

Writing English with Chinese characters: an eccentric view of what the written Japanese language is like

Nathan Hagen, 2024-Sep-02

Utsunomiya University, 7-2-1 Yoto, Utsunomiya, Tochigi, Japan 321-8585

E-mail: nh@hagenlab.org

Abstract: We make an attempt at showing an English speaking audience what the Japanese language feels like, without having to teach Japanese. We can do this by showing how to write English using Chinese characters instead of, or mixed in with, alphabetic characters. Doing this requires a lot of adjustments and difficult choices — choices that Japanese scholars must have had to make when adopting the Chinese writing system for their native language.

Japanese is widely regarded as being a difficult language for native English speakers to learn. In fact, the U.S. Foreign Service Institute, in its published ranking of world languages by their estimated time to achieve proficiency, puts Japanese in a special category entirely on its own. While European languages such as Norwegian and Spanish are estimated to require a native English speaker about 24 weeks of study (at 25 hours of study per week), German is estimated to require 30 weeks, Greek, Russian, and Turkish about 44 weeks, and Arabic, Chinese and Korean about 88 weeks. Japanese, on the other hand, is listed as “>88 weeks”, but how much greater than 88? If my own personal experience is any guide, then something like 160 weeks would be close to the number. But I suspect that I’m somewhat slow, so if we pick somewhere in the middle then perhaps 120 weeks is a better estimate.

So, why is the Japanese language so hard to learn? The spoken language is actually not so difficult. It has a range of sounds that resembles those of Spanish, and has a very limited use of tones (pitch accents). The real difficulty with becoming proficient lies in the writing system. So, how did the Japanese language acquire such a complex writing system?

Two of the prominent writing systems that exist today — the Western alphabetic system and the Chinese logographic system — take very different approaches to communicating language, and although a significant fraction of modern Chinese learn some minimal form about alphabetic writing, few people in the Western cultural sphere learn anything at all about the Chinese writing system. One goal of this article is thus to introduce this new system to those who have never seen it before, and to give a feel for what it is like to use it. And all the while, we can still stick to using English!

While the early Chinese writing system was extremely influential and spread throughout much of East Asia, most regions outside of the modern borders of China have since reduced or eliminated their use of logograms and have switched to alternative writing systems. Japan, however, is a prominent exception (the

only other such exceptions being Taiwan and Singapore). Perhaps one reason that the Chinese system has been less widely adopted than the alphabetic system is that it places heavier demands on raw visual memorization. Another reason, though this reason is equally shared by the alphabetic writing system, is that the system evolved from a language in which it is well-adapted. If we try to insert it into a different language, then we find the need to make messy adaptations.

It is probably not a surprise, therefore, that anyone learning the Japanese written language quickly comes to feel that the Chinese writing system did not fit the Japanese language very well. The two are not closely related linguistically, and have different grammatical constructions. As a result, the process of adopting the Chinese writing system must have required a lot of adjustment. We shall explore just how this must have felt — how to adopt a writing system into a language that is not well suited to it, and how this effects the adopting language. While we will be doing this here for English, it gives a good feel for what the Japanese language feels like from the inside.

1 Writing with logograms

Since we will be discussing a lot about how words and spoken and written, it will help to start by introducing the characters for speak (言) and write (書). But what if we want to use past tense for these verbs? While we can now use the character ^{writ}書 in place of the English for the present tense, it might not be clear how we can make use of this for *wrote*. But this is mainly because the verb is irregular. If it were regular, then the past tense would be *written* — the same thing but with a *-ed* at the end of it. So, let’s write the past tense as ^{writ}書^{ed} by adjusting the characters attached the end, just as Japanese likely did for at least some of its verbs. Similarly, the regularized past tense of ^{speak}言 would be *spoken*, and therefore ^{writ}書^{ten}

as ^{speak}言^{ed}. Or, if we want to preserve the sound structure of the language at all costs, we could write this as ^{speak}言ⁱⁿ present tense but ^{spok}言^e in past tense. If we don't keep the hint on top of the character, then we just need to remember which pronunciation form to use in each context. Japanese definitely makes a lot of use of that irritating approach.

This kind of inflection of characters to produce contextual endings is actually something that English already does, in the way that we use numbers. If we see "2" we think "two", but "20" is now "twenty", and "2nd" is "second", as in "He came in 2nd place." [1]

Next, we define the symbol for a "character" (or "symbol") (字). Whichever one of these two words this ^{symbol}字 represents, we could leave entirely up to context (as is sometimes done in Japanese), but let's change the ending to indicate the choice we want, as in ^{charact}字^{er} versus ^{symb}字^{ol}. In order to talk about languages (語), if we also define England (英), Japan (和), and China (漢), we can ^{writ}書 about the corresponding ^{language}語^s as ^{Eng. lang.}英語 and the ^{Jap. lang.}和語. (I am abbreviating in order to make the hinting fit.) Or, to refer to an ^{Engl}英^{ish} or ^{Japan}和^{ese} person (人), we write ^{Chin. charact}英^人 and ^和人. By combining ^漢字^{ers}, we can quickly modify words. So, if we want to ^{talk}言 about an ^{Eng. lang.}英語 ^{person}言^{ing}人, then we can write ^{Eng. speaker}英語^言人. You can see how the hints are ^{flex}柔^{ible}, and this too follows a similar path to ^{Japanese}和語 usage.

We will be comparing things, and so we need to ^{symb}定^{ine} ^{irregular}字^{ols} for good (良), bad (惡), and not (不). These first two are ^{not-regular}不^常 ("irregular" = 不^常) adjectives, and so we need to decide what to do with "better" and "best". We can ^{writ}書^e these as ^{good}良^{er} and ^{good}良^{est}, and just remember to pronounce (言音: talk-sound) them as "bet-^{symb}ter" and "best". In other words, we decide that the ^{not}字^{ol} does ^{pronunc}不^{define} the ^{pronunc}言^音iation, but that the ^{Eng. lang.}言^音iation is determined by context of what word (句) it is being used in. This is an all-too-familiar choice for ^{Eng. lang.}英^語 users. So, we follow the same process for defining bad, worse, and worst: ^{bad}惡, ^{bad}惡^{er}, ^{bad}惡^{est}. We keep in mind that each ^{charact}字^{er} has a base ^{pronunc}言^音iation, and an ^{irregular}不^常 location-dependent ^{pronunciation}言^音. While one could ^{writ}書^e hints like ^{wor}惡^{se} and ^{wor}惡st, this becomes quite messy as our vocabulary expands. Over time, as we all get used to ^{writ}書^{ing} in the new system, we can expect that people (^{plural-person}等^人) will slowly adjust to modifying their ^{pronunc}言^音iation to "baddest" and throw away the current ^{irregular}不^常 form of "worst". The ^{writ}書^{ing} system is therefore nudging the ^{language}語 to become more ^{regular}常. Perhaps it is all for the ^{good}良^{er}.

In my experience, when the ^{conversat}話^{ion} with a ^{Jap. person}和^人 turns to the ^{Jap. lang.}和^語, they like to claim that ^{Japanese}和語 is ^{bett}良^{er} because it allows for more ^{compress}縮^{ed} ^{writ}書^{ing}. Naturally, ^{Eng. speakers}英語^人 find this ^{un}不^{convincing}. First of all, this should be an argument for ^{us}用^{ing} ^{Chinese}漢語, which is ^{entire}全^{ly} ^{logogram}漢^字s, rather than ^{Japanese}和語, which mixes ^{logogram}漢^字s with other ^{symb}字^{ols}, just like we are doing now. ^{And not}及^不 only that, but why should ^{compress}縮^{ion} be the criterion for the ^b良^{est} ^{writ}書^{ing} system?

One advantage of our new ^{writ}書^{ing} ^{system}系 is that whereas alphabetic (alphabet = ^{sound-element-character-list}音^素字^表) ^{English}英語 ^{writ}書^e both the ^{person and language}人^及 the ^{writ}語^{as} "Japanese", here we ^{writ}書^e them as ^{Japanese}和^人 or ^{和語}, such that the ^{writ}書^{ten} form clearly distinguishes between the two even though the ^{alphabet}音^素字^表ic form does not. This is a ^{common}常^{feature} of ^{Japanese}和語, though it can sometimes be ^{difficult}難^{to} remember which ^{symb}字^{ol} to ^{us}用^e when.

^{English}英語 is ^{know}知ⁿ for being ^{flex}柔^{ible} with its nouns (名詞), verbs (動詞), and ^{and}及^{jectives} (形容詞). When ^{talk}言^{ing} about a ^{person}人 who supports the ^{red}赤^{team}, we can call him "a ^{red}赤", ^{convert}換^{ing} the ^{noun}名詞 into an adjective. Similarly, something that is "wet" can ^{adjective}濕^{et} the tongue (形容詞 → 動詞), and if we ^{say}言 "a ^{wet}濡", we can immediately ^{imagin}想^e that it is perhaps it means a ^{wet}濡^{place} or maybe a ^{wet person}濡^人. The ^{noun}名詞 form is clearly marked for us by the indefinite particle "a" — one of the particle's main functions. We also often use stress ^{pronunc}言^音iation to differentiate between ^{noun}名詞s, ^{adjective}形容詞s, and ^{verb}動詞s. For example, we ^{writ}書^e ^{house}家 for both a ^{noun}名詞 (the place we live) and a ^{and}及^{動詞} (to provide shelter), but ^{pronunc}言^音 them as ^{language speakers}haus and ^{many other}haus. Perhaps this is why many Asian ^{language speakers}語言^人 (among ^{little}多^他s, I assume) feel annoyed by the seemingly-random (well, to them it seems so) use of definite and indefinite particles in ^{English}英語. ^{Chin. lang.}漢語-influenced ^{writ}書^{ing} ^{system}系s, however, have ^{little}小^{need} for this because it is easy to differentiate ^{word}句 forms by attaching ^{character}字^s. I ^{know}知 almost nothing of the ^{Chin. lang.}漢語, but have ^{read}說 that ^{Chinese}漢語 can simply add ^{not}不 to negate something, use ^{little}過 to imply too much, or ^{too-many new charact}少 to imply too little. [2] If you think I am defining ^{too-quick}過^多 ^{new charact}新^字ers ^{too-quick}過^早ly, then you now ^{know}知 how to communicate that.

If we are feeling ^{enthusias}熱^{tic} here, we can ^{add}增^{some} more ^{charact}字^{ers} to reduce our dependence on the ^{alphabet}音^素字^表. We can define an "adjectivizing/adverbializing" ^{symb}字^{ol} 的, such

that strong (強) becomes 強的, giving us that -ly ending. Thus 音素字表 becomes 音素字表的, fortunate (運) becomes 不運及不運的. No more messy endings! Among the population (人數) of 英語言人, there may be some 少熱心 holdouts who will cling to the use of 音素字表的 endings, though. But an enthusiastic academic may want to charge through regardless of the consequences. This 新系 is giving us such 強! So we also 增 a “verbalizer” 字 of that 換 a 名詞 into a 動詞: 為. 日本語 itself does not make 多 of use of this, since it is already so 柔ible with 換 ing 名詞 into 動詞. But, hey, we are aiming for 縮 ion, right?

Someone resisting the 新 language will 泣外: this is going 過遠. These 新字 are complicated and take 過長 to 書. Just take a look at the 字 er we use for “use”. Just 書 ing the 音素字表的 字 ers u 及 s, they continue, is far 早 er than writing 用. Enthusiasts would reply, however, that the 實 強 here is the 縮 ion: we are replacing 二 字 ers with only 一.

2 Names (名 s) in the new system

One place where 漢語 字 ers do a 惡 job is dealing with foreign (外國: outside-country) 語 名 es. While you can get away with 名 ing cities, say, as northern capital (北京) or river metropolis (川都) in your own 國 y, continuing in this vein for 外 國 ies 早 ly becomes confusing. We 要 a way of 用 ing our 字 ter system to describe 名 es in some way that reflects what 外國人 s call them. If some far off 國 calls their 京 都 some strange 名 e like “London”, then how shall we deal with that?

In 和, their choice was extremely practical. They adopted some of the practices of 漢 — this was, after all, where they learned how to 書 e in the first place — but mostly they just threw up their 手 s and decided 不 用 the 漢語 字 ers for this at all. They came up with a native syllabary, something much like an 音素字表, and to 書 e 名 es just like they are 言 音 ed. Or, well, how they are 言 音 ed by 和 語 人, of course.

From what I 分, the 漢語 人 went 全-in. They have these 字 ers for 書 ing, and that’s 全 they have to work with. So they made the only choices available 下 those restrictions: they either invent 字 ols for each and every

new location, or they 用 e some of their 字 ols 全 the for their pronunc and 不 for their meaning. From what I can tell, both are involved: Vietnam = 越, United States = 米, France = 仏, Germany = 独, Iceland = 冰, Ireland = 愛, Norway = 諾, etc. If we did this in 英語 using our Latin (拉語) 及 Greek (希語) roots, then a 國 y called “Myrme- donia” would be 書 ten like myrme-dom-ia (ant-house-country). We just realize from context that the 句 is not referring to ants (*myrmex-*) and houses (*domus*) but rather to a 國 y.

3 Forming compound words (連句)

A feature of the 英語 that native 言人 s make 常 用 e of is the 多 數 of 拉語 及 希語 句 素 s that are 早 的 com- bined to create 連 句 s. Thus, if we were to 用 e the fa- miliar prefix pyro- 及 suffix -phobia, we create the word *pyrophobia* whose meaning is 早 的 分 ed even if the 讀 人 has never 見 n it before. More unusual 連 句 s, with much narrower meanings are also possible. Thus, even if we were to 見 the unfamiliar 句 *peripyrchromatic* we could already guess that this seems to refer to the color (色) at the edge (境) of a fire (火) — a guess we could easily improve on if the 書 er supplies the proper context. Maybe the 書 人 is some researcher who 学 ies 火, and who is trying to point out some feature of the color 色 ation at its 境 (which is a phenomenon that actually is 学 ied, by the way).

On reflection, we can 見 that the behavior of these 句 素 s, largely borrowed from 拉語 及 希語, perform a func- tion in 英語 which is quite similar to how the 漢語 字 ers are used in Japanese: since each 漢語 字 er (in principle, if not in actual practice) represents a concept, placing multiple 字 ers together generates a 連 句. Thus *peripy- rochromatic* would be 書 ten as 火 境 色 的 (fire-boundary-color-related).

We can also consider, however, that 不 全 句 素 are from 拉語 及 希語. There are a 大 數 of Anglo-Saxon roots as well (burn [燒], king [王], land [土]), though native 言人 s will generally 考 of them as full 句 s rather than word roots (sunburn [日燒], heartburn [心燒], eye- burn [目燒]; kingmaker [王作人], unkingly [不王的]; overland, landlocked). These are often so 早 的 stuck together, that it may be 不 透 whether one should 用 e a hyphen, as in double-cross or doublecross, half-wit or halfwit. Some native 英語 言人 s will come to 見

some as 句 roots if they happen to 知 something of another 独的語: (for-bid) (hus-band), (kinder-garten) (pumper-nickel).

As a result of its total fluency with creating compound words, 漢語 and Japanese speakers are often puzzled by the 英語言人's assertion about the 大数 of 句s in the 英語 vocabulary. But we can 始 to 見 the 元 of the confusion: if we can invent 句s like 不 王 的 語 縮 — and actually be 分 ed — then counting 句s starts to lose meaning. If we have, say, 4000 different characters available to choose from, and allow 句s of up to 5 characters long (a number which is much less than those languages limit themselves in practice) then we come up with $4000^5 \approx 10^{18}$. While this is something of an overestimate (we would not make any sense of 句s like 不不不不不) it is still so far beyond the number of 句s 英語 provides that the number of 句s in 英語 would seem close to zero in this context. Of course, the truth is that the comparison is quite unfair, since 英語 speakers count 句s from a dictionary — where one would search in vain for comprehensible (well, to some) 句s such as the 上 “peripyrchromatic”. If we want a fairer comparison, we would allow free use of the Latin and Greek word elements 拉語及希語句素 (should we call that a word)? If we know that *myrmex*- is the 拉語 root for *ants*, then we 知 too that a myrmecophiliologist is a 人 who is scared of 人人 that study ants.

4 Abbreviations in the logogram (漢字) system

A 常 complaint one hears (聞) about modern 英語 is about its profusion of acronyms. 人人 have learned the utility of 短ing 不要的 長 phrasing that AFAICT doesn't really disrupt the overall 語 much, IMHO, because acronyms are generally awkward to 言 音. They stand out as different from 常 句s. This 限 s their diffusion into the 語. Occasionally, however, one encounters an acronym that is so 良, and so 用 效, that it 換 s into a word of its own. 一 example is LASER, which also spawned a related 動詞 (“to lase”).

Since the 中 語 字 ers adopted directly from 漢, on the other 手, are so easy to clump together, one finds it very easy to invent 句s on the fly, or to throw together strings of 字 ers to create 長 句s, much in the way that

German 独語 is famous for doing among the 歐 語 s. 独語 is an “agglutinative” 語. However, much as we showed 上 with 拉語及希語句素, once the 句s become unwieldy then 人人 will 見 for a shortcut. In 和, these shortcuts take the form of removing 字 ers from a 長 句 — choosing only 二 to 四 字 ers that are distinctive (do 不 show 上 in other 句 s, or with a 言 音 iation that may be confused with some other 句 in the current context) and retain some of the feeling of the original 長 form. While this 音 s fine, its proliferation is at a scale far beyond that of acronyms in 英語, and there is no easy way to distinguish between the original forms of the 句s and their 短 ed forms. Eventually, 人人 will come to forget the original forms and remember only the 短 ion. But the problem with this is that we are left with a collection of 字 ers whose individual meanings are no 長 er 透的 connected to the 上全 meaning of the 句.

For example, in 英語 we 短 the United Nations to U.N., making sure to 用 e periods in order to indicate the 短 ion and not the 句 素 “un-”. In 和語, this is 國際連合 (international-connect-fit), which gets 短 ed to 国連, which literally means nation-connect. We lose a bit of the original sense there, but not 過 惡. However, more specialized 語 tends to be more 短 ed. At universities, a department of Applied Physics would be 応用物理学 (apply-use-physics) but usually 短 ed as 応物 (apply-thing), which 不 長 er makes sense by itself. Likewise, Applied Chemistry (应用化学, apply-use-chemistry) becomes 応化 (apply-change). Someone unfamiliar with these can easily mistake them as words rather than 短 ions, but if the sentence were something like “I am going to 言 to the head of apply-change” or “Apply-change is restructuring”, you might be forgiven for 考 ing you mis-聞 d what I said.

Over time, you can 想 e that such 短 ions take on a life of their own and become 句s. Indeed, the 句 for electronic calculator is the unwieldy 「電氣式卓上計算機」 (electric-style table-top calculation-machine) which 分 ably gets 短 ed, but becomes 「電卓」 (electric-table) which makes no sense whatever as a 句 that should be built from two concepts. Or how about the 短 ion 「国保連」 (nation-protect-connect) for the annoyingly 長 句 「国民健康保険団体連合会」, but which refers to an insurance 系 for national government employees. You can 見

how, if this got 外 of 手, it 始 to cause havoc with the logic of the 語. And so it does.


But the 惡dest perpetrators of these 語 atrocities are unquestionably the journalists, who go 全外 to 縮 their 書ing into as 小 a space as possible, to an extent that would likely make their 英 語 counterparts envious. For a 外国 讀人, this makes 讀ing newspaper articles quite a chore.

5 Going all out on compression

You can 見 that as we 增 more 及 more 字ers to our 系, we can reduce the 数 of 字ers we use for 書ing. Once we get 用ed to it, and we no longer 要 the hinting 上 the 字er, then maybe this will 短en the 数 of pages that books 要. However, with a 小 熱 m, we can milk this 書ing 縮 ion 物 even 長er. By condensing it down, we can create the *dulce leche* of 書 系s. And taking it 全 the way to its extreme, we can create the 白 矮 星 of 書ing systems. Unfortunately, we will see that 食ing 白 矮 星 diamonds does not make for a 大良 meal.

Since we are going to push way beyond what the ancient Chinese ever envisioned, we are now jettisoning their characters and building our own. We moderns walk around all the time with computers in our pockets, and no longer have any need to put pen to paper for writing, we can go far beyond what paper-based writing could ever achieve. Now, if we are aiming for making symbols that compress information down into a single symbol, well, what kind of symbol are we talking about? Any word, after all can be regarded as a symbol — and we can put a box around each word to reinforce the concept. These symbols might not all be the same size, but us alphabet users are already used to that — after all, an “i” is much narrower than a “w”.

So, if we think of each word as being an individual symbol, then the “compress individual concepts down into individual symbols” gang in East Asia should now see that we’re not so different after all. We just don’t fit our symbols into squares, and alphabet-derived symbols are much more flexible, in addition to giving strong hints to their pronunciation, so that the memorization requirements are much relaxed. Advanced technology, that. But if compression is what you want, compression you shall have. After all, why stop at a word as an individual symbol? What about a sentence? A paragraph? Even better, a page? Each page has the same shape,

right? So, if we want uniformly-shaped symbols, then we can take a page of text and, well, that’s our symbol. Is the symbol too big? No problem. We have computers — just tell your computer display to shrink it down to fit into whatever space you like. In fact, here are the pages of this article: . Five pages, five symbols. It’s all there — just zoom right in. And now our writing system has achieved a wonderful compression. One book now fits onto a single page. An entire library onto the pages of one book. But then again, we could just digitize it, converting text to a much more efficient format.

Now we have come full circle. Once we see words as individual symbols, we alphabetors can say that we have way more symbols than the logogram users do, and therefore more possible combinations. If they could achieve $4000^5 \approx 10^{18}$ combinations, the English dictionary with its million words, taken five at a time, can do $10^{6^5} = 10^{30}$ which is so many more than they can achieve that it makes their vocabulary look like zero. They reply, however, that this isn’t fair. Alphabetic words contain more than single logogram symbols do, so we should be comparing their “words” to our words, and so their vocabulary would then leap up to $4000^{5^5} \approx 10^{90}$. We counter, however, that 25 of their symbols is like a sentence in our alphabetic writing, and so that’s not a fair comparison. For a full sentence of 20 words, we should allow something like $10^{6^{5^{20}}} = 10^{600}$. Hey, that’s not fair, they retort. They have sentences too.... Through this mock debate, the pointlessness in the compressibility argument becomes clear.

6 Conclusion

The advantages and disadvantages of adapting the Chinese characters to writing English are evident. The trouble with memorizing all of the many various shapes. The inconsistency in the “hinting”, and the fact that different people can potentially pronounce the same character differently depending on their interpretation of the context. The need for a secondary writing system to allow verbs and adjectives to inflect. The fluidity with word-boundaries and the ease of creating new words. All of these are likewise features of the Japanese writing system. And many of the difficult decisions that we have seen would go into adapting the Chinese writing system to English likely reflect what early Japanese scholars had to do.

Finally, we should make sure not to go too far with showing the difficulties of using Chinese script, since similar adjustments are also required for languages adopting the alphabetic system. We can see this, for example, in the large number of diacritics used in modern Vietnamese writing, in order to force it to express the extra vowel sounds and tonal structures that are not present in the circle of Latin-based languages.

A Postscript

One inspiration for writing this article was a blog post “If English was written like Chinese” by Mark Rosenfelder. Rosenfelder’s post, however, focuses mainly on character creation, rather than adoption of writing from Chinese. So his approach show something more like what the English language writing system would have looked like if we had started out like the Egyptians did — from pictographics that then could have been simplified over time into logograms, as the Chinese system did (and as the demotic script evolved from the ancient Egyptian hieroglyphs, and bits of which survive even into modern times inside the Coptic script).

References

- [1] . This example closely follows that posed by a comment on Reddit. Unfortunately, I can no longer locate it to credit its author.
- [2] . This example is adapted (and closely follows) that from Tae Kim’s blog article “Which is harder? Japanese or Chinese?”, located at: <https://www.guidetojapanese.org/blog/2006/07/20/which-is-harder-japanese-or-chinese/>