L2/25-213

Title: Katakana *kun'yomi* in Unihan Database `kJapanese`
Date: 2025-08-05
Source: HarJIT *a.k.a.* Harriet Riddle
Action: For consideration by CJK WG

1. Background

The documentation for the provisional `kJapanese` Unihan database property states the following:

> The Japanese readings(s) for this ideograph expressed in Kana. Readings expressed in Hiragana are generally considered Kun-yomi (訓読み), and readings expressed in Katakana are generally considered On-yomi (音読み).

This is true in the most common cases, but some exceptions exist.

Sino-Japanese morphemes may be written in hiragana but, in a dictionary context, are often written in katakana to differentiate them from Japonic morphemes. As such, *on'yomi* readings listed in the `kJapanese` property are indeed listed in katakana.

This document reports the result of analysing exceptions to the other rule of thumb; that is to say, `kJapanese` property values where a *kun'yomi* appears to be included in katakana.

2. Detection

Although there are multiple phases of *on'yomi* and some irregular forms, they, when expressed in kana, almost all follow the following relatively constrained structure (see Myers):

- A full-sized kana.
- Optionally, a small kana *ya*, *yu* or *yo*.
- Optionally, a full-sized *i*, *u*, *ki*, *ku*, *chi*, *tsu* or *n*.

This is not a hard-and-fast rule (some common *kanji* have curriculum-standard, albeit irregularly-formed, *on'yomi* ending in a small *tsu*, for example), but applies for the majority of distinct *kanji* characters.

Comparing with the `kJapaneseKun` property value is also often illuminating. If a reading listed in katakana in the `kJapanese` property is listed in the `kJapaneseKun` property for the same codepoint, this suggests that that reading might be a *kun'yomi*.

3. Explanations for some of these cases

3.1. Japonic terms in katakana; taxonomic terms

By convention, a Japonic morpheme, if written in kana, is typically written in hiragana, but not always, and with a number of exceptions. For example, it is not uncommon for vernacular taxonomic terms to be written in katakana.

3.2. Characters used for writing non-Sino-Japanese loanwords

Characters with non-Japonic *kun'yomi* readings are a special case: *kun'yomi* are Japonic morphemes (*wago*) in the majority of cases, but not necessarily, particularly in the case of characters referring to units of measure originating outside of East Asia. Such non-Japonic, non-Sino-Japanese morphemes are seldom if ever written in hiragana, and the `kJapanese` property values accordingly show them in katakana.

An example may be in order. `U+7CCE` 糎 is a CJKUI representing the metric unit of "centimetre" (or, confusingly but more rarely, "millimetre"). Its readings currently listed in the Unihan database are as follows:

```
U+7CCE     kCantonese lei4
U+7CCE     kDefinition     (J) centimetre
U+7CCE     kJapanese  センチ センチメートル
U+7CCE     kJapaneseKun    SENCHIMEETO
U+7CCE     kMandarin  lí
```

Figure 1: Unihan database "Readings" category for U+7CCE

There is no established *on'yomi*, since the character was coined in Japan, and is not used in compounds; the Mandarin and Cantonese readings are (and `U+7CCE`'s purely theoretical *on'yomi* would be) synchronised with `U+5398` 厘, thus effectively making `U+7CCE`'s semantics in those languages equivalent to *"centi" in the sense of an abbreviation of "centimetre"*. Similarly, the *kun'yomi* reading センチ (*senchi*) is a contraction of the other *kun'yomi* センチメートル (*senchimētoru*); the `kJapaneseKun` value gives センチメート (*senchimēto*) as another such *kun'yomi*; all ultimately derive from French *centimetre* and, as such, are written in katakana as non-Japonic morphemes.

As such, both `kJapanese` property values are given in katakana, but both are *kun'yomi*; indeed, even without knowledge of their etymology, they can be detected as, at least, not being regularly-formed *on'yomi* due to having three or more full-sized kana.

4. Occurrence

4.1. Katakana readings which do not match the regular *on'yomi* structure

In total, some 136 kJapanese values across 127 codepoints are written in katakana but do not conform to the structure for a regularly-formed *on'yomi* as given by the following regular expression:

```
^[アイウエオ-モヤユヨ-ロワ][ャュョ]?[イウキクチツン]?$
```

Some of these are phonetically irregular *on'yomi*; some are Japonic *kun'yomi* written in katakana for some other reason (e.g. nouns which can be used as taxonomic vernacular names). Some may follow archaic orthography, and some may be errors. However, a majority (roughly 75 codepoints) of these are non-Japonic (loanword) *kun'yomi*.

Table 1: Katakana kJapanese values which do not conform to the usual on'yomi structure

Codepoint		kJapanese value	O/K	Comments
U+34DD	荆	ケィ	O	ケイ
U+3B8B	榔	イタ	K	(taxonomic)
U+3BC9	樗	クワ	K	(taxonomic)
U+41C6	斗	デカリットル	K	Decalitre
U+41C9	竓	ヘクトリットル	K	Hectolitre
U+45B3	蛇	タタ	K	(taxonomic)
U+4787	贊	ニエル	K	?
U+495B	鏍	ニッケル	K	Nickel (=鎳)
U+4ED9	仙	セント	K	Cent (contrast U+8056 聖)
U+4EEB	仫	ムー	O	Mulao (ethnonym)
U+509D	傽	タフ	O	(irregular *on'yomi*)
U+5159	兙	デカグラム	K	Dekagram
U+515B	兛	キログラム	K	Kilogram

Codepoint		kJapanese value	O/K	Comments
U+515D	勉	デシグラム	K	Decigram
U+515E	匙	ミリグラム	K	Milligram
U+5161	匁	ヘクトグラム	K	Hectogram
U+5163	匣	センチグラム	K	Centigram
U+51C3	凃	トオ	O	Tu (a Chinese surname)
U+5337	匷	キヨ	O	キョ
U+5341	十	ジッ	O	(modified *on'yomi*)
U+5408	合	ガッ	O	(irregular *on'yomi*)
U+5408	合	カッ	O	(irregular *on'yomi*)
U+540B	吋	インチ	K	Inch
U+542C	听	ポンド	K	Pound
U+544E	呎	フィート	K	Feet
U+544F	呏	ガロン	K	Gallon
U+5491	呑	ダース	?	(musical instrument)「呑噂」???
U+54E9	哩	マイル	K	Mile
U+55B5	喵	ミャオ	O	Meow
U+55E7	嗀	ガロン	K	Gallon
U+561D	噝	ブッシェル	K	Bushel
U+565A	噚	ファゾム	K	Fathom
U+56CD	囍	ヨシ	K	(unclear why this isn't in hiragana)
U+574A	坊	ボッ	O	(modified *on'yomi*)
U+5A47	婇	ウネメ	K	(unclear why this isn't in hiragana)
U+5A47	婇	ウネベ	K	(unclear why this isn't in hiragana)

Codepoint		kJapanese value	O/K	Comments
U+5BEA	寪	イギ	O	イ *or* ギ (two readings)
U+5C01	封	ポンド	K	Pound
U+5CC5	峅	クラ	K	(unclear why this isn't in hiragana)
U+5E0B	帋	カミ	K	(unclear why this isn't in hiragana)
U+5EFF	廿	ニユウ	O	ニュウ
U+5F17	弗	ドル	K	Dollar
U+6070	恰	カッ	O	(irregular *on'yomi*)
U+62C6	拆	チヤク	O	チャク
U+65E9	早	サッ	O	(irregular *on'yomi*)
U+69CF	槚	カネ	K	(unclear why this isn't in hiragana)
U+6BF9	毹	シム	?	?
U+6CD5	法	ハッ	O	(irregular *on'yomi*)
U+6CD5	法	ホッ	O	(irregular *on'yomi*)
U+6CD5	法	フラン	K	Franc
U+6D6C	浬	ノット	K	Knot (unit)
U+7339	猹	チャー	O	(Chinese-language *hapax legomenon*)
U+73CA	珊	サンチ	?	(taxonomic); サン, サチ
U+74E6	瓦	グラム	K	Gram
U+74E7	瓧	デカグラム	K	Dekagram
U+74E9	瓩	キログラム	K	Kilogram
U+74F0	瓰	デシグラム	K	Decigram
U+74F1	瓱	ミリグラム	K	Milligram
U+74F8	瓸	ヘクトグラム	K	Hectogram

Codepoint		kJapanese value	O/K	Comments
U+7505	瘭	センチグラム	K	Centigram
U+78B5	磧	ヒロ	K	(unclear why this isn't in hiragana)
U+78C5	磅	ポンド	K	Pound
U+7ACB	立	リットル	K	Litre
U+7ACD	竍	デカリットル	K	Decalitre
U+7ACF	竏	キロリットル	K	Kilolitre
U+7AD3	竓	ミリリットル	K	Millilitre
U+7AD4	竔	リットル	K	Litre
U+7AD5	竕	デシリットル	K	Decilitre
U+7AE1	竡	ヘクトリットル	K	Hectolitre
U+7AF0	竰	センチリットル	K	Centilitre
U+7C17	簗	ヤナセ	K	(unclear why this isn't in hiragana)
U+7C73	米	メートル	K	Metre
U+7C75	籵	デカメートル	K	Decametre
U+7C7F	籿	デシメートル	K	Decimetre
U+7C80	粀	デカメートル	K	Decametre
U+7C81	粁	キロメートル	K	Kilometre
U+7C8C	粌	ヘクトメートル	K	Hectometre
U+7C8D	粍	ミリ	K	Milli(metre)
U+7C8D	粍	ミリメートル	K	Millimetre
U+7C8E	粎	メートル	K	Metre
U+7CA8	粨	ヘクトメートル	K	Hectometre
U+7CB4	粴	キロメートル	K	Kilometre

Codepoint		kJapanese value	O/K	Comments
U+7CCE	糎	センチ	K	Centi(metre)
U+7CCE	糎	センチメートル	K	Centimetre
U+7D0D	納	ナッ	O	(irregular *on'yomi*)
U+7D4B	紘	ヒロシ	K	(unclear why this isn't in hiragana)
U+7D4B	紘	ヒロム	K	(unclear why this isn't in hiragana)
U+7D4B	紘	ヒロ	K	(unclear why this isn't in hiragana)
U+8056	聖	セント	K	Saint (contrast U+4ED9 仙)
U+87CE	蟎	ダニ	K	(taxonomic)
U+933B	錻	ブリキ	K	*blik* (Dutch; sheet metal)
U+9801	頁	ページ	K	Page
U+99D8	駘	ダデ	?	?
U+203F9	𠏹	ホトケ	K	(unclear why this isn't in hiragana)
U+20490	𠒐	ミリグラム	K	Milligram
U+20498	𠒘	ヘクトグラム	K	Hectogram
U+20499	𠒙	キログラム	K	Kilogram
U+204AD	𠒭	ミリグラム	K	Milligram
U+204B2	𠒲	デカグラム	K	Dekagram
U+204C8	𠓈	ミリヤグラム	K	Myriagram (i.e. 10^4 grams)
U+204CF	𠓏	クウイントル	K	Quintal (i.e. hundredweight)
U+204CF	𠓏	キンタル	K	Quintal
U+20959	𠥙	シヨ	O?	ショ?
U+21A5B	𡩛	ヘクタール	K	Hectare
U+21AB8	𡪸	センチアール	K	Centiare

Codepoint		kJapanese value	O/K	Comments
U+21B5A	𡭚	マバ	O	マ *or* バ (=麼)
U+21CB2	𡲲	クワ	O	クヮ (=靴)
U+237FF	𢟿	ナギ	K	(taxonomic; =梛)
U+247F1	𤏱	タジヒ	K	(taxonomic; =蝮)
U+24CBE	𤲾	アール	K	Are (unit)
U+25A58	𥩘	センチリットル	K	Centilitre
U+25A7B	𥩻	デシリットル	K	Decilitre
U+25A95	𥪕	キロリットル	K	Kilolitre
U+25AB3	𥪳	ミリリットル	K	Millilitre
U+25E2F	𥸯	ミリアメートル	K	Myriametre (i.e. 10^4 metres)
U+26B22	𦬢	ボダイ	O/K	Bodhi (=*on'yomi* of 菩提)
U+26B35	𦬵	ボダイ	O/K	Bodhi (=*on'yomi* of 菩提)
U+29E2D	𩸭	ライヒー	O/K	(from *on'yomi* of 雷魚)
U+2A602	𪘂	ゲッ	O	ゲツ
U+2B679	𫙹	ブリザード	K	Blizzard
U+2B78B	𫞋	シヨウ	O	ショウ
U+2B7BF	𫞿	ヤナセ	K	(unclear why this isn't in hiragana)
U+2B81D	𫠝	ミノガメ	K	(taxonomic)
U+2B877	𫡷	ヘクタール	K	Hectare
U+2CABF	𬪿	センチアール	K	Centiare
U+2CF01	𬼁	ダラー	K	Dram
U+2CF04	𬼄	オンス	K	Ounce
U+2CF17	𬼗	グレーン	K	Grain (unit)

Codepoint		kJapanese value	O/K	Comments
U+2D160	亢	ヨシ	K	(unclear why this isn't in hiragana)
U+2D260	壽	シユ	O	シュウ、ジュ (=寿)
U+2D505	竡	ヘクタール	K	Hectare
U+2D516	竰	センチアール	K	Centiare
U+2DECC	畱	リユウ	O	リュウ
U+2E0F0	穌	クワ	O	クヮ (=和)
U+2E215	縫	シヨウ	O	ショウ
U+2E494	繭	ツタ	K	(taxonomic; =蔦)

4.2. Other places where katakana in kJapanese matches kJapaneseKun

Two additional entries are worth highlighting. U+74F2 甴 and U+8422 蒢 both have no kJapaneseOn property, and have a kJapanese property which is entirely in katakana and overlaps the kJapaneseKun property. In the case of U+74F2 甴, the reading (トン) is a loanword for a unit of measure (ton).

In the case of U+8422 蒢, the reading (ヤチ) is also in fact a *kun'yomi* despite being in katakana (for unclear reasons): it is not cognate to the kMandarin or kCantonese readings, and matches that of the variant U+2BBA4 𦰡, which is given in hiragana (やち) in its kJapanese property; the other readings (ヤツ (やつ) in kJapanese and YAJI (やぢ, やじ) in kJapaneseKun) are related to やち, and similarly unrelated to the kMandarin and kCantonese readings. Since its IRG sources are limited to J0 and GE (i.e. JIS-X-0208 and GB/T-16500, the latter of which derives from Unicode 1.1), its usage seems to be specific to Japan; hence, the absence of an *on'yomi* is to be expected.

U+5A42, U+5FA4, U+7E90, U+84D9, and probably also U+6FF9 and U+7874, none of which has a kJapaneseOn property, all appear to include one or more *on'yomi* (cognate to readings in their kMandarin or kCantonese properties) in their kJapaneseKun properties; their kJapanese properties correctly use katakana for these pronunciations. What, if anything, to do about these cases is beyond the scope of this document.

5. Suggestions

For consistency, it seems sensible for *kun'yomi* taxonomic vernacular terms to be recorded in hiragana; the few noted above are the exception. For other readings which are unambiguously *kun'yomi* but in katakana for unclear reasons, it seems better to change them to hiragana to differentiate them from *on'yomi*—not counting those which represent (non-Sino-Japanese) loanwords, which would seldom if ever be written in hiragana, and so remain in katakana. Per helpful feedback from Ken Lunde, the minor benefit of differentiating them from *on'yomi* by other means would not outweigh the drawbacks of complicating the `kJapanese` property syntax.

No action is needed, nor appropriate, for *on'yomi* which are merely irregularly formed (readings marked "O" above). Properly assessing or correcting potential errors in the *on'yomi* listed above is outside the scope of this document. It is also unclear how to handle readings which have a Sinitic etymology or Sinitic etymological intermediates, but which take said reading from a Chinese phrase other than that character itself (those marked "O/K" above); leaving them as-is seems as good as any other course of action.

5.1. Changes to kJapanese values

The following are the suggested changes to `kJapanese` values. For 21 of these characters, the existing values have Japonic or mixed Japonic/Sino-Japanese *kun'yomi* listed in katakana, or twice in both katakana and hiragana. The suggested property value lists them in hiragana only.

For others, *kun'yomi* readings have been added sourced from the `kJapaneseKun` property; this is true of `U+4ED9` 仙 (せんにん), `U+544F` 咿 (クオト), `U+565A` 嚊 (えいびろ), `U+6CD5` 法 (おきて), and `U+7ACB` 立 (たちどころ).

In the cases of `U+7D4B` 紱 (わた, わたいれ) and `U+8422` 蒑 (やぢ, やじ), both of these changes have been made.

Table 2: Suggested changes to `kJapanese` values (changes underlined)

Codepoint		Suggested kJapanese value
U+3B8B	榔	<u>いた</u>
U+3BC9	槫	ゲ　チョ　<u>くわ</u>
U+45B3	蛇	タ　ダ　チャク　<u>たた</u>
U+4787	贊	シ　<u>にえる</u>

Codepoint		Suggested kJapanese value
U+4ED9	仙	セン　セント　せんにん
U+544F	吽	ショウ　ガロン　クオト
U+565A	噚	ファゾム　えいびろ
U+56CD	囍	よし　そうき
U+5A47	婇	サイ　うねめ　うねべ
U+5CC5	峅	くら
U+5E0B	帋	シ　かみ
U+69CF	槏	ケン　タイ　かね
U+6CD5	法	ホウ　ハッ　ホッ　フラン　のっとる　のり　おきて
U+78B5	碵	セキ　ひろ
U+7ACB	立	リツ　リュウ　イ　リットル　たつ　たてる　たちどころ
U+7C17	簗	やな　やなせ
U+7D4B	紘	ゲン　コウ　ひろし　ひろむ　ひろ　わた　わたいれ
U+8422	蒮	やち　やつ　やぢ　やじ
U+87CE	蟒	だに
U+203F9	儞	ほとけ
U+237FF	櫟	なぎ
U+247F1	猴	ワ　たじひ
U+2B7BF	簗	やな　やなせ
U+2B81D	鸝	ベツ　みのがめ
U+2D160	冗	よし
U+2E494	緇	つた

5.2. Definition additions for non-Japonic (loanword) *kun'yomi*

Since the presence of non-Japonic *kun'yomi* values implies specific definitions of the characters in question (mostly, but not all, units of measure), suggested values of `kDefinition` are also given below for certain such characters, primarily those where:

- There is no pre-existing `kDefinition` value.
- The existing `kDefinition` value does not encompass the sense in question (excluding `U+74E6` 瓦, `U+7ACB` 立 and `U+7C73` 米, which are basic characters for which the sense of a unit of measure is comparatively uncommon).
- The existing `kDefinition` value appears to be erroneous—affecting only `U+7C75`, where the existing `kDefinition` value ("deci̲meter") conflicts with both the `kJapanese` and `kJapaneseKun` properties as well as the character's construction, all of which suggest a meaning of "deca̲metre".
- The existing `kDefinition` value lacks certain obvious search keywords, e.g. "cubic metre" is somewhat of a more common term than the synonymous "kilolitre".
- The existing `kDefinition` value is ambiguous—impacting e.g. `U+544E`, where "foot" is ambiguous between the anatomical and metrological senses of the English word "foot".

Additionally, like those existing for `U+2CF01` and `U+2CF04`, a `kStrange` value of category `O` might also be considered for `U+2CF17`, the shape of which may be related to that of the Latin letter `G`.

Table 3: Suggested `kDefinition` additions and changes related to non-Japonic kun'yomi

Codepoint		Suggested `kDefinition`
`U+4ED9`	仙	`Taoist super-being, transcendent, immortal; cent`
`U+5161`	甶	`mercury (chemical element); hectogram`
`U+542C`	听	`hear; understand; obey, comply; one pound`
`U+544E`	呎	`foot, twelve inches`
`U+561D`	噠	`bushel (unit of measure)`
`U+5C01`	封	`letter, envelope; feudal; one pound`

Codepoint		Suggested kDefinition
U+5F17	弗	not, negative; dollar
U+6D6C	浬	nautical mile; knot, nautical mile per hour
U+74F2	瓲	ton
U+7ACF	竡	kiloliter, cubic meter
U+7C75	籵	decameter, ten meters
U+7C7F	籿	decimeter, one tenth of a meter
U+7C80	粀	decameter, ten meters
U+7C8C	粌	hectometer
U+7CB4	粴	kilometer
U+8056	聖	holy, sacred; sage, saint
U+20490	𤒐	milligram
U+20498	𤒘	hectogram
U+20499	𤒙	kilogram
U+204AD	𤒭	milligram
U+204B2	𤒲	decagram, ten grams
U+204C8	𤓈	myriagram, ten kilograms
U+21A5B	𡩛	hectare
U+21AB8	𡪸	square meter, centiare
U+24CBE	𤲾	are, one hundredth of a hectare
U+25A58	𥩘	centiliter
U+25A7B	𥩻	deciliter
U+25A95	𥪕	kiloliter, cubic meter; classifier
U+25AB3	𥪳	milliliter

Codepoint		Suggested kDefinition
U+25E2F	籵	myriameter, ten kilometers, Swedish mile
U+2B679	鱳	blizzard
U+2B877	頡	hectare
U+2CABF	龘	square meter, centiare; one hundredth of a 公畝
U+2CF17	匹	grain (unit of weight)
U+2D505	窟	hectare
U+2D516	竈	square meter, centiare

6. *Moji Jōhō Kiban* considerations

UTC L2/22-181 notes that:

> I propose that at first we include only readings that are provided in the Moji Jōhō Kiban database, which will serve as a very stable baseline for future property changes and additions.

As such, there would appear to be no expectation that the kJapanese property remain in sync with the *Moji Jōhō Kiban* database indefinitely. However, the CITPC may wish to update their own data in response to these observations.

7. Bibliography

- **Myers**, Robert T. *A New Kana: A Modest Proposal Based on Statistical Analysis of Chinese Loanword Pronunciations.*
  - Since this work seems to be difficult to locate, a copy is attached to this document.
- **UTC L2/22-181**: Lunde, Ken (2022.08.23). *Proposal for new provisional Unihan database property: kJapanese.* https://www.unicode.org/L2/L2022/22181-unihan-kjapanese.pdf