

# Japanese Etymologies

RFD	1
AUTHOR	Theodore Nze
UPDATED	21 Mar 2025

This dictionary lists in each entry the following, in this order:

**【lemma】** pinyin, definition, [language of origin] date of first attestation in Chinese, text name: quotation

Similarly (but with some key differences) contributions on GitHub should look like this:

lemma, definition in English, word class, year of first attestation (or year given by Huang), quotation, source of quotation, note (e.g. language of origin, nota benes etc.)

e.g.

白內障,cataract,noun,1906,白內障(白翳)者。因于水晶體之溷濁。 , 『中學生理衛生教科書』 100,from Japanese: 1867 (NDL); but cf. 白色內障 attested by 1644

Let's look at some potential problems:

## Lemma

Huang uses Simplified Chinese, which we will not be using in HuangSupplement.

However, be cautious as to which traditional variant you are using e.g. X 线 → X 線 not X 綫.

## Definition

Many words can have different shades of definition, e.g. 班長 may mean 'military squad leader' or 'class group leader'. Keep your eyes peeled for the black circles with white numbers which list different senses. ❶❷

## Word class

The majority of words you encounter will be nouns. Do not assume however that the word class will be the same as that of the language of origin.

## Year of first attestation

Huang often fails to consult pre-modern texts. Definitely do not take Huang's date for granted. Unless the word is something which is definitely a modern invention e.g. X 线, definitely consider consulting sources pre-Meiji/pre-Qing. Example is 罷免 which Huang attributes to Japanese, but is much older with a slightly different pragmatic sense.

## Notes

Note should be structured as followed

from language of origin if applicable spelling in source text if different from Chinese:  
year (source); any other information

e.g. from Japanese 驅虫剂: 1886 (Nikkoku); kana reading is incorrect – should be うち  
ゆうざい\*

\*Kana reading in Huang is not actually incorrect.

## Making commits

Best practice is to make a commit for each entry, using the imperative, e.g. 'Add 愛人'.

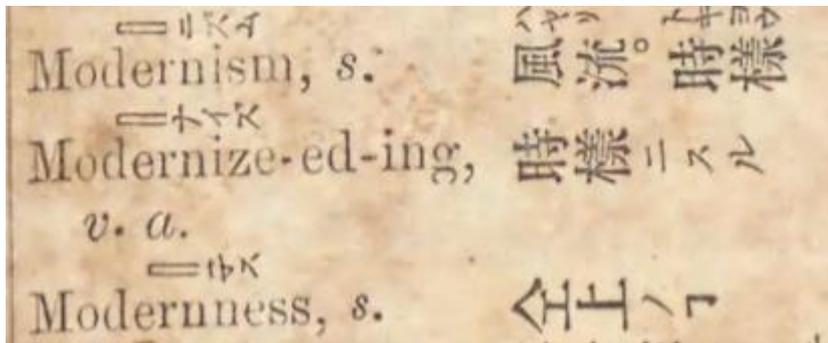
## Obsolete words

The majority of words in Huang's dictionary (or at least for those with Japanese etymologies) are obsolete and should be skipped. Keep in mind that just because a word is obsolete in Japanese, that doesn't mean it is obsolete in Chinese and vice-versa. To see if it is an obsolete word, I would recommend consulting 《現代漢語詞典》 (7th edn), available online [here](#).

## General things to be aware of when looking through archives

- If you're looking through archives which date prior to character simplification and the results are giving simplified characters, it's probably not worth your time searching through them.
- Be suspicious if the text dates pre-1910 and is written left to right. Are you sure you're looking at the actual manuscript and not a modern addition to it?
- Dealing with a database which doesn't sort chronologically but you want to check for Meiji sources? Add a ㄨ/ㄛ to your search.
- Pre-war texts are often filled with ryakuji and hentaigana. The Wikipedia article for hentaigana has a list of the most common ones, although you'll need a font e.g. BabelHan to read it. Definitely familiarise yourself with the hentaigana for

か、こ、し、す、た、つ、な、の、は and る as these are the ones likely to come up the most. Also, the ryakuji 仝 means 同 and is often used like “ or ibid. in English. And the character 丿 means こと . So the following reads: 同上のこと “same as above”.



## For those looking at Japanese etymologies

The main resources I would recommend using are:

1. 日本国語大辞典 (Nikkoku) – page 5
2. 日本語歴史コーパス (CHJ) – page 5
3. 国立国会図書館デジタルコレクション (NDL) – page 6
4. Yomiuri Database Service (Yomidasu) – page 7
5. Asahi Shinbun Cross Search (Asahi) – page 7
6. HathiTrust – page 7
7. Others – page 8

## 日本国語大辞典 (Nikkoku)

Nikkoku should be your first port of call. Not because it gives the earliest attestations, but because it's the only one which gives you a word's specific historical definition.

If the attestation given dates back to pre-Meiji, 9 times out of 10 Nikkoku is going to be the earliest you're going to find. If you really think you could find earlier in other sources, go for it but it's highly unlikely. Japanese manuscripts pre-Meiji are hard to find online and even if you do find something, because of the pervasive use of cursive script and variant forms pre-Meiji, it will likely be very hard to verify that you're looking at the correct character. And unless you're well versed in classical Japanese and classical Chinese, it will be hard to verify that the term is used in the same sense as in Huang's. This is not to mention that kango attested pre-Meiji are unlikely to be Japanese creations. If they are, then their attestation is likely to be in Nikkoku anyway.

For Meiji and post-Meiji attestations, always check in at least CHJ and NDL.

## 日本語歴史コーパス (CHJ)

Truth be told, CHJ is unlikely to have what you're looking for. It's particularly weak when it comes to words which are in any way obscure or technical. However, it takes only a few seconds to get what you want, so this should be your next port of call after Nikkoku. 7 times out of 10, its earliest attestation is going to be post-1900, but sometimes it will surprise you.

Make sure with CHJ that you're looking in the **文字列検索** with the following settings.



Also with CHJ, if nothing is coming up, try using either the traditional variant or the shinjitai as it is sensitive to this.

## 国立国会図書館デジタルコレクション (NDL)

This is what is most likely going to give you the earliest results. If Nikkoku gives you anything Meiji/post-Meiji, this is where you need to go. Especially if Huang gives a Chinese quotation from 1900s or earlier.

Although NDL may give attestations prior to what is immediately available online, since you cannot verify them, limit your search to the ログインなしで閲覧可能 as below.

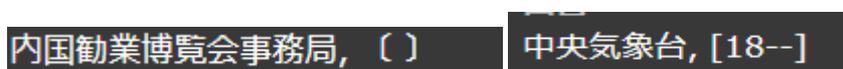
Unless of course you have open access to 国立国会図書館, but I think you need to actually be present in one of the Japanese national libraries for that...



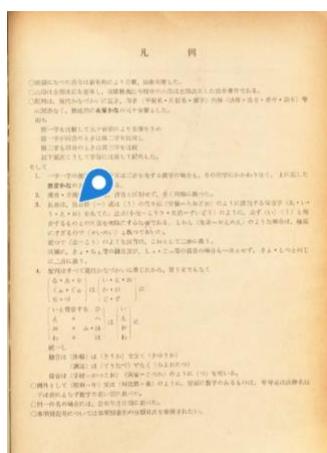
Of course, set the listing to oldest first as below.



However, what is very annoying is that a lot of the first results will be undated as below. Sometimes, they have the date in the title or on the first page, but mostly you can just ignore them.



NDL's results are based on AI generated transcriptions of the texts. These are often inaccurate so you'll have to look for yourself. However, this is relatively easy as there is an arrow which points to the relevant word when you search for the word in the document as below.



For NDL, you don't need to worry about character variants as it looks for all variants regardless. However, this does mean you should be vigilant that the manuscript variant matches what you searched.

## Yomiuri Database Service and Asahi Shinbun Cross Search

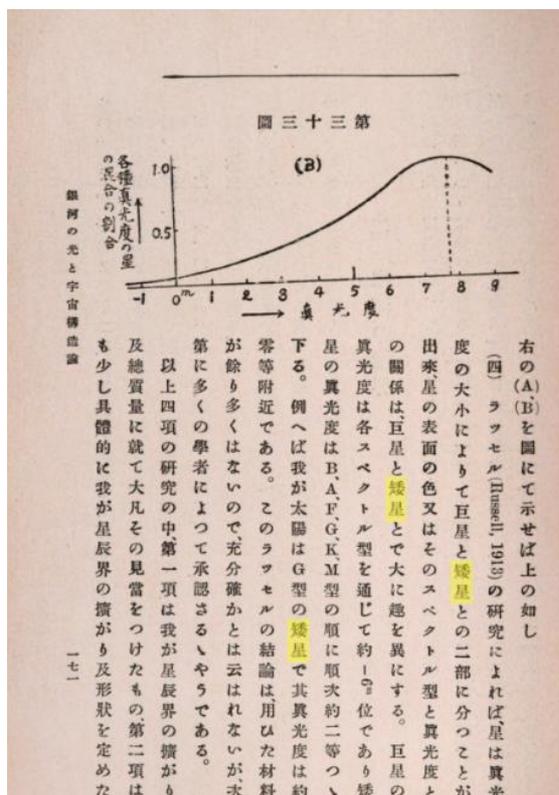
If you can't find anything relevant on Nikkoku, CHJ or NDL prior to 1900, definitely check out both Yomidasu and Asahi.

Yomidasu begins in 1874 and Asahi from 1879.

However, be cautious with both of these as the headlines and description given for an article, might not actually appear in the article itself. This is actually more the rule than the exception. Think of the keyword as a rough guide to get what you're looking for. Best practice is to look within the article itself.

## HathiTrust

If you have a particularly difficult word, HathiTrust is oftentimes the last port of call. HathiTrust has a massive amount of material pre-Meiji which is searchable thanks to AI generated transcriptions which are by and large unreliable. Sometimes, it highlights the text you searched for.



However, if the word only shows up in the text followed by a load of nonsense, it's probably not worth your time, as below.

## #28 - 1 matching term

...羽後交通脱之付き 西野花 愛妻;み奪取資...

### Other resources

The following resources should be used as a last resort or if there is some exceptional circumstances which requires pooling sources.

- Maisaku 每索 – Mainichi Shinbun Database

I don't know why but its search feature never seems to go back very far.

- Aozora Reader

Japanese literature from Meiji onwards, mostly onwards. Also no way of ordering texts oldest to newest.

- Google Books

Very similar results to HathiTrust with a lot more material to sift through (which is however likely unreliable or inaccessible).

- Wikisource

Not very reliable for original orthography.