History of Chinese characters (汉字) and simplification

Standardization of the written system
In Japan there has been an effort on the part of the government to standardize language use, particularly kanji use. This is a timeline of those efforts:

- **1923**: Ministry of Education specified 1,962 kanji and 154 simplified characters for standard use. These were called tōyō kanji.
- **1931**: The former tōyō kanji list was revised and 1,858 characters were specified.
- **1942**: 1,134 characters were specified as standard tōyō kanji and 1,320 characters as sub-tōyō kanji were specified.
- **1946**: The Ministry removed 88 characters from the list but added 249 others, for a total of 1295 standard characters. The list was not adopted, however, until later in the same year when the number of characters was upped to 1850. 881 of those were expected to be learned by the end of elementary school (6th grade).
- **1981**: The list was revised to have 1,945 characters.
- **2009**: The list is expected to be revised to include an additional 186 characters

So, the number of kanji that one was supposed to learn in school was established as part of the government’s educational policy. However, there were also changes in the way characters were written. This is usually referred to as simplification.

Simplification of the written system: Kanji
Many kanji, prior to the end of the war, were written in their “complex” or “traditional” forms (if they had one). For example, 学 was written 學. As you can see, the “complex” form has more strokes. But for centuries, calligraphers had been abbreviating characters. So, if you look at a hand-written document from the early 20th century or before, it is not unusual to see the simplified form of the kanji. But, in print, the traditional form was always used.

After the war publishers decided that it was time to adopt the simplifications formally, and typeface was changed accordingly. More about the mechanics of simplification follows below.

Methods of Simplification
In principle, there are four methods of simplification.

1. **Calligraphic simplification**
   These simplifications are meant to reflect what “grass script” or cursive writing looks like. 
   Examples:

<table>
<thead>
<tr>
<th>学 (gaku)</th>
<th>觀 (kan)</th>
<th>転 (ten)</th>
<th>数 (sū)</th>
<th>党 (tō)</th>
</tr>
</thead>
</table>

2. **Partial simplification**
   In this case, only part of a complex character is used to represent the whole character. Examples:

<table>
<thead>
<tr>
<th>声 (koe)</th>
<th>獨 (doku)</th>
<th>医 (i)</th>
<th>県 (ken)</th>
<th>号 (gō)</th>
</tr>
</thead>
</table>
3. **Phonetic simplification**

In these simplifications, a component is substituted because it is pronounced the same as the original kanji (even though it doesn’t look the same, or have the same meaning)

<table>
<thead>
<tr>
<th>證 証 (shō)</th>
<th>囲 囲 (i)</th>
<th>缺 欠 (ketsu)</th>
<th>厅 庭 (chō)</th>
</tr>
</thead>
</table>

4. **Anomalous simplification**

These simplifications best simply memorized. Examples:

<table>
<thead>
<tr>
<th>體 体 (karada)</th>
<th>與 与 (ataeru)</th>
<th>竽 尽 (tsukusu)</th>
<th>龍 龍 (tatsu)</th>
</tr>
</thead>
</table>

**Simplification across lines**

There are some characters that are not simplified the same in all countries. Examples:

<table>
<thead>
<tr>
<th>Taiwan</th>
<th>傳</th>
<th>經</th>
<th>貿</th>
<th>関</th>
<th>芸</th>
</tr>
</thead>
<tbody>
<tr>
<td>Japan/Korea</td>
<td>伝</td>
<td>経</td>
<td>貿</td>
<td>関</td>
<td>芸</td>
</tr>
<tr>
<td>PRC</td>
<td>传</td>
<td>经</td>
<td>卖</td>
<td>关</td>
<td>芸</td>
</tr>
</tbody>
</table>

**Simplification of the written system: Kana**

Also immediately after the war, we see the implementation of the new orthography (現代仮名遣い) Almost all works printed AFTER 1946 use these systems. Works printed BEFORE 1946, sometimes even when reprinted, use the old system. The old orthography “spells” sounds differently than what you learned in Japanese 101. Here is a chart of the old orthography (From Nelson's *Japanese-English Character Dictionary*, Appendix 7):
Old-style kana.

There are a few kana that ceased to be used after the language reformation following the war. For our purposes, you should be familiar with the following:

ゐ (i), now written い. Used primarily in the auxilliary verb iru, which denotes continuous action. For example, tabete iru would be written 食べてゐる.

ゑ (e), now written え. Used in a variety of words.

を (o), now written お. While in modern Japanese を is used only for the particle that marks the direct object, in pre-war Japanese it can be used embedded in words, much like お.

ひ (i), now written い. When embedded in a word, ひ is usually pronounced i, not hi.

は (wa), now written わ. When embedded in a word, は is usually pronounced wa, not ha.

ふ (u), now written う. When embedded in a word, ふ is usually pronounced u, not fu.

There were no small kana in the old orthography. So, 買った appears as 買つた.

Finally, over time writers have moved away from using some kanji for common words like どこ, これ, or それ.

Here is a sample passage (a few sentences from Natsume Sōseki) written in first the complex kanji and old orthography and then the simplified kanji and new orthography:
うとうとしたと思ふうちに眼が覚めた。すると、隣の室で妙な音がする。始めは何の音とも又何處から来るとも判然した見当が付かなかったが、聞いてゐるうちに、段々耳の中へ纒まった観念が出来てきた。何でも山葵卸しで大根かなにかをごそごそ擦つてゐるに違ない。自分は確に左様だと思つた。夫にしても今頃何の必要があつて、隣りの室で大根卸を捱えてゐるのだと想像が付かない。

Even if you can’t understand the passage, you can see how the characters and orthography changed.

One more thing to keep in mind is that we take all these sets of characters for granted a solid, relatively unchanging group. But they’re not! For example, in the 18th century it was common to see texts printed (not hand written) with variant forms of kana called hentaigana. There is a link off of the index page to a chart of these, if you’re interested.

Why do hentaigana exist? Contrary to popular belief, the development of kana was not a neat and clean affair. Sure, they came from cursive forms of kanji used for their phonetic value back in the early days of ancient Japan. But it’s not as if there was a unified effort or authoritative voice in the process. Let’s start back at the beginning:

During the 6th-8th centuries (more or less) in Japan, people were using two writing systems: classical Chinese (hereafter referred to with the Japanese term, kanbun) and a curious mix of kanji used for their phonetic value and some for their semantic value. In particular, the poems of the poetry anthology Man’yōshū were written largely with kanji used for their phonetic value. It was a cumbersome system, and most certainly not standardized. Here is a sample:

[原文]箇毛與 美箇母乳 布久思毛與 美夫君志持 此岳爾 菜採須兒 家吉聞名 告<紗>根 虚見 津 山跡乃國者 押奈戶手 吾許薙居 師<吉>名倍手 吾己曽座 我<許>背歯 告目 家呼毛名雄母

The text above is how the first poem of the Man’yōshū appeared in the original. Incomprehensible, right? Below is the same poem in a more readable form—the kanji used for phonetic values have been changed to kana or kanji, as appropriate.

[訓読]箇もよ み箇持ち 堀申しよ み堀申し持ち この岡に 菜摘ます子 家聞かな 告らさね そらみつ 大和の国は おしなべて 我れこそ居れ しきなれて 我れこそ座せ 我れこそば 告ら め 家をも名をも

Finally, here is the same poem, entirely in kana:

[仮名]こもよ、みこもち、ふくしもよ、みぶくしもち、このをかに、なつますこ、いへきかな、のらさね、そらみつ、やまとのくには、おしなべて、われこそを、しきなれて、われこそませ、われこそば、のらめ、いへをもなをも
Let’s focus on one point: notice that the syllable *mo* can be represented by more than one *kanji*: sometimes it is 母 and sometimes it is 毛. And, it’s not just *mo*. Indeed, every syllable has multiple representative *kanji*. Why? Linguists theorize that Japanese was a more sound-rich language back then, and there was a distinct difference between one *mo* and the next. But, of course, we can’t know for certain because there are no sound recordings from that time. In the end, as you might guess, 毛 became the kana も and the kana that evolved from 母 (yes, there was one!) fell out of use. But, those kana that fell out of use took their time in doing so. Some hung out until the earliest 20th century, which is why we have *hentaigana*. Eventually, *hentaigana* stopped appearing in print except for trendy uses, such as on business signs.

The kanji used for phonetic value have a categorical name: *man’yōgana*, as in “kana used in the *Man’yōshū*”. It isn’t much a leap to recognize that some *ateji* are actually *man’yōgana*. The Wikipedia page in English on *man’yogana* is actually relatively informative (as of December, 2008).

**Modernization of the grammar: Classical Japanese to Vernacular Japanese**

So far we’ve been covering the writing system. But, there is another large linguistic change of which you should be aware: in the late 19th century and early 20th century a remarkable shift occurred. Japanese stopped writing with classical grammar and shifted to the vernacular. Let’s start again, at the beginning:

In the earliest days of written Japanese records, one had a choice of writing in 漢文 *kanbun* (more on that later) or in the grammar of the day (we’re talking 9th century or so). Back then, we assume, people wrote more or less the way they spoke. However, as time passed there became a disconnect between the written form of the language and the spoken form. Although both forms evolved, the did not do so together. By the end of the Tokugawa period, people were inflecting verbs and adjectives and adverbs in a different way from how they spoke. Here is a sample of a simple sentence, one in “classical” (the Japanese call it 古語 or 文語) grammar and one in the vernacular (what you might think of as “modern Japanese”):

<table>
<thead>
<tr>
<th>Classical Grammar</th>
<th>Vernacular Grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>いま は 昔、竹取の 翁 といふ もの ありけり。野山 にまじ りて、竹 を 取りつゝ、万 づ ち つ ことを 使 ひけり。</td>
<td>いま は もう 昔 の こと にな る が、竹取 の 翁 と い う 者 が あ った。 野や 山 に 分 て ん て 竹 を 取 り 竹 を 取 り し て は、い ろ ろ ろ の 物 を 作 る の に 使 っ て い た。</td>
</tr>
</tbody>
</table>

Not only has vocabulary changed, but also verbal inflection. For example, ありけり became いた. It is beyond the scope of this class to cover all of classical grammar, but you should know that when Japanese students study classical Japanese, their focus is on verbal, adjectival, and adverbial inflections, and they use the same paradigms as they do for modern language.

Back to *kanbun*, which is Classical Chinese as the Japanese wrote it. This really doesn’t have much to do with the topic of “simplification,” but we’ll throw it in here for lack of a better place. This language is an odd linguistic product. Don’t assume that a scholar of Chinese can read the *kanbun* texts of a Japanese scholar! Often there is enough linguistic drift to cause confusion. Although whole courses can be, and are, taught on *kanbun*, here are the few aspects of it that I want you to know for this course:
First, the usage of *kanbun* was largely reserved for official texts, such as government documents, and religious texts, such as sutra. Occasionally Japanese will compose Chinese poetry (漢詩), too. So, unless you end up in religious studies or historical studies at the graduate level, you’ll probably not need *kanbun*. That said, if you have Japanese friends who went to high school in Japan, they’ll undoubtedly have memories of sitting through *kanbun* class.

Given that Japanese and Chinese are such different languages, how do the Japanese manage *kanbun* verbally? Do they read it with the on-yomi? Not really. Traditionally, when reading a sentence in *kanbun* out loud, the reader will change the order of the sentence to match Japanese grammar, and when appropriate, change the reading from on to kun. In effect, it is like translating on the fly. Many texts are annotated to help the reader, giving him/her guidelines about where to jump next, etc. Here is the explanation from Wikipedia:

Inasmuch as Classical Chinese was originally unpunctuated, the *kanbun* tradition developed various conventional reading punctuation, diacritical, and syntactic markers.

- **kunten** (訓点, kunten² "explanation mark") "guiding marks for rendering Chinese into Japanese"
- **kundoku** (訓讀, kundoku² "explanation reading") "the Japanese reading/pronunciation of a kanji character"
- **kanbun kundoku** (漢文訓讀, kanbun kundoku² "Chinese writing Japanese reading") "a Japanese reading of a Chinese passage"
- **okototen** (乎古止点, okototen² "inflectional dot marks") "diacritical dots on characters to indicate Japanese grammatical inflections"
- **kutōten** (句読点, kutōten² "phrase reading marks") "punctuation marks (e.g., , comma and . period)"
- **kaeriten** (返り点, kaeriten² "return marker") "marks placed alongside characters indicating their Japanese ordering is to be 'returned' (read in reverse)"

*Kaeriten* grammatically transform Classical Chinese into Japanese word order. Two are syntactic symbols, the | *tatesen* (縦線, tatesen² "vertical bar") "linking mark" denotes phrases and the  뢿 *reten* (レ点, reten² "[katakana] re mark") denotes "return/reverse marks". The rest are *kanji* commonly used in numbering and ordering systems: 4 numerals *ichi* 一 "one", *ni* 二 "two", *san* 三 "three", and *yon* 四 "four"; 3 locatives *ue* 上 "top", *naka* 中 "middle", and *shita* 下 "bottom"; 4 Heavenly Stems *kinoe* 甲 "first", *kinoto* 乙 "second", *hinoe* 丙 "third", and *hinoto* 丁 "fourth"; and the 3 cosmological *sansai* (三才, sansai² "three worlds", see Wakan Sansai Zue) 天 "heaven", *chi* 地 "earth", and *jin* 人 "person". For written English, these *kaeriten* would correspond with 1, 2, 3; I, II, III; A, B, C, etc.

As an analogy for how *kanbun* numerically marks Chinese sentences with Subject Verb Object (SVO) word order into Japanese Subject Object Verb (SOV), John DeFrancis (1989:132) gives this English (another SVO language) literal translation of the Latin (another SOV) Commentarii de Bello Gallico opening.

<table>
<thead>
<tr>
<th>Gallia</th>
<th>est</th>
<th>omnis</th>
<th>divisa</th>
<th>in</th>
<th>partes</th>
<th>tres</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Gaul</td>
<td>is</td>
<td>all</td>
<td>divided</td>
<td>into parts</td>
<td>three</td>
<td></td>
</tr>
</tbody>
</table>