Lecture 25

-Augmented grammar

Treebank

- Treebank is a “bank of trees”
- It is a corpus of correctly parsed sentences
- Multiple uses of treebank
- Words have POS tags
- From treebank you can count (overall) the number of times a production/rule has been used ➔ PCFG probabilities
- Example: Penn treebank (Marcus *et al.,* 1993)

Penn treebank

S	Simple clause (sentence)	CONJP	Multiword conjunction phrases
SBAR	S′ clause with complementizer	FRAG	Fragment
SBARQ	*Wh*-question S′ clause	INTJ	Interjection
SQ	Inverted *Yes/No* question S′ clause	LST	List marker
SINV	Declarative inverted S′ clause	NAC	Not A Constituent grouping
ADJP	Adjective Phrase	NX	Nominal constituent inside NP
ADVP	Adverbial Phrase	PRN	Parenthetical
NP	Noun Phrase	PRT	Particle
PP	Prepositional Phrase	RRC	Reduced Relative Clause
QP	Quantifier Phrase (inside NP)	UCP	Unlike Coordinated Phrase
VP	Verb Phrase	X	Unknown or uncertain
WHNP	*Wh*- Noun Phrase	WHADJP	*Wh*- Adjective Phrase
WHPP	*Wh*- Prepositional Phrase	WHADVP	*Wh*- Adverb Phrase

Table 12.1 Abbreviations for phrasal categories in the Penn Treebank. The common categories are gathered in the left column. The categorization includes a number of rare categories for various oddities.

```
( (S (NP-SBJ The move)
     (VP followed
         (NP (NP a round)
             (PP of
                 (NP (NP similar increases)
                     (PP by
                         (NP other lenders))
                     (PP against
                         (NP Arizona real estate loans)))))
         ,
         (S-ADV (NP-SBJ *)
                (VP reflecting
                    (NP (NP a continuing decline)
                        (PP-LOC in
                                (NP that market))))))
     .))
```

Figure 12.2 A Penn Treebank tree.

1) Augmented grammar: ADDING THE PROBABILITIES→PCFG

PCFG problems

- PCFG results in shorter sentences
- PCFG is context-free.
- Neighboring words are not taken into account while deciding on a rule/production
- S→NP VP→Det N VP→the N VP→the fox VP→ the fox Verb→ the fox reads

2) Augmented grammar: ADDING THE HEAD VARIABLES

Lexicalized PCFG: Augmented grammar by adding head words to the rules

- Probabilities depend on relations between words in the parse tree
- How to achieve that?
- Identify the head (word/variable) of a phrase (NP, VP, PP): the most important word in the phrase
- Identify both on the left and right side of the rule
- Head of a Verb is the Verb only
- When defining probabilities define it for the combo of heads in the RHS of a rule

Identifying the head phrase

- Step 1 is to Identify the head of phrase

  Consider allocating probability to the rule

VP(v,n)→Verb(v) NP (n) [P(v,n)]

eg. VP: ate a mango

VP→Verb NP : ate a mango

where NP→Det Noun : a mango

- Head words associated with VP→Verb NP are (ate, mango)
- (ate, mango) equivalent to (in general): (Verb, Noun)

The Augmented grammar: Augment the grammar with the head variable

- Let the VP whose head word is the verb v, be denoted by VP(v)
- Let the NP whose head word is the noun n, be denoted by NP(n)
- Augmented grammar (PCFG): VP (v)→Verb(v) NP(n) $[P_1(v,n)]$

  NP(n)→Det(d) Noun(n) $[P_2(d,n)]$
- The probability $[P_1(v,n)]$ would be high for v=ate, n=mango
- The probability $[P_1(v,n)]$ would be low for v=ate, n=table
- Use the Penn treebank for counting the probability
- This time ate gets higher weightage with food items

3) Augmented grammar: ADDING THE CASE AGREEMENT (SUBJECTIVE, OBJECTIVE)

Case agreement

- Add S/SBJ and O/OBJ
- Problems solved by case agreement:

distinguish between….

I smell flowers

Me smell flowers

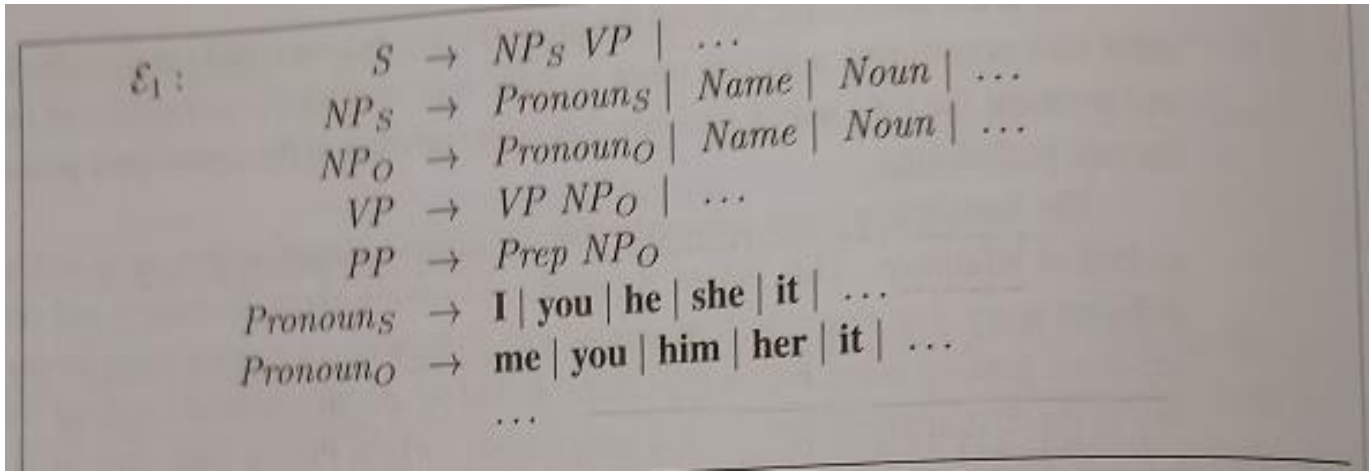
S→NP-SUBJ VP

NP-SBJ→Pronoun-SBJ

VP→Verb NP-OBJ

NP→Noun

An example of augmented grammar



$\mathcal{E}_1$:

$S \rightarrow NP_S\ VP \mid \cdots$

$NP_S \rightarrow Pronoun_S \mid Name \mid Noun \mid \cdots$

$NP_O \rightarrow Pronoun_O \mid Name \mid Noun \mid \cdots$

$VP \rightarrow VP\ NP_O \mid \cdots$

$PP \rightarrow Prep\ NP_O$

$Pronoun_S \rightarrow$ **I | you | he | she | it** $\mid \cdots$

$Pronoun_O \rightarrow$ **me | you | him | her | it** $\mid \cdots$

$\cdots$

4) Augmented grammar: ADDING THE SUBJECT-VERB AGREEMENT

Subject-verb agreement

- Add pn: 1S, 1P, 3S, 3P, (2S, 2P)
- 1S: 1st person singular (myself, i)
- 2S: 2nd person singular (you)
- 3S: 3rd person singular (She)
- 3P: 3rd person plural (They)
- Problems solved by subject-verb agreement: distinguish between-

She smells flowers

They smells flowers

An example of augmented grammar **head + CASE c (with sbj-obj) + pn (person, number)**

-All 3 augmentations are there

$\mathcal{E}_2$:

$S(head) \rightarrow NP(Sbj, pn, h)\ VP(pn, head) \mid \ldots$

$NP(c, pn, head) \rightarrow Pronoun(c, pn, head) \mid Noun(c, pn, head) \mid \ldots$

$VP(pn, head) \rightarrow VP(pn, head)\ NP(Obj, p, h) \mid \ldots$

$PP(head) \rightarrow Prep(head)\ NP(Obj, pn, h)$

$Pronoun(Sbj, 1S, \mathbf{I}) \rightarrow \mathbf{I}$

$Pronoun(Sbj, 1P, \mathbf{we}) \rightarrow \mathbf{we}$

$Pronoun(Obj, 1S, \mathbf{me}) \rightarrow \mathbf{me}$

$Pronoun(Obj, 3P, \mathbf{them}) \rightarrow \mathbf{them}$