

# Computational Linguistics - 1, Assignment 2

Zubair Abid, 20171076

The three tasks of the assignment:

1. Comparing Indian POS tagger output on previous assignment data to own output, and analyzing and commenting on the differences.
2. Chunk given data with proper labels (English)
3. Chunk Indian language data with proper labels

## 1 Compare Indian POS tagger output to own output

Running the test data from Assignment 1 through the LTRC Shallow Parser, analysing and commenting on the differences between the two.

### 1.1 Side-by-side comparison

1. Given sentence: আমার আজ দুপুর থেকেই বেশ ঘুম পেয়েছে

Tags (form: word/Own tag /LTRC):

আমার/PR\_PRP/PRP আজ/N\_NN/NN দুপুর/N\_NN/NN থেকেই/PSP/untagged বেশ/RP\_INTF/INTF ঘুম/N\_NN/NN পেয়েছে/V\_VM/VM

2. Given sentence: এই কাজ জিনিষটা আমার একদম ভাল লাগে না

zTags (form: word/Own tag /LTRC):

এই/DM\_DMD/DEM কাজ/N\_NN/NN জিনিষটা/N\_NN/NN আমার/PR\_PRP/PRP একদম/RB/RB ভাল/JJ/JJ লাগে/V\_VM\_VF/VM না/RP\_NEG/untagged

3. Given sentence: পরীক্ষাটা একদম জঘন্ন হতে চলছে

Tags (form: word/Own tag /LTRC):

পরীক্ষাটা/N\_NN/NN একদম/RB/RB জঘন্ন/JJ/JJ হতে/V\_AUX/VM চলছে/V\_VM\_VF/VM

4. Given sentence: আমার বাড়িতে একদিন আয়, বেশ মজা হবে

Tags (form: word/Own tag /LTRC):

আমার/PR\_PRP/PRP বাড়িতে/N\_NN/NN একদিন/N\_NN/NN আয়/N\_NN/NN ,./SYM বেশ/RP\_INTF/INTF মজা/N\_NN/NN হবে/V\_VM\_VF/VM

### 1.2 Analysis

Differences in the given data are minimal, primarily because most of the words in the bengali lexicon used had a single tag, so disambiguation was not a major issue. However there do exist some differences, enumerated below:

1. General vs Specific tags

Like in এই/DM\_DMD/DEM or লাগে/V\_VM\_VF/VM or হবে/V\_VM\_VF/VM or চলছে/V\_VM\_VF/VM. The difference in tag here is the specificity applied.

In the first example the LTRC tagger assigns just a "DEM" tag for Demonstrative instead of the specific deictic "DMD" tag applied by my rules.

In the next three, all the finite main verbs are just tagged as main verbs (VM) by the LTRC tagger.

2. হতে/V\_AUX/VM

The LTRC tagger considers the bengali হতে as a main verb, whereas mine classifies it as an auxiliary. It relies on the assumption in my ruleset that given two verbs in consecutive sequence only one of them is marked as a main verb, the other usually gets categorized as auxiliary.

## 2 English Chunking

Labels used: NP, VP, ADJP, ADVP (given in question).

PP is not tagged.

1. *She has been absent since last Wednesday.*

(( She )) NP  
 (( has been )) VP  
 (( absent )) NP  
 (( since ))  
 (( last Wednesday )) NP

2. *That movie is getting scarier and scarier.*

(( That movie )) NP  
 (( is getting )) VP  
 (( scarier and scarier )) ADJP

3. *What do you do in Japan?*

(( What ))  
 (( do ))  
 (( you )) NP  
 (( do )) VP  
 (( in ))  
 (( Japan )) NP

4. *The Handmaid's Tale is an awesome piece of dystopian fiction.*

(( The Handmaid's Tale )) NP  
 (( is ))  
 (( an awesome piece )) NP  
 (( of ))  
 (( dystopian fiction )) NP

5. *OK. Now what?*

(( OK )) ADJP  
 (( Now )) ADVP  
 (( What )) NP

6. *Take this medication for pain as often as needed.*

(( Take )) VP  
 (( this medication )) NP

(( for ))  
 (( pain )) NP  
 (( as often as )) ADVP  
 (( needed )) ADVP

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.*

(( There )) NP  
 (( were )) VP  
 (( people )) NP  
 (( everywhere )) ADJP  
 (( covering )) VP  
 (( the roads )) NP  
 (( along ))  
 (( the route )) NP  
 (( from ))  
 (( the BJP headquarters )) NP  
 (( to ))  
 (( the Smriti Sthal )) NP  
 (( from ))  
 (( side to side )) ADVP  
 (( with ))  
 (( security personnel )) NP  
 (( maintaining )) VP  
 (( strict vigil )) NP  
 (( to ))  
 (( ensure )) VP  
 (( that nothing )) NP  
 (( goes wrong )) VP

8. *The movie was not too terrible.*

(( The movie )) NP  
 (( was )) VP  
 (( not too terrible )) ADJP

## 3 Indian Language (Bengali) Chunking

Labels used NP, VGF, VGNF, VGINF, VGNN, JJP, RBP, NEGP, CCP, FRAGP, BLK.

1. আমার খুব খিদে পেয়েছে

(( আমার )) NP  
 (( খুব খিদে )) NP  
 (( পেয়েছে )) VGF

2. ভেজা মেঝেতে দৌড়োতে হয় না বাবা

(( ভেজা মেঝেতে )) NP  
(( দৌড়োতে )) NP  
(( হয় )) VGF  
(( না )) NEGP  
(( বাবা )) NP

3. পরীক্ষায় ভালো না করলে ঘর থেকে বার করে দেব

(( পরীক্ষায় )) NP  
(( ভালো )) JJP  
(( না করলে )) VGINF  
(( ঘর )) NP  
(( থেকে ))  
(( বার )) NP  
(( করে দেব )) VGF

4. খাবারে বিরিয়ানি না দিলে বিয়েতে কেউ আসবে না

(( খাবারে )) NP  
(( বিরিয়ানি )) NP  
(( না দিলে )) VGINF  
(( বিয়েতে )) NP  
(( কেউ )) NP  
(( আসবে )) VGF  
(( না )) NEGP

5. আমার আজ দুপুর থেকেই বেশ ঘুম পেয়েছে

(( আমার )) NP  
(( আজ )) NP  
(( দুপুর থেকেই )) NP  
(( বেশ ঘুম )) NP  
(( পেয়েছে )) VGF

6. এই কাজ জিনিষটা আমার একদম ভাল লাগে না

(( এই কাজ )) NP  
(( জিনিষটা )) NP  
(( আমার )) NP  
(( একদম ভাল )) NP  
(( লাগে )) VGF  
(( না )) NEGP

7. পরীক্ষাটা একদম জঘন্ন হতে চলছে

(( পরীক্ষাটা )) NP  
(( একদম জঘন্ন )) NP  
(( হতে চলছে )) VGNF

8. আমার বাড়িতে একদিন আয়, বেশ মজা হবে

(( আমার বাড়িতে )) NP  
(( একদিন )) NP  
(( আয় )) VP  
(( বেশ )) JJP  
(( মজা )) NP  
(( হবে )) VGF

9. অতিরিক্ত কথা বললে আমি তোদের দুজনকে ক্লাস থেকে বার করে দেব

(( অতিরিক্ত কথা )) NP  
(( বললে )) VGINF  
(( আমি )) NP  
(( তোদের দুজনকে )) NP  
(( ক্লাস থেকে )) NP  
(( বার )) NP  
(( করে দেব )) VGF

10. সকালে ঘুম থেকে উঠতে গিয়ে আমার দম বেরিয়ে যায়

(( সকালে )) NP  
(( ঘুম থেকে )) NP  
(( উঠতে গিয়ে )) VGINF  
(( আমার )) NP  
(( দম )) NP  
(( বেরিয়ে যায় )) VGF

11. কলেজ আমরা একটু বেশি খাবার বাইরে থেকে কিনে খাই কারণ যেটা কলেজ দেয় সেটা বেশ জঘন্ঘ খেতে

(( কলেজ )) NP  
(( আমরা )) NP  
(( একটু বেশি )) JJP  
(( খাবার )) NP  
(( বাইরে থেকে )) NP  
(( কিনে খাই )) VGF  
(( কারণ )) CCP  
(( যেটা...সেটা )) FRAGP  
(( কলেজ )) NP  
(( দেয় )) VGF  
(( বেশ জঘন্ঘ )) RBP  
(( খেতে )) VGNF

12. এইটা হলো আমার এসসিজিনমেন্টের শেষের সেন্টেন্স

(( এইটা )) NP  
(( হলো )) VGF  
(( আমার এসসিজিনমেন্টের )) NP  
(( শেষের )) JJP  
(( সেন্টেন্স )) NP