

## Working with:

- **Glossing and translation tools**
- **The BCCWJ *Syoonagon* interface**

---

### *Background:*

- *Handout - Guide to using the BCCWJ Syoonagon corpus*

# 0. Today's plan

- Glossing (and understanding) Japanese text
  - JDIC Text Glossing + Google Translate
- Querying the BCCWJ
  - Some basic concepts in corpus linguistics
  - The Syoonagon interface to the BCCWJ
- Group work time on HW #4

# 1. Glossing (and understanding) Japanese text

- In linguistics writing, when you are discussing data from a language other than the one you are writing in, you typically need to provide:
  - the language **data** (in romanization)
  - a morpheme-by-morpheme **gloss**
  - when needed, a more colloquial or naturalistic **translation**

# 1. Glossing (and understanding) Japanese text

- Why the morpheme-by-morpheme gloss?

Mir      ist      kalt      (data)

*me.DAT is.3SG cold*      (gloss)

'I am cold'      (translation)

- How would a non-speaker of German be misled if given the naturalistic translation only?
- Japanese word and phrase structure is even more different from English — **glosses** and **translations** are both important to provide

# 1. Glossing (and understanding) Japanese text

- How can a non-speaker of Japanese use **online tools** to understand Japanese language data?
  - Google Translate provides a decent translation (especially for short texts)
  - But we need a **glossing tool** as well — otherwise, the structure of the Japanese example will remain opaque

# 1. Glossing (and understanding) Japanese text

- The [JDIC online Japanese dictionary](#) has a relatively useful [Text Glossing](#) tool
  - Paste the Japanese text into the box
  - The glossing tool breaks the text into units and provides dictionary info for the content words
  - Case markers and other grammatical morphemes are not glossed
  - The interface is in Japanese orthography: use the class [kana charts](#) if necessary

# 1. Glossing (and understanding) Japanese text

Try it: 日本の美しい情景はここに極まる。 [[source](#)]

- Paste this sentence into the [Text Glossing](#) tool
  - Which elements are glossed? Which of the meanings provided seem plausible?
  - Which elements are not glossed?
  - Use the class [kana charts](#) if necessary
- Compare [Google Translate](#)'s output
  - Can you put the two outputs together to come up with a gloss for this sentence?

# 1. Glossing (and understanding) Japanese text

Some commonly encountered grammatical morphemes

は	wa	TOP	topic
が	ga	NOM	nominative (subject)
を	o ("wo")	ACC	accusative (object)
に	ni	DAT/LOC	dative/locative ( <i>to, at, in, ...</i> )
で	de	(various)	<i>of / with (using) / from /</i> continuative form of copula
の	no	GEN	genitive ( <i>X's, of X</i> )
な	na	NA	with AN; also N-na + extended pred
から	kara		from; because



# 1. Glossing (and understanding) Japanese text

Sometimes, the tools do misparse the data!

- Try this sentence:

普段、葉が緑色に見えるのはクロロフィルが含まれるからである。 [[source](#)]

- What is going on with the verb morphology?
- Does Google Translate help clear things up?

## 2. Querying the BCCWJ

- Handout - [Guide to using the BCCWJ 少納言 \(Syoonagon\) corpus](#)
  - See also NINJAL's [English-language description](#) of the BCCWJ
- Corpus linguistics basics
  - What you can do with a corpus is **compare frequencies**
  - Sometimes you have to think carefully about what frequencies to compare in order to answer your question

## 2. Querying the BCCWJ

Example research question:

- We saw last time that N and AN are said to be distinguished by how they are used to modify (describe) N: **N-no** N vs. **AN-na** N
- Can we confirm that this distinguishes N and AN?
  - What are some examples we could try?
  - How can we make sure we are comparing frequencies meaningfully?

## 2. Querying the BCCWJ

- If we get surprising results, we can look at the individual examples more closely
  - This is a good application of the Text Glossing tool (with Google Translate) for support in reading Japanese text

### 3. Today's key points and skills

- 1 Understand the difference between a **translation** and a (morpheme-by-morpheme) **gloss**
- 2 Try out some tools for collecting and interpreting Japanese-language data online
- 3 Consider how to answer a research question in terms of relative frequencies of different constructions