

Computational Linguistics 1

Assignment 2: POS + Chunking

Deadline: 11:55 pm, 6th October

The assignment consists of three tasks, with detailed subtasks as mentioned below.

Task 1: Comparing Indian POS tagger output with your own [9 marks]

- Take the untagged test data from Task 2 of Assignment 1 and tag it using the Indian language POS taggers available [here](#).
- Analyze and comment on the differences in POS tags assigned by the tagger with respect to your own output that you had obtained using the rules.

Task 2: Chunk the following with proper labels: NP, VP, ADJP, ADVP (English) [16 marks]

1. *She has been absent since last Wednesday.*
2. *That movie is getting scarier and scarier.*
3. *What do you do in Japan?*
4. *The Handmaid's Tale is an awesome piece of dystopian fiction.*
5. *OK. Now what?*
6. *Take this medication for pain as often as needed.*
7. *There were people everywhere, covering the roads along the route from the BJP headquarters to the Smriti Sthal from side to side, with security personnel maintaining strict vigil to ensure that nothing goes wrong.*
8. *The movie was not too terrible.*

Task 3: Chunk with proper labels (Indian Language Data) [25 marks]

- Pick sentences in an Indian language (Marathi, Bengali, Hindi, Telugu, Malayalam, Urdu, Kannada) consisting of atleast 50 words and chunk them using the labels specified in the Section **13.2 of Chunking guidelines for Indian Languages** uploaded on moodle. See Section 8 for detailed assessment.