

# Assignment 5 – Chunking

*-Havish Chennamaraj  
(201201131)*

*-Pradeep  
Maripi(201201108)*

## **Aim:**

Writing chunk rules for data in English and Hindi. Generating chunks from a sentence with appropriate chunk tags using the former.

## **Description:**

1. About training and testing data: 20 POS tagged sentences in English and Hindi as training data and 10 POS tagged sentences as testing data.
2. Writing chunk rules: Chunk rules are written manually for both languages, constructed from the training data.
3. Code description: Applies the chunk rules written manually to chunk the test sentences. Basically it tries to make the largest possible chunk using the rules.

## **Observations:**

- Very less number of rules were obtained from the training data. So, almost all chunks were consisted of only one word.
- All kinds of verb forms were always single chunks. This may also be attributed to the lack of chunk rules containing verbs.

## **Results:**

### **English**

- [Those] [are] [not] [pencils] [.]
- [Are] [those pencils] [?]
- [What] [are] [you] [afraid] [of] [?]
- [How] [many] [sisters] [do] [you] [have] [?]
- [This] [is] [a book] [.]
- [Milk] [is] [good] [for] [you] [to eat] [.]
- [Who] [are] [you] [?]
- [I] ['m] [afraid] [.]
- [I] [have] [toothache] [.]
- [There] [are] [two] [pencils] [in] [my box] [.]

### **Hindi**

- [आकाश धूल से] [नहाया हुआ था] [।]
- [गाँव में शादी] [का] [सीजन ज़ोरों पर] [था] [।]
- [इसे] [नवाब] [शाहजेहन ने] [बनवाया था] [।]
- [इसके] [चारों] [ओर] [दीवार] [हैं] [और] [बीच] [में] [एक] [तालाब] [है] [।]
- [इसका] [प्रवेश] [द्वार] [दो] [मंजिला] [हैं] [।]
- [असम में विधानसभा] [के] [दूसरे] [चरण का मतदान] [सोमवार को] [होगा] [।]
- [जहाँ] [सोने सी] [चमकती है] [रेत...]
- [आकाश में] [आवारा] [बादल] [छा रहे थे] [।]
- [साहब से फटकार] [पड़ गयी] [क्या] [?]
- [कोई] [काम] [है] [तो] [बोलो] [।]