

Linguistic Data 2: Collection & Modeling

Assignment 2

15 February, 2020

Each one of the course participants have been allotted two articles from their Indian Language Wikipedia pages.

The task is as follows:

- From your 2 articles, take any 100 sentences.
- Annotate each sentence as per the AnnCorra format.
- For reference, the Treebanks of the Indian languages are also given with this assignment.

Resources

- Your respective article allocation can be found [here](#).
- The Hindi sample TreeBank can be found [here](#).
- The Bengali sample TreeBank can be found [here](#).
- The Telugu sample TreeBank can be found [here](#).

Post any and all doubts on the Moodle thread for this assignment.

DEADLINE: 29 February 2020, 11:55 PM