Session 11: Hands-on Activity

Masaki EGUCHI, Ph.D.

Housekeeping

Session overview

Learning Objectives

By the end of this session, students will be able to:

- Understand NLP tasks such as POS tagging and dependency parsing
- Understand how automated parsing works
- Conduct multi-lingual Part-Of-Speech (POS) tagging using TagAnt
- Conduct POS tagging using spaCy library in Python (through Google Colab)
- Conduct Dependency parsing using spaCy library in Python (through Google Colab)

Hands-on Activity

Task 1: POS tagging with TagAnt

Task 2: POS-sensitive frequency list

Task 3: Understanding dependency grammar through visualization

POS tagging with TagAnt

Tagging with TagAnt

1. Open TagAnt
2. Select `Input Files`
3. Select `Language`
4. Select Display information (see next)

Display setting info in TagAnt

Followings are basic selection in TagAnt.

	Menu	Function	Example
word		tokenization	dogs, ran
pos		POS tag (simple)	NOUN, VERB

Menu	Function	Example
pos_tag	POS tag (detailed)	NNS, VBD
lemma	lemmatized word	dog, run

Other Diaplay settings

Menu	Function	Example
word+pos	tokenization and POS	dogs_NOUN, ran_VERB
word+lemma +pos_tag	token+lemma+POS	dogs_dog_NN, ran_run_VERB

Task 1: Annotating Japanese text (10 mins)

- Annotate 50 Japanese text files with TagAnt.
- Create frequency list for `aozora_50`

Task 1: Answer

Before

After

`_ - _ _- --_ _ - _`

`_SPACE_ _ _ _- _ _- --_ _- -_ _- _ _- -_ _- _ _- -_ _-`

Task 2: Frequency-list by POS tags (10 mins)

- Using `AntConc`, create following frequency lists:
  - Create a frequency list of `-`
- If you are done, please create another frequency list with different search terms.