EMJ period are annotations on texts, the so-called '*kunten*' materials (see 6.2.2). The study of the development of *kana*, especially *katakana*, and of *dakuten* (below) is intimately linked to the study of *kunten* materials. There exists a large amount of such materials in the possession of Buddhist temples in Japan, exhibiting great diversity in traditions of annotation, including letter shapes. Tsukishima (1986) reproduces in several hundred tables examples of *kana* shapes and other writing conventions drawn from such *kunten* materials. In Table 6.2 (taken from Tsukishima 1977: 98) are shown *kana* shapes (and a few other annotation conventions) from annotations (dating from 883) of the *Dìzàng shílún jīng* (地蔵十輪経, Japanese *Jizō-jūrin-gyō*), a Chinese translation of the 'Sutra of the Ten Cakras of Kṣitigarbha'. Even a cursory

Table 6.2 *Early* kana *shapes (from Tsukishima 1977: 98)*

| イフ | ン | ワ | ラ | ヤ | マ | ハ | ナ | タ | サ | カ | ア | ③ 平安初期の例㈢――「地蔵十輪経」(聖語蔵及び東大寺図書館蔵) 元慶七年(八八三)点 |

(中田祝夫博士調査を参照す)

inspection gives a clear impression of the diversity and also reveals that both *hiragana-* and *katakana-*like letter shapes are found in a single set of annotation. It is important to emphasize that all the letter shapes in Table 6.2 are from a single set of annotations and do not represent different steps in the development of letter shapes.

As such materials are further studied and published, the details of our knowledge about the development of the use of *hiragana* and *katakana* will change, but the general picture is as explained in this section.

The different strategies of reduction eventually resulted in two distinct sets of *kana*, *hiragana* and *katakana*, which were associated with functional differences: Parts of *man'yōgana*, which yielded *katakana*, were practical to use in annotations on texts where space is limited; in addition to that use, *katakana* also later became used in *kanji-kana majiribun*. Throughout the premodern period *katakana* maintained some connection with *kanbun* or writing incorporating *kanji*. Cursively written whole *man'yōgana*, eventually giving the modern *hiragana*, were thought aesthetically pleasing and were associated with literature and calligraphy. The *hiragana* were also practical for writing running text, as strings of letters could be written without lifting the brush from the paper. When talking about writing in the Heian period, *hiragana* has become known, and was also at the time occasionally referred to, as '*onna-de*' ('woman's hand'). Whereas academic, intellectual or public writing was mostly done in Classical Chinese (or in *hentai kanbun*, see 9.1.2.1), personal and private writing in Japanese in *hiragana* was associated with the leisure of the (female) aristocracy and as is well known, much of the prose literature in Japanese from this period was written in *hiragana* by women. One exception was composition and writing of poetry in *hiragana*, which to some extent was public and in which men certainly took part. As noted above it is wrong to view writing in *hiragana*, or in Japanese, during the Heian period as the exclusive domain of women.

It must further be noted that just as there was a gradual transition and long period of coexistence between *kana* and *man'yōgana*, so the letters which we today identify as *hiragana* and *katakana* were not functionally differentiated from the outset, or conceived of as belonging to different sets of letters; thus among the *hentaigana* we find *katakana*(-like) letter shapes, and conversely the *kunten* materials also hold many *hiragana*(-like) letter shapes mixed in with *katakana*, see Table 6.2. The *hiragana* and *katakana* for *e* illustrate this well: they originate in different *man'yōgana* which represented distinct syllables in OJ and early EMJ, え< 衣 (OJ /.e/) and エ < 江 (OJ /.ye/), and they can both be found within single bodies of text. Overall, however, the later functional differentiation between *hiragana* and *katakana* has its origin and basis in the two different strategies of simplification and tendencies in use.

### 6.1.2.1   Sei'on *and* daku'on

In OJ, different *man'yōgana* were generally used for *sei'on* and *daku'on* (i.e. syllables with initial tenuis (/p, t, k, s/) or *media* (/b, d, g, z/), respectively; see 2.2.2). Thus, for example, the syllables *ka* and *ga* could be written with different characters (e.g. *ka* 加; *ga* 我). However, even in OJ this phonological distinction was sometimes ignored, and for example a *man'yōgana* normally

used for *ka* could be used for *ga* (e.g. 加 sometimes used for the genitive
particle *ga*). In early EMJ the *sei–daku* distinction came to be consistently
ignored in general writing and the same *man'yōgana* were used for *pa/ba, ta/
da, ka/ga, sa/za*, etc. Accordingly, as the abbreviated *hiragana* and *katakana*
forms developed, separate letters for *sei'on* and *daku'on* did not evolve and
this phonological distinction remained unnoted through most of the MJ period
in general writing. Thus, although the tenues and mediae clearly were phono-
logically distinct in MJ, as in OJ and in NJ, the *kana* letters from the *pa, ta,
ka*, and *sa*-columns were used to represent /pV, bV/, /tV, dV/, /kV, gV/, and
/sV, zV/, respectively, such that for example は was used for both *pa* and *ba*,
た for *ta* and *da*, か for *ka* and *ga*, and さ for *sa* and *za*. This is reflected even
today in the absence of separate *kana* letters for *sei'on* and *daku'on*, with
*daku'on* being indicated by adding a diacritic, the *dakuten* , ゛, to the top right
corner of *kana* for *sei'on*, e.g. か *ka*, が *ga*.

Such a development towards a more underspecifying script is not unique,
but paralleled, for example, in the transition from the older (Germanic) to the
younger (Scandinavian) runes (completed by the middle of the ninth century).
As opposed to the older runes (with twenty-four letters), the younger system
of sixteen letters did not have separate letters for tense (unvoiced) and lax
(voiced) stops, in addition to other instances of equivalence. Thus there was
in the Scandinavian runes only one letter for /p, b/, /t, d/, and /k, g/, respec-
tively, although these sounds were and remained distinctive in the Scandinavian
languages and had had separate orthographic representation in the Germanic
runes. It is not difficult to understand that native readers have few problems
with a phonologically underspecifying script in general writing, as they know
the words in the language and can guess from context which words are meant,
even in the absence of an absolute phonological identification, in the same
way that phonologically or phonetically underspecified speech is usually
readily understandable. Thus, for example, most scripts have no expression of
suprasegmental features such as accent or tone. Note finally that although
tenues and mediae have been phonologically distinct through the history of
Japanese, a sizeable proportion of mediae derives from tenues either in mor-
phophonemic alternations as a result of *rendaku* (see 2.6.2), e.g. *sakura*
'cherry' + *pana* 'flower' => *sakurabana* 'cherry flower', or by automatic
phonological neutralization of tenues as mediae after nasals (cf. 7.1.2.2), e.g.
(*yom-* 'read' =>) *yoN-* + *-te* => *yonde* 'read.GER'. In such cases no ambiguity
in morpheme identification arises, even if the phonemic shape is not unam-
biguously noted.

### 6.1.2.2  Dakuten
In specialized writing, however, various means were used, when it was thought
necessary, to give a more precise indication of the phonological shape of a
word, noting whether a syllable was *sei* or *daku*, or its phonological pitch.

This was mostly done by diacritics, but there are also cases of, for example, inverted *kana* being used to specify a *daku'on*. The earliest attested use of diacritics to mark *sei* or *daku* on *man'yōgana* is from the late ninth century, and on *kana* from the eleventh century. Diacritics include for example 丷 , short for 濁(音) *daku(on)* (or the verb *nigoru* 濁 'pronounce as a *daku'on*'), for example 婆丷 to specify *ba* (and not *pa*). Usually, however, diacritics were dots, circles, lines, even triangles, or combinations of these. They could be added to *man'yōgana* or *kana*, but they were also added to logographically used *kanji* to give a hint to their reading. EMJ texts in which *sei-daku* (or pitch) was noted were almost exclusively annotations (*kunten* materials) or dictionaries, and *sei-daku* thus remained un-noted in general writing in the EMJ period. During the early LMJ period (Kamakura), annotation of EMJ texts in Japanese to specify features of pronunciation along those lines became current, but it was not until the NJ period, from the beginning of the Edo period, that the *dakuten* we know today became established and widespread in general writing.

The *dakuten* eventually settled upon appears to have had two sources: (Chinese) tone dots and the letter *anusvāra* from the *Siddham* script. In some documents, including the dictionary *Ruiju-myōgi-shō* from 1081 (6.2.3), *sei-daku* was noted in conjunction with pitch by means of tone dots which mark tone or pitch by their position next to a *kanji*; see 7.4.1. When extended to note *sei-daku*, the tone dots indicate pitch by *position*, while their *shape* indicate *sei-daku*, usually using single dots or circles for *sei'on*, but double dots or circles for *daku'on*.

However, separate from the notation of pitch was a diacritic of the shape ⸜ which in early annotations was placed next to or below a letter, but later customarily on the right hand top corner, the position in which the present-day *dakuten* is placed. This diacritic is thought to have its origin in the letter *anusvāra* from the Indic script *Siddham*. The *Siddham* script was used to write Sanskrit and was brought to Japan in the early ninth century, most likely by the priest Kūkai when he returned in 806 from his study tour to China where he also studied Sanskrit. *Siddham* was, in India and other places, later generally replaced by the *Devanagari* script to write Sanskrit, but *Siddham* has remained in use in Shingon Buddhism in Japan, mainly for copying out sutras and mantras. The letter *anusvāra* represents a nasality feature of Sanskrit. It is usually romanized as -ṃ, but in *Siddham* it appears as a dot above a syllable, e.g. *sa* स versus *saṃ* सं. The diacritic ⸜ was used with *kana*, *man'yōgana*, or with *kanji* used logographically for Chinese words. In addition to marking *daku'on*, it is also in some annotations used to mark syllables with an initial nasal (*n-*, *m-*); there are also cases of it being added to *kana* for .*u* or .*i* to note a nasalized vowel [ũ] or [ĩ] which were renditions of Chinese /-ŋ/, e.g., ʔ [ũ] or イ [ĩ]. This diacritic was in other words used mainly as a nasality mark;

this provides further evidence that the mediae phonetically were pre-nasalized in EMJ (see 2.2, 7.1.4.3, and 11.1) and were associated with the phonetic feature of nasality also in metalinguistic consciousness.

### 6.1.2.3  Handakuten

OJ /p/ was lost in most contexts in the course of sound changes which occurred through the MJ period, merging with /-w-/ in medial position after vowels (but preserved after /Q/) in EMJ (7.3.1.2), and changing to /f-/ in initial position in LMJ; see 11.3 for details and exceptions. The limited contexts in which /p/ was preserved were readily recognizable and a specific way of writing /p/ as distinct from /f/ did not develop in the MJ period. However, towards the end of the LMJ period Portuguese Jesuit missionaries instituted the use of a circle on the top right corner of a *kana* for contemporary *fV* (present-day *hV*) in order to write unambiguously *pV*, e.g. は (*fa/ha*), ぱ (*pa*). This was first used in the *Rakuyōshū*, a *kanji* dictionary published by the Jesuit press in Amakusa in 1598. Since then this practice gradually spread and is, of course, today a fully integrated feature of Japanese writing.

### 6.1.2.4  Orthographic categories; the Iroha-uta, the Japanese 'alphabet'

Although there was a large inventory, including variant shapes, of *man'yōgana* and early *kana* letters with widespread phonographic equivalence (i.e., many different letters for the same sound), the establishment and awareness of distinct orthographic categories is evident from three mnemonic word lists or poems from the first half of the Heian period in which each distinct letter category occurs only once: the *Ame-tsuchi no kotoba* (or *Ame-tsuchi no uta*), *Taini-uta*, and *Iroha-uta* (see (1)–(3) on pp. 166–7). These lists/poems functioned as a kind of ABC to remember the distinct letter categories and were also used for writing practice. The *Ame-tsuchi no kotoba* is basically a list of words, but the *Taini-uta* and *Iroha-uta* are organized into phrases or clauses conforming to some extent to the native poetic metre of alternating seven- and five-mora-long verse lines. The *Ame-tsuchi no kotoba* has forty-eight letters, but the latter two have the forty-seven *kana* letter categories which, with the later addition of ん/ン, are in use today.

The three lists concur in not having distinct letters or representation for *sei'on* and *daku'on*. Chronologically, they all reflect a stage of the language *after* the merger of the *kō–otsu* distinctions (as there is no representation of these distinctions), but *before* the merger of non-initial /-p-/ with /-w-/ (950–1000, see 7.3.1.1), as some *pV kana* are illustrated in positions in which /p/ later merged with /w/, e.g. かは *kapa* 'river' (> *kawa*) distinct from ゆわ *yuwa* 'sulphur', as well as before the merger of /.wo/ and /.o/ (*c.* 1000, see 7.3.2.3), as both を *wo* and お *o* find distinct representation. The *Ame-tsuchi no kotoba* alone reflects a stage of the language prior to the merger

(1)    *Ame-tsuchi no kotoba*

あ め つ ち ほ し そ ら や ま か は み ね た に く
も き り む ろ こ け ひ と い ぬ う へ す ゑ ゆ わ
さ る お ふ せ よ え の え を な れ ゐ て

| あめ | | つち | ほし | そら | やま | | かは | みね |
|---|---|---|---|---|---|---|---|---|
| *ame* | | *tuti* | *posi* | *sora* | *yama* | | *kapa* | *mine* |
| heaven | | earth | star | sky | mountain | | river | peak |

| たに | | くも | きり | むろ | こけ | | ひと | いぬ |
|---|---|---|---|---|---|---|---|---|
| *tani* | | *kumo* | *kiri* | *muro* | *koke* | | *pito* | *inu* |
| valley | | cloud | mist | room | moss | | man | dog |

| うへ | | すゑ | ゆわ | さる | おふ | | せよ | |
|---|---|---|---|---|---|---|---|---|
| *upe* | | *suwe* | *yuwa* | *saru* | *opu* | | *seyo* | |
| above | | end | sulphur | monkey | grow | | do! | |

| え | | の | え | を | なれ | | ゐて | |
|---|---|---|---|---|---|---|---|---|
| *e* | | *no* | *ye* | *wo* | *nare* | | *wite* | |
| hackberry tree | | GEN | branch | ACC | get used to | | sitting | |

(2)    *Taini-uta*

た ゐ に い て な つ む わ れ を そ き み め す と
あ さ り お ひ ゆ く や ま し ろ の う ち ゑ へ る
こ ら も は ほ せ よ え ふ ね か け ぬ

| たゐ | に | いて | | な | つむ | われ | を | そ |
|---|---|---|---|---|---|---|---|---|
| *tawi* | *ni* | *ide* | | *na* | *tumu* | *ware* | *wo* | *zo* |
| paddy | DAT | go out | | greens | pick | I | ACC | FOC |

| きみ | めす | と | | あさり | | おひゆく | | |
|---|---|---|---|---|---|---|---|---|
| *kimi* | *mesu* | *to* | | *asari* | | *opi-yuku* | | |
| you | see.RESP | when | | hunt | | chase-after | | |

| やましろ | | の | | うちゑへる | | こら | | |
|---|---|---|---|---|---|---|---|---|
| *yamasiro* | | *no* | | *utiweperu* | | *kora* | | |
| Yamashiro | | GEN | | is very drunk | | dear.girl | | |

| もは | | ほせよ | え | | ふね | かけぬ | | |
|---|---|---|---|---|---|---|---|---|
| *mopa* | | *poseyo* | *e-* | | *pune* | *kakenu* | | |
| seaweed | | dry! | (un)able- | | boat | doesn't anchor | | |

(3)   *Iroha-uta*

い ろ は に ほ へ と ち り ぬ る を わ か よ た れ
そ つ ね な ら む う ゐ の お く や ま け ふ こ え
て あ さ き ゆ め み し ゑ ひ も せ す

| いろ | は | にほへと | | | | ちりぬる | を |
|------|-----|---------------|---|---|---|-----------|------|
| iro | pa | nipopedo | | | | tirinuru | wo |
| colour | TOP | although shines beautifully | | | | will scatter | EXCL |

| わ | か | よ | たれ | そ | | つね | ならむ |
|-----|-----|------|------|-----|---|-------|--------|
| wa | ga | yo | tare | zo | | tune | naramu |
| I | GEN | world | who | FOC | | always | will be |

| うゐ | | の | おくやま | | けふ | こえて |
|------|---|-----|----------|---|------|--------|
| uwi | | no | okuyama | | kepu | koete |
| material.world | | GEN | deep.mountain | | today | crossing |

| あさき | ゆめ | みし | ゑひ | も | せす |
|--------|------|--------|------|-----|------|
| asaki | yume | mizi | wepi | mo | sezu |
| shallow | dream | won't see | get drunk | ETOP | doesn't |

'Colors are fragrant, but they fade away. In this world of ours none lasts forever. Today cross the high mountain of life's illusion [i.e. rise above this physical world] and there will be no more shallow dreaming, no more drunkenness [i.e. there will be no more uneasiness, no more temptations].'
(Translated by Andrew N. Nelson (1974: 1014))

of the syllables /.ye/ and /.e/ (before 950, see 7.3.2.2). The *Ame-tsuchi no kotoba* is not attested until the *Minamoto Shitagō-shū*, compiled 967, in *hiragana*, but it has two occurrences of *hiragana* え *e*, interpreted as representing /.e/ and /.ye/, respectively, as there is no other good explanation for the double occurrence of え. The current *kana* inventory of course does not include representation of that earlier distinction, but in addition to *man'yōgana*, *kana* annotations from the ninth and early tenth century have distinct letters for /.e/ and /.ye/, shown under 衣 in the *a*-column and under 江 in the *ya*-column in Table 6.2 above, and as mentioned the current *hiragana* and *katakana* in fact derive from early *kana* for /.e/ and /.ye/, respectively: え< 衣 /.e/, while エ < 江 /ye/.

The *Taini-uta* is attested first in 970 (in the *Kuchizusami*, a first primer for the education of boys of the nobility, compiled by Minamoto no Tamenori, ?–1011), which is probably fairly close to the time it was made, but the first attestation of the *Iroha-uta* is in the *Konkōmyō saishōō kyō ongi* from 1079, although it must have been made earlier. Nothing is known about the

authorship of the *Iroha-uta*. It has, like the *hiragana* script itself, been attrib-
uted to Kūkai, but the phonological system reflected in the *Iroha* postdates
Kūkai, who died in 835, by a century and a half.[1] The *Iroha-uta* soon super-
seded the *Ame-tsuchi no kotoba* and *Taini-uta*, which today are not generally
known, and from the late Heian period became widely and commonly used to
remember the *kana* categories and for writing practice.

It is difficult to exaggerate the importance of the establishment and dis-
semination of the *Iroha-uta*. It is a defining event in the history of Japanese
script and writing, and more generally in Japanese culture, linguistic and
otherwise, and for example played an essential role in the spread of literacy
in Japan through the LMJ and NJ periods. The *Iroha-uta* finally defined the
forty-seven *kana* letter categories which, with the sole addition of ん/ン for
the moraic nasal, still are in use today, enshrined the principle of not having
separate letters for *sei'on* and *daku'on*, and fixed the sequence of the letters
in a list. Phonographically, the *Iroha* inventory may be seen to have some
shortcomings, viz. the systematic underspecification of the *sei'on–daku* dis-
tinction, and a lack of distinct, single-letter representation of 'new' sounds in
the language (see 6.1.2.5), so that as the *Iroha* letter categories became set in
stone, the relation between sound and written representation over time was
obscured by sound changes (see 6.1.3). However, the *Iroha* represents a good
phonological analysis of the number of (free) moras in the language at the
time it came into being, and it successfully established and fixed a small
manageable inventory of letter categories, thus providing a writing system
which was learnable for larger groups of people. The fact that the letter cat-
egories of the *Iroha* by and large are those used to write Japanese today is
testimony to its success.

In addition to establishing the letter categories and being used for writing
practice, the *Iroha-uta* also in more general cultural terms assumed the func-
tions of an *alphabet* or *ABC*. It came to be used as a sequence for enumeration,
and is still today sometimes used in lists, or in subclassification within diction-
ary entries. One enormous impact of the *Iroha* was in its use for ordering
('alphabetizing') dictionaries from as early as the end of the Heian period
(used first in the *Iroha jiruishō* compiled between 1144 and 1181), making
possible the compilation of dictionaries with easy, phonologically based look-
up. The *Iroha* continued to function as the Japanese alphabet, as a linguistic
and meta-linguistic organizing principle until it eventually was replaced by
the *gojūonzu* in the middle of the Meiji period (6.1.4). The word *Iroha* (written
variously as いろは, 伊呂波, 色葉) also came to be used as a name for the

---

[1] It is interesting to note that *iropa* is a word meaning 'natural mother' which could be taken in
the sense of 'origin, beginning', here of writing. The similarity between the words *alphabet* and
*iropa* is surely fortuitous.

*kana* inventory and by extension to mean 'ABC' in the sense of 'rudiments, basics (of something)', cf. living expressions such as *iroha no 'i' no ji mo shiranai* (いろはの 「伊」 の字も知らない) 'be utterly illiterate, or ignorant; lit.: not even know the 伊 [first] letter of the *Iroha*'.

### 6.1.2.5 New sounds

In the transition between OJ and EMJ and during the EMJ and LMJ periods, a number of sound changes occurred. Some of these were mergers, or loss of phonological distinctions, for example the loss of the *kō–otsu* distinction, resulting in fewer orthographic distinctions being necessary. Others, however, introduced entirely new phonemes into the language, all related to syllable structure: *bound moras* (6.1.2.5.1), *complex syllable onsets* (6.1.2.5.2), *and SJ syllable final* -t (6.1.2.5.3). As the *kana* letters were reduced forms of *man'yōgana*, which reflected the phonology of OJ, they did not readily provide for the representation of sounds which were new to EMJ, so although some of these new sounds were present in the language at the time the *Iroha* was established, the letter categories of the *Iroha* do not include distinct representation of these new sounds.

**6.1.2.5.1 Bound moras** Through the *onbin* sound changes (7.1.4), Japanese became quantity sensitive and acquired long syllables which included bound moras (7.1.1): *moraic consonants* /C, Q, N/ and *moraic vowels*, both *oral* /I, U/, and *nasal* /Ĩ, Ũ/. Of these only the moraic nasal /N/ eventually got its own letter, the only addition to the letter categories of the *Iroha*. Various means were employed to write the *moraic consonants* /C, Q, N/ which appeared in the language in the transition between OJ and EMJ: Apart from not being represented (e.g. *moQ-te* (<*motite*) 'holding; hold-GER' written as モテ *mo.te* (in annotations from the late ninth century) or *siN-zi* (< *sini-si*) 'dead; die-SPST.ADN', written (in the *Tosa nikki*, 935) as しし *si.si*), they have been written with a number of different *kana* for CV syllables. There was some overlap – within individual texts – in representation, and for example *man'yōgana* 牟 (*mu*) or its reduced *katakana* shape ム have both been used to represent both *N* and *Q*; this reflects the phonemic identity between *N* and *Q* (as /C/ unspecified for nasality) in morpheme-internal position (cf. 7.1.2.1). The current use of *kana* for *tu* (つ) to write the moraic obstruent /Q/ was originally inspired by the *kana* transcription of syllable final -*t* as つ (*tu*) (cf. 6.1.2.5.3) and was settled upon relatively late. The convention of writing this letter in small size dates from the script reform of 1946. The *hiragana* letter ん for the nasal moraic obstruent /N/ is thought to derive from the *man'yōgana* 无 (*mu*), though some scholars believe it to be a further development from (a precursor of) *katakana* ム (itself < 牟). Table 6.2 above shows two letter shapes for /N/, one like present-day *hiragana* ん and the other

seemingly a precursor of *katakana* ン. However, the *katakana* letter ソ seems to be first attested as such in the eleventh century. It appears to have no recognizable source in a *man'yōgana*, but to have been invented specifically for the purpose of representing /N/. The dot in ソ may well reflect the nasality diacritic ˇ which was also a source of the *dakuten* (6.1.2.2).

The bound moraic vowels, /I, U, Ĩ, Ũ/, emerged through the *onbin* sound changes, and also came to be used in SJ loanwords. These sounds were written with *kana* for the free moras /i, u/, i.e. い, う, with no differentiation of oral and nasal, apart from the use in some annotations of the nasality diacritic ˇ with *kana* for .*i*, .*u* to specify /Ĩ, Ũ/ in renditions of Chinese /-ŋ/.

**6.1.2.5.2 Syllables with complex onsets** Syllables with complex onsets, consisting of consonant + (palatal or labial) glide, represented in the moras *Cya*, *Cyo*, *Cyu* and *kwa*, *gwa*, *kwe*, *gwe* (7.2), entered the language in the EMJ period through the large-scale adoption and adaptation of SJ vocabulary. Somewhat later *Cya*, *Cyo*, *Cyu* also arose outside SJ vocabulary through sound changes.

Also these sounds were not written with distinct single letters, but instead as combinations of *kana* for *Ci* or *Cu* with *yV* and *wa*, i.e., *Ciya*, *Ciyo*, *Ciyu* (e.g. きや *kya*, きよ *kyo*, きゆ *kyu*) and *kuwa*, *guwa*, *kuwe*, *guwe* (くわ *kwa*, ぐわ *gwa*, くゑ *kwe*, ぐゑ *gwe*), respectively. Since the script reform in 1946 the glide initial letter has been written in smaller size: きゃ, きょ, きゅ; くわ, ぐわ.

**6.1.2.5.3 Syllable final /-t/** A sound found only in SJ loanwords was syllable final /-t/ (cf. 7.1, 7.2). This remained pronounced as /-t/, i.e. without a following vowel, until well into the NJ period (see further 11.4). In annotations it was noted in a number of different ways, but in general writing it was transcribed by *kana* つ or sometimes ち. It has recently been shown, however, that some LMJ sources made an orthographic distinction between /-tu/ and /-t/ by means of variant *kana* (*hentaigana*) which were originally used as equivalents for /tu/. In NJ final /-t/ acquired an epenthetic vowel to give *tu* or *ti*.

**6.1.2.5.4 Recent Modern Japanese loanwords** Finally, we should note 'new sounds', and especially new combinations of sounds, which have come into the Japanese language in the contemporary period, since the Meiji restoration, through the intake of large numbers of western loanwords (see 14.6). Like moras with complex onsets, the moras containing these sounds have been written with combinations of *kana*, e.g., *fan* ファン, *paatii* パーティー 'party', *sherii* シェリー 'sherry', *tsaitogaisuto* ツァイトガイスト 'zeitgeist', *jerii* ジェリー 'jelly', *vandaru-jin* ヴァンダル人 'Vandals (Germanic

tribe)'. The moras /wi, we, wo/ have reappeared in the language, but they are today written as <u>wirusu</u> ウイルス 'virus', <u>wesutan</u> ウエスタン '(a) western', <u>wokka</u> ウオッカ 'vodka', rather than using the old *kana* for these moras (\*ヰルス, \*ヱスタン, \*ヲッカ).

### 6.1.3 Orthographic norms

After the establishment of the orthographic categories in the *Iroha*, a number of sound changes took place through the MJ period which rendered several originally phonographically distinct *kana* categories equivalent, in some cases unconditionally and in others only in word non-initial position: /-p-/ merged in intervocalic position with /-w-/ as /-w-/, but remained /-p-/ after /Q/ (7.3.1), and /w/ was lost in both word initial and non-initial position, first before /o/, and later before /i, e/ (7.3.2.3). These changes are often stated as syllable mergers, as shown in (1) which lists the merged syllables, the outcome of the merger, and the previously phonographically distinct *kana* categories which could now be used to represent the outcome of the merger.

(4)     Sound changes resulting in *kana* equivalence

|  |  |  |  |  |
|---|---|---|---|---|
|  | **In initial position** |  |  |  |
| a. | /.o/, /.wo/ | > | /.o/ | お を |
| b. | /.i/, /.wi/ | > | /.i/ | い ゐ |
| c. | /.e/, /.we/ | > | /.e/ | え ゑ |
|  | **In non-initial position** |  |  |  |
| d. | /-.o/, /-.wo/, /-po/ | > | /-.o/ | お を ほ |
| e. | /-.i/, /-.wi/, /-pi/ | > | /-.i/ | い ゐ ひ |
| f. | /-.e/, /-.we/, /-pe/ | > | /-.e/ | え ゑ へ |
| g. | /-.wa/, /-pa/ | > | /-.wa/ | わ は |
| h. | /-.u/, /-pu/ | > | /-.u/ | う ふ |

These sound changes had two consequences for the representational value of *kana*: First, the *kana* categories お を, い ゐ, and え ゑ, respectively, became phonographically equivalent in all contexts. Second, the representational value of the *p*-column *kana* changed in non-initial position, but the fact that there was no orthographic distinction between *sei 'on* and *daku 'on* and that the *kana* from the *p*-, *t*-, *k*-, *s*- columns were used to represent both *sei 'on* and *daku 'on* (/pV, bV/, /tV, dV/, /kV, gV/, and /sV, zV/, respectively) complicated the situation further. In initial position, as before, は could represent /pa/ or /ba/ (mainly in SJ loanwords which had no restrictions on the occurrence of word initial media, see 7.2), and likewise ひ could stand for /pi/ or /bi/. However,

in non-initial position は could now represent /wa/ or /ba/, or, after /Q/, /pa/. Likewise, ひ came to represent /i, bi/ (or /pi/), へ could represent /e, be/ (or /pe/), ほ /o, bo/ (or /po/), and ふ /u, bu/ (or /pu/). The result of this was partial phonographic overlap, but not full equivalence, of non-initial ほ, ひ, へ, は, and ふ with お/を, い/ゐ, え/ゑ, わ, and う, respectively. See (2) which shows the phonographic value of はひふへほ, and the partial phonographic equivalence with other *kana* in non-initial position:

(5)

|  | Initial position | | Non-initial position | | /Q/ __ |
|---|---|---|---|---|---|
| は | /pa/ | /ba/ | /ba/ | /wa/ (= わ) | /pa/ |
| ひ | /pi/ | /bi/ | /bi/ | /i/ (= い, ゐ) | /pi/ |
| ふ | /pu/ | /bu/ | /bu/ | /u/ (= う) | /pu/ |
| へ | /pe/ | /be/ | /be/ | /e/ (= え, ゑ) | /pe/ |
| ほ | /po/ | /bo/ | /bo/ | /o/ (= お, を) | /po/ |

Thus, for example, *kawa* 'river' could now be written かわ or かは, whereas *kaba* 'Japanese white birch' would only be written かは. Likewise, *kai* 'shellfish' could be written かい, かゐ, or かひ, whereas *kabi* 'mildew' would only be written as かひ. Conversely, from the reader's point of view, かわ spelled *kawa*, whereas かは spelled *kawa* and *kaba*; and かい and かゐ unambiguously spelled *kai*, whereas かひ spelled *kai* and *kabi*. Taking account of the spelling of words with /-Qp-/ complicates the situation further. For example, *kappa(-to)* 'with a thud' could also be spelled かは (depending on how /Q/ was represented). In (6), the spellings of some words which illustrate the polyvalence are listed, showing the potential difficulties for both readers and writers:

(6) *kawa* 'river'　　　　　　　　　　かは　かわ
　　*kaba* 'Japanese white birch'　かは
　　*kappa(-to)* 'with a thud'　　　かは

　　*kabi* 'mildew'　　　　　　　　かひ
　　*kai* 'shellfish'　　　　　　　　かひ　かゐ　かい

　　*kae* 'change (tr.); inf.'　　　　かへ　かゑ　かえ
　　*kabe* 'wall'　　　　　　　　　　かへ

　　*io* '500'　　　　　　　　　　　いほ　いを　いお
　　*ibo(ziri)* 'praying mantis'　　いほ

　　*kau* 'change (tr.); concl.'　　かふ　かう
　　*kabu* 'head'　　　　　　　　　かふ

Clearly, for the *kana* categories おを𛀙, いゐひ, えゑへ, うふ, わは, the mapping between sound and writing became much less straightforward than it had been at the time the *kana* letter categories were established and they more closely reflected the phonology of the language, and as may be expected, this led to efforts to regularize the use of these *kana* letters. Rather than reforming the inventory and use of *kana* letters along simple phonographic lines (for example, by dropping the letter categories をゐゑ altogether and specializing non-initial はひふへほ to represent /bV/ and /pV/ after /Q/), an etymological spelling principle gained ground from around the beginning of the thirteenth century, eventually resulting in the so-called *rekishi-teki kana-zukai* (歴史的 仮名遣い) 'historical *kana*-usage, historical spelling', which is still in use today in the presentation of premodern Japanese texts and in writing Classical Japanese. The simple principle is that any word is to be spelled the way it was, or would have been, spelled within the *kana* categories of the *Iroha*, before the sound changes outlined in (4) took place. Thus *kai* 'shellfish' (< OJ *kapi*) is spelled か̲ひ̲, *ai* 'indigo' (< *awi*) is あ̲ゐ̲, and *kai* 'rudder' (< *ka.i*) is か̲い̲. In addition to the notion of faithfulness to the shape of earlier texts, the adoption of the etymological spelling principle seems to show that the orthographic categories in the *Iroha-uta* were taken as given and not to be tinkered with. The widespread adoption of the etymological spelling principle means that many text copies as well as much writing preserves orthographic distinctions long after the phonological distinctions they originally represented had been lost, making such materials useless for the dating of sound changes.

At the beginning of the NJ period further sound changes took place which resulted in phonographic equivalence between more *kana* categories: /z/ and /d/ merged (as /z/) before the high vowels /i, u/ (see 14.1), rendering the *kana* pairs じ/ぢ and ず/づ phonographically equivalent. By then the etymological principle had long been established and so the principle for which *kana* to use was fairly clear, although usages which can be thought of as 'spelling mistakes' (for example writing *mizu* 'water' < *midu* as み̲ず̲ rather than the etymologically correct み̲づ̲) are frequent in Edo period NJ texts.

The *rekishi-teki kana-zukai* was only abolished as the norm with the orthographic reforms in 1946 when the *gendai kana-zukai* ('present-day *kana* usage, present-day spelling') was adopted, in which the etymological principle largely has been abandoned, with a few well-known exceptions. For example, the second half of the long vowel /oo/ is mostly written う (e.g. *gakkoo* 'school' が̲っ̲こ̲う̲, *oogi* 'fan' お̲う̲ぎ, *toota* 'asked' と̲う̲た), but お when having earlier been spelled ほ or を (e.g. *too-* 'far' と̲お̲ (< とほ) or *too* 'ten' と̲お̲ (< とを)); and the grammatical particles *wa* 'topic', *e* 'allative', and *o* 'accusative' retain their earlier spelling は, へ, and を; finally, /zu/ and /zi/ (/ji/) are usually spelled ず and じ, except where they derive from *rendaku* of /tu/ (/tsu/) and /ti/ (/chi/) in transparent compounds (e.g., *mi-zika* 'close, familiar'

み<u>ち</u>か, cf. *mi* 'body' and <u>*tika-*</u> 'close' or *kana-<u>zukai</u>* '*kana*-usage' かな<u>づ</u>か
い, cf. <u>*tsukai*</u> 'usage').

The etymological spelling principle was first explicitly proposed by the poet
and scholar Fujiwara no Teika (1162–1241) as part of his recommendations
of the spelling of individual words set forth in works such as *Gekanshū* (?1241,
a book on poetics) or *Hekianshō* (1226, annotations on poems from earlier
poetry anthologies). His primary concern was the production of faithful
editions of earlier texts, preserving their original shape, and he based his
spelling proposals on inspection of earlier manuscripts and text versions in
order to establish original, 'correct' spellings. It seems that there was some
consensus about the spelling of individual words before Teika, at least to some
extent on an etymological basis, as shown for example in the surviving hand-
written letters of the wife of Fujiwara no Tamefusa (1049–1115) in which the
distinctions between /-p-/ and /-w-/ and between /.wo/ and /.o/ are maintained
orthographically although the sound changes in (4) had already taken place,
but it was Teika who explicitly established the etymological spelling principle
which, together with his specific spellings of individual words, came to be
known as *Teika kana-zukai* 'Teika *kana*-usage, Teika spelling' and soon
acquired the status of a spelling norm, not just for copying, but also for com-
posing text. Thus, from Kamakura to early Edo most literary writing was
spelled according to the *Teika kana-zukai*.

There are inevitably not a few mistakes in Teika's etymological spellings
(for example *too-* 'far' < mid EMJ *towo* < OJ *topo-* spelled とを 'to.wo'
instead of etymologically correct とほ 'to.po'), first of all because the copies
of texts he had access to and worked from already contained errors.
Furthermore, his recommendations concern only a relatively small number of
words and only address the use of えゑへいゐひおを.[2] Therefore later schol-
ars elaborated on and amplified Teika's work. The *Kanamoji-zukai* (completed
after 1363) by the monk Gyōa (fourteenth century, dates unknown, born
Minamoto no Tomoyuki) is the first large-scale application of Teika's princi-
ples. It lists spellings for more than a thousand words and for a long time
served as the major normative spelling reference. In addition to unsuccessfully
attempting to credit his own grandfather, rather than Teika, with the establish-
ment of the etymological spelling principle, Gyōa takes the important step of
incorporating also the letter categories ほわはうふ (as well as む which was

---

[2] However, with regard to お and を, Teika partly employed a phonographic, rather than etymo-
logical, spelling principle, following the example of earlier dictionaries like *Iroha jiruishō* (late
Heian) in assigning letters in word initial position according to phonetic pitch, such that /.o-/
would be written with を if it was [high], e.g. をく *oku* (< OJ *oku*) 'fall, descend (of dew, snow,
frost); put', but with お if it was [low], e.g. おる *oru* (< OJ *woru*) 'break'. This caused much
confusion later, as pitch changed, obscuring the basis for the spellings which had become fixed
and normative.

problematic for entirely different reasons). Mention must also be made of the *kokugaku* scholar Keichū (1640–1701) who provided the basis for the *rekishi-teki kana-zukai* in use today. In his *Waji shōranshō* (1695) he set the etymological spelling principle on a firm philological footing with systematic citation for specific spellings of sources which even by today's standards mostly are reliable.

### 6.1.3.1 Undoing the etymological kana-spelling

In printed editions premodern texts are usually presented in the etymological *kana*-spelling, so that for example the word for 'front' is written まへ regardless of the date of the text. Table 6.3 is a simple guide to phonemic transcription of EMJ and LMJ texts, giving the phonemic shapes represented by the *kana* letters えゐゑをはひふへほ in different periods, reflecting the sound changes which took place during the EMJ and LMJ periods (see 7.3, 11.2, 11.3). As these sound changes generally are not systematically reflected in the copies available to us today of most of the literary texts (due to the application of the etymological spelling principle in copying the texts), the assignment of texts to definite stages within these sound changes is to some extent arbitrary. This holds particularly for the texts dating just after 1000, including important literary texts such as the *Genji monogatari*, *c.* 1001–1010, and the *Makura no sōshi*, *c.* 1000. It is clear that these texts postdate the change of /-p-/ to /-w-/ (cf. 7.3.1) and that what is written かは in editions of these texts represents /kawa/. However, it is difficult to say for certain that the change of /wo/ to /o/ (cf. 7.3.2.3), which is thought to have been complete *c.* 1000, was in fact entirely complete by the time these texts were written, and it is therefore less certain that what is written かほ in editions of these texts represents /kao/ and not /kawo/; it is quite certain, though, that it does not represent /kapo/. Note also that the difference between /.e/ and /.ye/, which pertains only to morpheme initial position, is not represented in the standard inventory of *kana* letter categories (cf. 6.1.2.4), but it is relevant to the earliest *kana* materials, such as *Tosa nikki* or *Kokinwakashū*; it can easily be verified in good dictionaries whether words written with initial え had /.ye/ (e.g. *yeda* 'branch') or /.e/ (e.g. *e-* 'to get'); in morpheme non-initial position え always represents /.ye/ before *c.* 950 (e.g. *nuye* 'thrush').

Through the EMJ and LMJ periods, し, ち, つ, じ, ず, ぢ, づ straightforwardly represent /si, ti, tu, zi, zu, di, du/ and should be transcribed *si, ti, tu, zi, zu, di, du*, e.g. つち *tuti* 'earth', しづか *siduka* 'quiet', みず *mizu* 'doesn't see', みじ *mizi* 'won't see', and やまぢ *yamadi* 'mountain path'. Only in texts dating from the early NJ period do じ and ぢ both come to represent /zi/ and should be transcribed *zi*, and ず and づ both come to represent /zu/ and should be transcribed *zu*. In terms of the etymological *kana* spelling (which in text editions supplies *dakuten* although these were not used in general writing, cf.

Table 6.3 *Phonemic transcription of the historical kana-spelling*

| Kana | Examples | <950<br>*Tosa nikki, Kokinwakashū, Ise monogatari* | c. 950<br>*Taketori monogatari** | 950 ><br>*Kagerō nikki, Utsuho monogatari, Ochikubo monogatari* | 1000 ><br>*Genji monogatari, Makura no sōshi* | 1100 ><br>*Konjaku monogatari, Heike monogatari, Hōjōki* | 1300 ><br>*Tsurezure-gusa, Soga monogatari* |
|---|---|---|---|---|---|---|---|
| え | え | e | | | | | |
| | えだ | yeda | eda | | | | |
| | ぬえ | nuye | nue | | | | |
| ゐ | ゐ | wi | | | | | i |
| | まゐる | mawiru | | | | mairu | |
| ゑ | ゑみ | wemi | | | | | emi |
| | こゑ | kowe | | | | koe | |
| を | をか | woka | | | oka | | |
| | あを | awo | | | ao | | |
| は | はら | para | | | | | fara |
| | かは | kapa | | kawa | | | |
| ひ | ひら | pira | | | | | fira |
| | かひ | kapi | | kawi | | kai | |
| ふ | ふね | pune | | | | | fune |
| | たふ | tapu | | tau | | | |
| へ | へ | pera | | | | | fera |
| | まへ | mape | | mawe | | mae | |
| ほ | ほか | poka | | | | | foka |
| | かほ | kapo | | kawo | kao | | |

Glosses: え 'to get', えだ 'branch', ぬえ 'thrush', ゐ 'well', まゐる 'goes humbly', ゑみ 'smile', こゑ 'voice', をか 'hill', あを 'blue', はら 'stomach', かは 'river', ひら 'plain', かひ 'shell', ふね 'boat', たふ 'stupa', へ 'shovel', まへ 'front', ほか 'other place', かほ 'face'.
*Note that the dating of the *Taketori monogatari* is very tentative.

6.1.2.2), other *kana* letters than those discussed above have not changed in representational value since the beginning of the EMJ period and can be transcribed for premodern texts as they are for modern texts.

### *6.1.4 Sound tables*

Today the *kana* letters are not presented in the *Iroha* sequence, but arranged in the *gojūonzu*, as shown in Table 6.1 above. This arrangement of the *kana* letters is now used universally in school education and for dictionary entries; it also influenced the *katsuyōkei* system for describing verbal inflection (see 3.4.6). *Kana* lists (*onzu* 音図 'sound tables') which formed the basis of the current grid arrangement began appearing from the beginning of the eleventh century. The first known such *onzu*, which is incomplete consisting of only eight columns, appears at the back of the *Kujakukyō-ongi* (from around 1000) and the first complete *onzu* is found in Meigaku's *Han'on-sahō* (1093).

As opposed to the *Iroha*, the *onzu* went beyond simply enumerating the distinct orthographic categories: the *onzu* involve a substantial phonographic and phonological analysis and systematization of the letter categories defined in the *Iroha*. They provide a phonological analysis into consonants and vowels of the moras represented by the *kana* letters, such that *kana* in a column share the same initial consonant and *kana* in a row share the same vowel, i.e. an analysis of a syllable into subsyllabic parts, or a mora into submoraic parts. In the early *onzu* the *kana* were not actually arranged in grids with columns next to each other, but as lists with the columns following each other, but the principle is the same, of columns sharing an initial consonant and the vowels appearing in a fixed order within the columns. The grid system, incidentally, also makes it possible to refer in Japanese to consonants, which have no separate *kana* representation, with reference to this arrangement, e.g. the 'consonant of the *ka*-column' (*ka-gyō no shi'in* カ行の子音), viz. *k*. Ingenious as they were, the *onzu* remained in the realm of scholastic and academic writing and annotations until late in the Edo period; in the Meiji period the *gojūonzu* was adopted for educational purposes and eventually replaced the *Iroha* as the presentation of the *kana* categories and as an organizing principle for dictionaries. The first Japanese dictionary ordered in the *gojūonzu* sequence was the *Genkai* (言海) published in 1891 by Ōtsuki Fumihiko, who was also an important figure in the development of a 'standard' Japanese, see 13.2.2.

The phonological analysis of the *onzu* was principally inspired by Chinese rhyme tables and in particular by the *fănqiè* (反切) spelling system, in which the 'reading' of a *kanji* is described in terms of two other *kanji*, in a form like '*x yz*' to be understood as '*x* has the initial of *y* and the rhyme of *z*', for example, 東 德紅 which means that 東 has the initial of 德 (EMC *tək, LMC

\*təðk) and the rhyme of 紅 (EMC \*ɣəwŋ, LMC \*xɦəwŋ), and thus that 東 has the readings EMC \*təwŋ, LMC \*təwŋ.

However, knowledge of Indic script, again in particular the *Siddham* script (cf. 6.1.2.2), exerted an important influence on the development of the sound tables. The sequence of vowels (あ *a*, い *i*, う *u*, え *e*, お *o*) as well as consonants (か *k-*, さ *s-*, た *t-*, な *n-*, は *p-* (>*f-* > *h-*), ま *m-*, や *y-*, ら *r-*, わ *w-*) in the *gojūonzu* corresponds to that of the letters in the *Siddham* script. The letters representing true consonants come first, followed by approximants and the liquid: *y, r, w*. For the true consonants, this arrangement moves from velar to labial place of articulation, with nasals following oral consonants. The position of *s-* in the arrangement has been taken to support the view that early EMJ /s/ phonetically was not a simple sibilant (which in *Siddham* are placed at the end together with approximants and liquids), but rather an affricate (2.2.3); the arrangement also clearly shows that the reflex of OJ /p/ was an obstruent, most likely a stop, /p/, and certainly not an approximant such as bilabial [ɸ] (7.3.1.3). Between the earliest *onzu* from the beginning of the eleventh through the twelfth century, there was considerable variation in the order of vowels or consonants. The first *onzu*, from the *Kujakukyō-ongi*, has the vowels in the order *i, o, a, e, u* and the consonants *k, s, t, y, m, p, w, r*. The *onzu* in Meigaku's *Han'on-sahō* has the vowels in the current order and the consonants ordered strictly by place but not manner of articulation: *k, y, s, t, n, r, p, m, w*. Eventually the order we know today, which is first found in the *Shittan-yōjū-ki* (1075) by Kanchi (1045–1111), was settled upon.

The sound changes mentioned in (4) above also affected the phonological analysis and arrangement of the *kana* in the *onzu*. In particular, when the *kana* pairs おを, いゐ, and えゑ became phonographically equivalent, standing for /.o, .i, .e/, respectively, the historically correct assignment of these letters to the *a-gyō* or *wa-gyō* within the *onzu* became far from obvious, and until quite late there are many examples of *onzu* with incorrect assignments. In his 1682 *Shittansanmitsushō*, the monk Jōgon finally established い and ゐ in the *a-gyō* and *wa-gyō*, respectively, and Keichū (1640–1701) did the same for え and ゑ in 1693 in the *Wajinshōranshō*. It was not until 1776 that お and を were finally fixed in the *a-gyō* and *wa-gyō*, respectively, in Moto'ori Norinaga's *Jion kana-zukai*.

## 6.2    Sources

The sources available to study EMJ are far more extensive and varied than the sources for OJ. In particular, we have from the middle of the period a large set of prose texts, written in *kana* in a language form which is thought to be close to the vernacular of the nobility and officials at court and which gives a more comprehensive picture of the contemporary language than we have for any other time in premodern Japanese. Sources are scarce until the early tenth

century, and there is thus a substantial gap of more than one hundred years in the solid attestation of Japanese. This is usually attributed to the fact that written Chinese almost succeeded in establishing itself as *the* written language in Japan at the expense of Japanese, particularly outside poetry, but thankfully, from a linguistic and cultural point of view, that did not happen. Apart from the first one hundred years, the Heian period is characterized by a lively culture of writing in Japanese which produced some of the masterpieces of Japanese literature, such as the *Genji monogatari* and the *Makura no sōshi*. After around 1100 the language in the written sources increasingly fossilized and towards the end of the Heian period the written norm known as 'Classical Japanese' (*bungo* 文語 'written, literary language') became relatively fixed and served as the dominant base for writing in Japanese until the beginning of the twentieth century. The sources for this period, in addition to the prose and poetry texts (6.2.1), include a body of annotated Chinese texts, the so-called *kunten shiryō* (6.2.2), as well as dictionaries (6.2.3) and Sanskrit-studies (6.2.4). Table 6.4 is a chronological list of some of the important sources.

### 6.2.1    Prose and poetry

Prose texts are by far the most valuable material for studying the language of this period. There exist a number of poetry anthologies and many poems within prose texts from the Heian period, but while the literary merit of some of these poems is considered high, their value for linguistic studies is more limited: a norm concerning grammar and lexis soon became established in poetry, and most of the poetic texts are therefore conservative in these regards, for example avoiding both SJ vocabulary and native forms with bound moras (in particular *onbin* forms). In this respect the Heian period poetry is in stark contrast with the poetry from the Nara period, which constituted the major source of OJ. Among the poetry anthologies from the period, the first one, *Kokin wakashū* (古今和歌集 'Collection of Japanese poetry'), stands out, but in particular for its Japanese preface (*kanajo* 仮名序 '*kana*-preface') on poetics, written in *hiragana* by the editor Ki no Tsurayuki (?868–?945). This is the first piece of prose writing in *hiragana* and it contrasts with most later *hiragana* writing from the period by being expository and non-fictional.

The prose material includes *monogatari* ('stories') of various sorts, ranging from collections of short stories, or even what would today be called 'short-short stories', e.g. *Ise monogatari* (伊勢物語 'Tales from Ise'), to Murasaki Shikibu's very long novel *Genji monogatari* (源氏物語 'Tale of Genji'); other well-known works are *Taketori monogatari* (竹取物語 'The tale of the bamboocutter'),[3] *Utsuho monogatari* (宇津保物語 'The tale of the hollow tree'),

---

[3] *Taketori monogatari* is thought to date from the middle of the tenth century, but the earliest extant copies date from the Edo period, making it less reliable as linguistic evidence than the other EMJ texts.

Table 6.4 *Important EMJ sources*

Ninth century
> *Konkōmyō saishoō kyō* (annotations *c.* 830)
> *Tōdaiji fujumonkō* (before 834)
> *Zaitōki* (*c.* 842)
> *Shinsen jikyō* (898–901)

Tenth century
> *Kokin wakashū* (914)
> *Ise monogatari* (early tenth century)
> *Wamyō-ruiju-shō* (*c.* 934)
> *Tosa nikki* (935)
> *Taketori monogatari* (mid tenth century)
> *Kagerō nikki* (second half of tenth century)
> *Utsuho monogatari* (970s)
> *Ochikubo monogatari* (late tenth century)
> *Sanbō ekotoba* (984)

Eleventh century
> *Hokke mongu* (annotations *c.* 1000)
> *Makura no sōshi* (*c.* 1000)
> *Genji monogatari* (1001–10)
> *Tsutsumi chūnagon monogatari* (*c.* 1055)
> *Sarashina nikki* (1059–60)
> *Konkōmyō saishoō kyō ongi* (*c.* 1079)
> *Han'on sahō* (1093)
> *Ruiju-myōgi-shō* (*c.* 1100)

Twelfth century
> *Daijionji sanzōhōshi-den* (annotations 1099–1116)
> *Konjaku monogatari-shū* (*c.* 1120)
> *Shittan yōketsu* (1101)
> *Ōkagami* (*c.* 1119)
> *Iroha jiruishō* (compiled between 1144 and 1181)
> *Kohon setsuwa-shū* ((late?) twelfth century)
> *Shittan kuden* (*c.* 1180)

*Ochikubo monogatari* (落窪物語 'The tale of Ochikubo'), *Tsutsumi chūnagon monogatari* (堤中納言物語 'The tales of the Tsutsumi middle counsellor'), and the historical novel *Ōkagami* (大鏡 'The great mirror'). Other genres include *nikki* ('diaries'), which mostly were not private, but written for circulation, such as Ki no Tsurayuki's *Tosa nikki* (土佐日記 'The Tosa diary'), the *Kagerō nikki* (蜻蛉日記 'Gossamer diaries') or the *Sarashina nikki* (更級日記 'The Sarashina diary'); and *zuihitsu* (随筆 'jottings'), especially Sei Shōnagon's famous *Makura no sōshi* (枕草子 'The pillow book'). Much of this literature was written in *hiragana* and by women in a free-flowing, lively, elaborate and sometimes playful language.

The so-called *setsuwa* (説話 'tales, legends') literature comprises a variety of short tales ranging from didactic Buddhist stories set in India or China to local folktales and ghost stories. The *setsuwa* literature is generally written in *kanji-kana majiribun* (6.1.1), i.e. with a large proportion of *kanji*. The language in the *setsuwa* and other writing in *kanji-kana majiribun* is relatively simple and straightforward compared to the *hiragana* literature, but it is also more heavily and obviously influenced by Chinese, because of the influence of *kanbun-kundoku* (see 9.1) on *kanji-kana majiribun*. The most famous anthology is the *Konjaku monogatari-shū* (今昔物語集 'A collection of tales of times now past'); others from this period are the *Kohon setsuwa-shū* (古本説話集 'A collection of old tales') and *Sanbō ekotoba* (三宝絵詞 'Illustrated stories about the three jewels'). As the written language fossilized, the narrative was kept in the classical written language, but dialogue was in a form close to the vernacular. From the end of the period, and especially in the LMJ period, the dialogue passages in the *setsuwa* literature are important sources of changes in the language, as most other writing was done in the by then fossilized classical written language.

These types of prose texts together present a comprehensive view of the contemporary language and its use. As with materials from the Old Japanese period, the text portions written logographically are less useful for studying the lexicon and grammar of the language than those written in *kana*. However, because of the volume of text this is in practice not a severe problem for studying the grammar. The texts from this period have come down to us in later copies, subject to scribal error and to conservative and normative redaction, especially in applying the etymological spelling principle (6.1.3), making them unreliable as guides to the dating of sound changes. However, much effort has been invested in producing reliable critical editions of the texts. Most editions present the texts in etymological spelling and, for the texts written in *kana*, with many interpolated *kanji* (for the convenience of modern readers), but good editions will also indicate the written form of the base text.

### 6.2.2   Annotated texts

*Kunten shiryō* (訓点資料 'materials with reading marks') are texts in Classical Chinese with annotations which provide a guide to the pronunciation of Chinese or SJ words, or to understanding or rendering the texts into Japanese. See below (9.1) for details about *kunten*, *kuntengo* (the language used in *kunten* texts), and *kanbun-kundoku*, the practice of rendering Chinese into Japanese. A great amount of *kunten* texts exist, mostly unpublished and in the possession of Buddhist temples. A great effort of both fieldwork and philology has been and is being invested in gaining access to, charting, cataloguing and studying these materials. Impressive advances have been made in

recent years, but there is still enormous scope for further research, in addition to the sheer volume of material also because the material is complex and difficult to interpret. The study of *kunten* texts and *kuntengo* has become a major focus for research for Japanese linguists and philologists since the 1950s, more recently attracting overseas scholars as well. *Kunten* texts are important, primary sources. Often the date of annotation and the identity of the annotator are noted in the text. It is in particular thought that early *kunten* texts from the late eighth and ninth centuries reflect some form of contemporary vernacular language and that they thus can contribute to filling the gap left by the lack of other types of sources between late OJ and 900. Glosses provide valuable information about vocabulary not attested elsewhere and often provide an accurate indication of pronunciation and therefore of contemporary phonology, making them important material for dating sound changes, as opposed to the versions of the prose texts which we have today. It must be kept in mind, however, that *kunten* texts are annotations and therefore orthographically severely underspecified and on some points very difficult to interpret, so their evidential value is not straightforward. In addition, strict, dogmatic norms of rendition and annotation arose, so that *kunten* texts from after the middle of the EMJ period generally cannot be taken to reflect contemporary language, but especially in grammar and vocabulary had become conservative and even archaic. As mentioned above (6.1.2), the *kunten* material is also very important to the study of the history of the development of the *kana* scripts. The amount of published studies and material increases steadily, but well-publicized *kunten* texts include *Konkōmyō saishōō kyō* (金光明最勝王経 'Sutra of the golden light and most victorious king', annotated *c.* 830), the *Hokke mongu* (法華文句 'The sentences and phrases in the Lotus sutra', *c.* 1000, see 9.1.1), and the *Daijionji sanzōhōshi-den* (大慈恩寺三蔵法師伝, *Kōfukuji* manuscript, annotated 1099 and 1116).

### 6.2.3    Glossaries and dictionaries

Closely related to the *kunten* materials are so-called *ongi* ('glossaries', 音義 'sound-meaning') which are lists of words extracted from individual texts in Classical Chinese, explaining the pronunciation or meaning of words or characters and aiding the reading of individual texts. There are a number of *ongi* from the period. An important example is the *Konkōmyō saishōō kyō ongi* (Glossary for the 'Sutra of the golden light and most victorious king', 1079) which also contains the first attestation of the *gojūonzu* (6.1.4) and of the *iroha-uta* (6.1.2.4). Originating as combinations of glossaries for single texts, the period also saw the compilation of the first *dictionaries*, which were independent of individual texts. These include the *Shinsen jikyō* (新撰字鏡), *Wamyō-ruiju-shō* (倭名類聚抄 or *Wamyōshō* 倭名抄), *Ruiju-myōgi-shō* (類

聚名義抄), and the *Iroha jiruishō* (色葉字類抄 or 伊呂波字類抄), compiled in the second half of the twelfth century and the first dictionary ordered according to the sequence of letters in the *Iroha-uta*. Meigaku's *Han'on sahō* (反音作法) explains the pronunciation of *kanji* by the *fǎnqiè* spelling principle (and by *kana* glosses).

### 6.2.4    Sanskrit studies

The term *shittangaku* (悉曇学), which literally means the study of the Indic script *Siddham*, is used to refer to the study of Sanskrit and materials in Sanskrit related to Buddhism. It comprises a number of materials which are important to the study of Japanese, especially its phonetics, because they explain or exemplify features of Sanskrit by comparison with Japanese, although the interpretation of the material in many cases is far from straightforward. They include the *Zaitōki* (在唐記 'Record of a sojourn in China'), the priest Ennin's notes on the pronunciation of Sanskrit (which he studied in China) illustrated by means of sinograms used as *phonetic* (rather than phonemic) symbols, based on their sound values as *man'yōgana* or their Japano-Chinese readings (see 9.2.1), supplemented with comments; Meigaku's *Shittan yōketsu* (悉曇要訣 'Essentials of Sanskrit learning') which has many valuable observations on Japanese in order to illustrate points about Sanskrit pronunciation; and Shinren's *Shittan kuden* (悉曇口伝 'The secrets of Sanskrit learning') which exemplifies Sanskrit sounds by notes and by Japanese sounds.

### REFERENCES

Writing, general: Kōno 1969, Mabuchi 1993, Ohno 1977, Seeley 1991, Tsukishima 1977: 52–120, Tsukishima 1986. Runes: Elliott 1959: 14ff., 21ff.. *Hentaigana* for /-t/: Sugahara 2000: 23ff. Sources: Tsukishima 1987: 12ff. *Kunten* materials: Kasuga 1985, Nakada 1969, Nishizaki 1992, Tsukishima 1986, Yoshida *et al.* 2001. Dictionaries: Nishizaki 1995.