

Andy Cherney  
14341303  
CS501 - 001

## Japanese Word Annotations

In Japanese, there are 3 writing styles: Hiragana, Katakana, and Kanji. The first 2 represent the sounds of the language, which I have learned to read quite effectively. Kanji are borrowed Chinese characters that represent ideas and distinguish words. They can have multiple readings depending on the word and context, which is why the only way to truly learn them is through vocabulary. As a learner, I love placing annotations (also known as Furigana) on top of the Kanji in sentences, so that I know how they are read. This is not the easiest task to do on a computer, though, because not many note taking apps support these kinds of annotations. The editor that I use for my notes is Obsidian, which does support it using HTML syntax in combination with Markdown. To achieve this, HTML uses the `<ruby>` and `<rt>` tags. Let's demonstrate with an example:

Suppose the word I wish to annotate is 世界 (meaning "the world"). The Hiragana reading for this word is せかい, and the English phonetic reading is "sekai." Then, the HTML ruby annotation would be:

```
<ruby>世界<rt>せかい</rt></ruby>
```

せかい

This renders as 世界 in Obsidian and on the web (The annotation could use the phonetic reading in place as well). For sentences, I would simply need to chain these annotations together with the non-Kanji characters. For example,

日本語が面白いです。 (meaning Japanese is interesting)

This sentence reads phonetically as "nihongo ga omoshiroi desu" with the Kanji words in this sentence being 日本語 and 面白い. The HTML sentence string annotated with Hiragana would look like

```
<ruby>日本語<rt>にほんご</rt></ruby>が<ruby>面白<rt>おもしろ</rt></ruby>いです。
```

にほんご おもしろ

rendering as 日本語が面白いです。 . Again, I could have also used the phonetic readings for the Kanji parts, which are "nihongo" and "omoshiro."

I would like to create a program that takes in a Japanese sentence by the user, pulls out the Kanji, allows the user to then put in the readings manually for the Kanji or Kanji sequence (if there is Kanji next to other Kanji) in that sentence, and builds the full HTML string with the provided readings. Automatic dictionary lookups are more ideal but are also more complicated to handle since certain words and conjugations can have both Kanji and Hiragana scattered in them. Regardless, once the full string is built, I would simply need to copy and paste it into Obsidian, rendering instantly and saving me time from typing the whole string out manually.

The inputs would be the Japanese sentence and the Hiragana readings for each Kanji part. The output would be the HTML ruby annotated sentence string.