

Overview of Natural Language Processing

Part II & ACS L390

Assignment 1: Syntactic Annotation

Weiwei Sun and Yulong Chen

Department of Computer Science and Technology
University of Cambridge

Michaelmas 2024/25

Assessment

Your marks are based on three practicals:

- Assignment 1: 10%; submitted through Moodle by Thursday 31 October at 12:00
- **Assignment 2: 25%; submitted through Moodle by Thursday 14 November at 12:00**
- Assignment 3: 65%; submitted through Moodle by Thursday 5 December at 12:00

Assignment II

Goal: Analyse automatically parsed results for child speech.

- Input: You are given 1000 sentences/sentence fragments of child speech.
- Subtask 1: Run Stanford CoreNLP (<https://stanfordnlp.github.io/CoreNLP/parse.html>) and Berkeley Neural Parser (<https://github.com/nikitakit/self-attentive-parser>) to obtain two sets of phrase structure trees.
- Subtask 2: Compare the two results with evalb (<https://nlp.cs.nyu.edu/evalb/>); report the overall difference in terms of bracketing scores.
- Subtask 3: Select 10 sentences to cover as more interesting phenomena as possible. Discuss the interesting phenomena. The phenomena can be linguistically interesting or computationally interesting. For example, you can choose a sentence that looks simple to you but receives wrong parsing prediction.

Assignment II (cont)

- Bonus mark: Students who choose sentences that are different from others will receive a bonus mark of 2%. The measurement of 'being different from others' is based on how many times the sentence selected by you is also selected by others. An extreme case is that all students choose totally different sentences. Then every student will receive this 2%.
- You should submit the following:
 - A .txt file to include the results from Stanford CoreNLP.
 - A .txt file to include the results from Berkeley Neural Parser.
 - A .pdf file to report the results for subtask 2 and 3. Word limit: 2000.