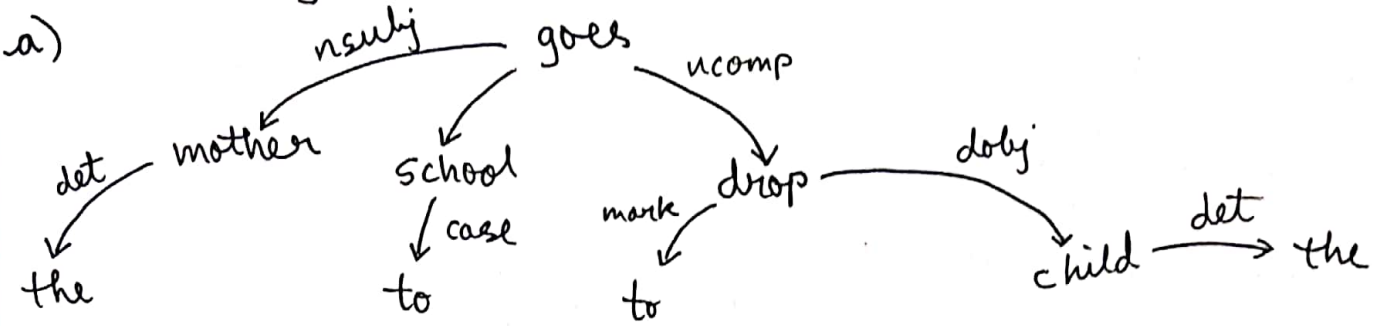
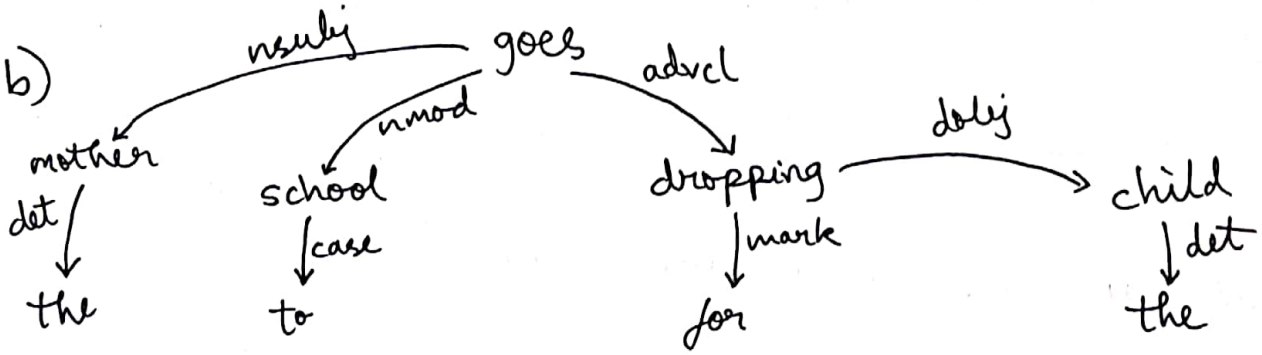
NLP - Assignment 2 2017 1047

1) Dependency Parsing

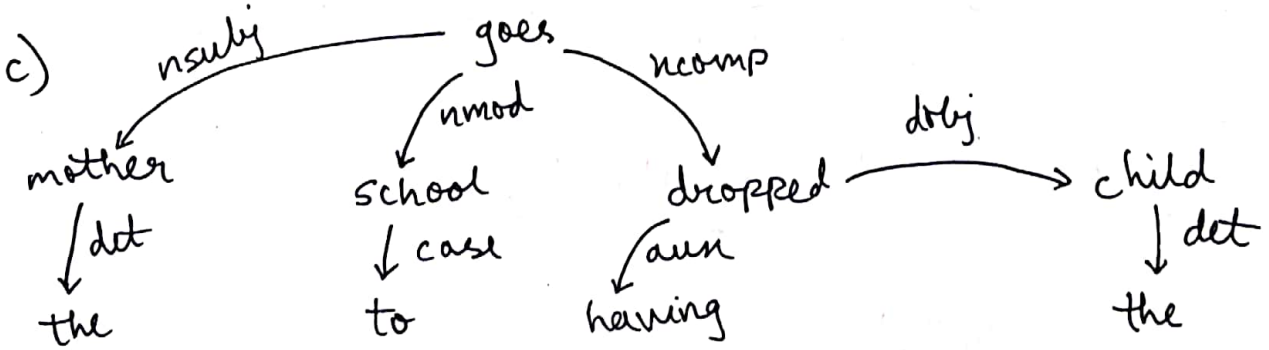
a)




b)

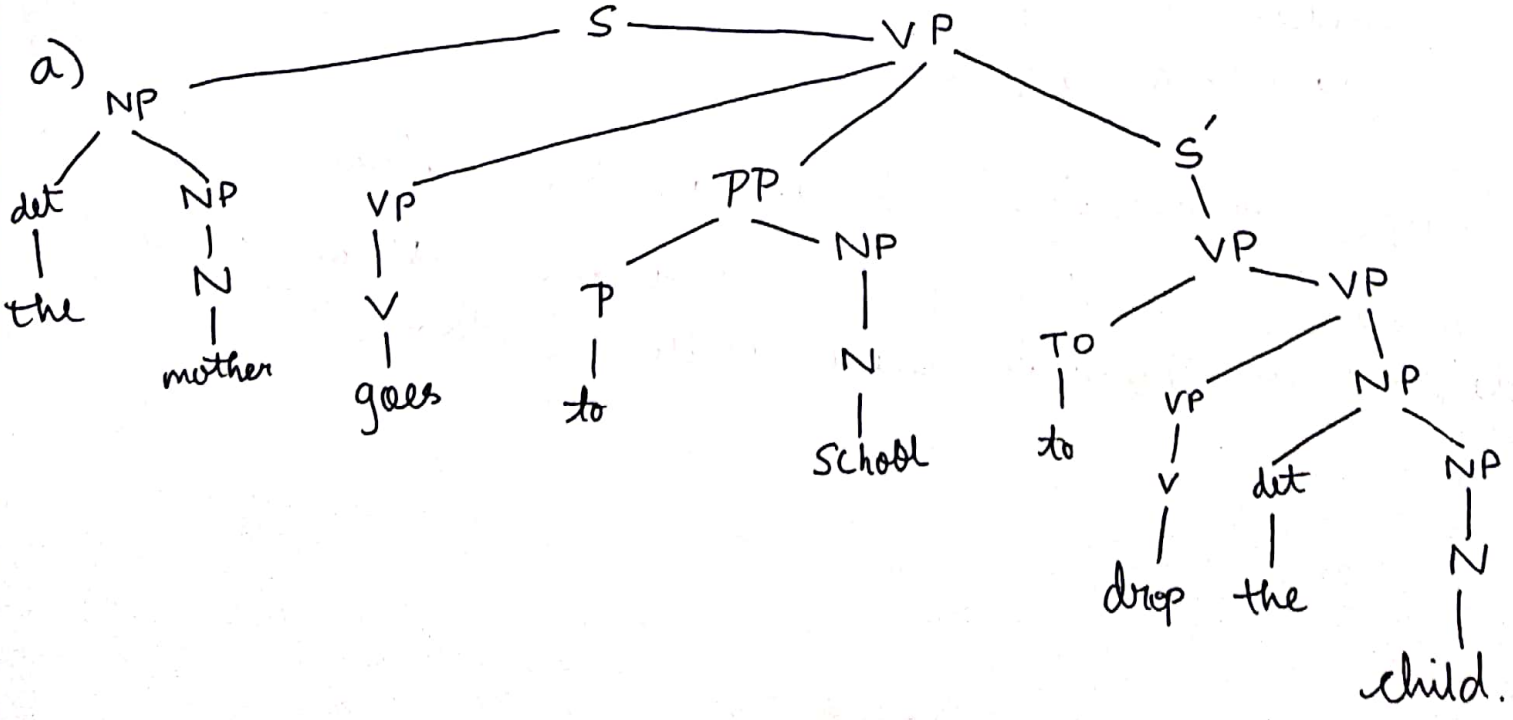



c)



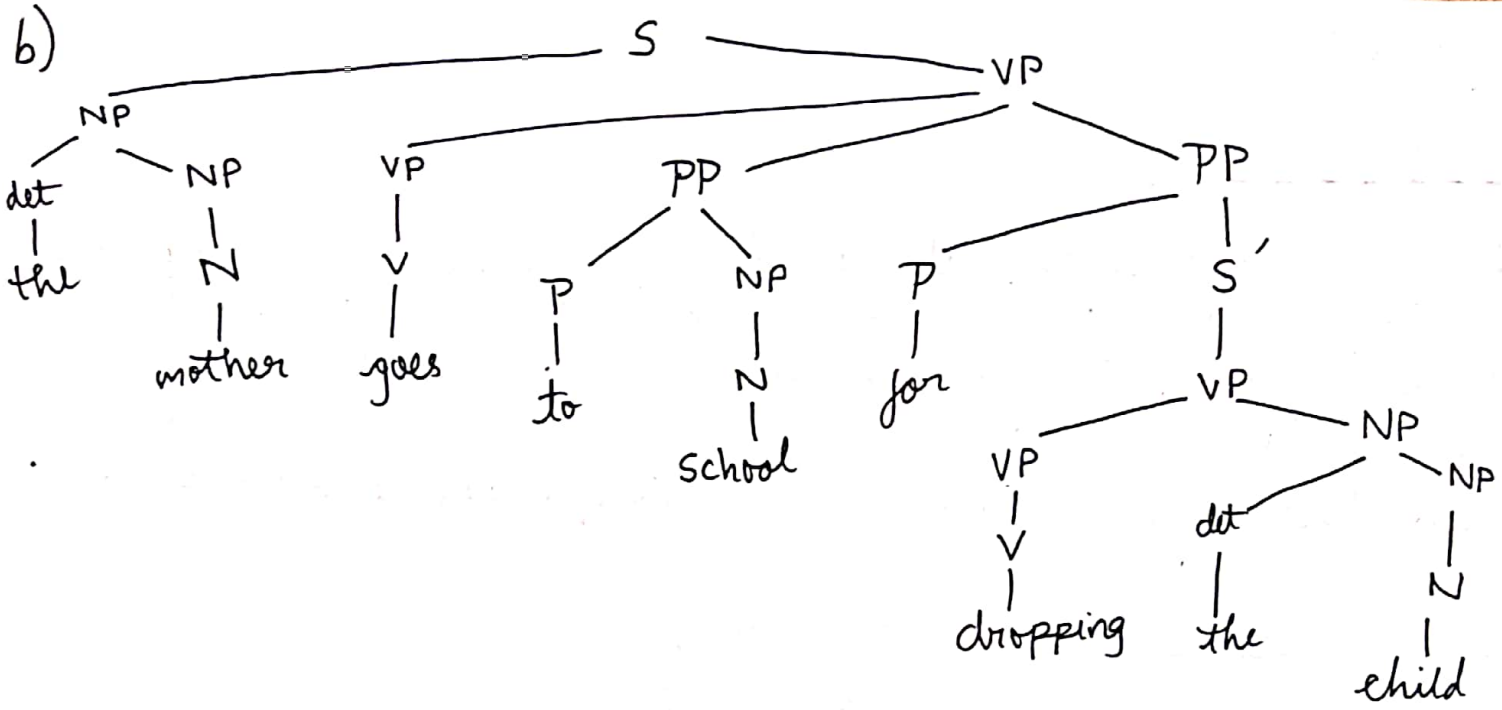

Constituency Parsing :

a)

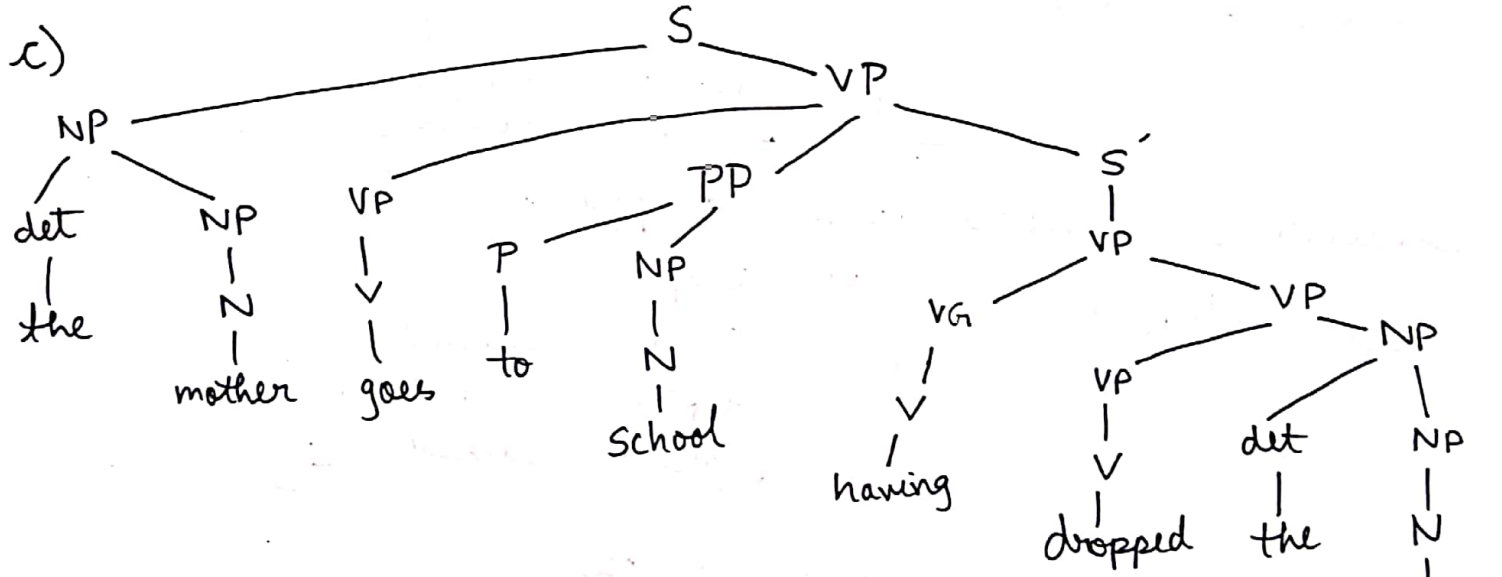



Scanned by CamScanner

b)




c)




2) Content Free Grammar: Rule Set:

S → NP VP

NP → det(?) NP | N

VP → VP PP+ S'(?) | VP NP | TO VP | VG VP | V

PP → P NP | P S'

S' → VP

VG → V

---

Part-of-speech → height 1 nodes:

N : noun

V : verb (not differentiating between verb forms.

det : determinant

Scanned by CamScanner

P : preposition
TO : "to"

---

NOTE : Using universal dependency notation for the dependency parsing & a modified version of standard constituency parse labels. Refer to the CFG for better understanding. A similar notation was followed in courses like 'Introduction to Linguistics', & 'Computational Linguistics - I'.

3) Sentence 1 : Word Form : drop

Feature Structure :

$$\begin{bmatrix} \text{root} & \text{drop} \\ \text{agr} & \begin{bmatrix} \text{tense} & \text{Present} \\ \text{aspect} & \text{Simple} \\ \text{voice} & \text{Passive} \end{bmatrix} \end{bmatrix}$$

Sentence 2 : Word Form : dropping

Feature Structure :

$$\begin{bmatrix} \text{root} & \text{drop} \\ \text{agr} & \begin{bmatrix} \text{tense} & \text{Present} \\ \text{aspect} & \text{Progressive} \\ \text{voice} & \text{Passive} \end{bmatrix} \end{bmatrix}$$

Sentence 3 : Word Form : dropped

Feature Structure :

$$\begin{bmatrix} \text{root} & \text{drop} \\ \text{agr} & \begin{bmatrix} \text{tense} & \text{Past} \\ \text{aspect} & \text{Simple} \\ \text{voice} & \text{Passive} \end{bmatrix} \end{bmatrix}$$

& Clause : having dropped :

Feature Structure :

$$\begin{bmatrix} \text{root} & \text{drop} \\ \text{agr} & \begin{bmatrix} \text{tense} & \text{Present Perfect} \\ \text{aspect} & \text{Simple} \\ \text{voice} & \text{Passive} \end{bmatrix} \end{bmatrix}$$

Scanned by CamScanner

- The roots for all the verbs is the same. We can differen-tiate between the verb forms using 'tense' & 'aspect' features.

- Having feature specifications reduces the number of CFG rules. Lets say our CFG would capture all the granularity as seen by the feature specification.

This leads to having 3 different tags / labels for height 1 nodes in the CFG trees (leading to leaves) for the 3 different verbs. e.g: VB for 'drop', VBG for 'dropping', VBD for 'dropped', etc.

But in feature specification; verbs can be captured by only 1 rule. The feature matrix handles the syntactic granularity. As one moves bottom-up from leaf nodes, overtly specific features & predicates maybe pruned out, further reducing the complexity. e.g: We may choose to prune out JJ-phrases & ADVCLs. Clearly, the number of rules will be less while using feature specifications.

4) For the 3 sentences:
Word Form: mother, clause: 'The mother'

feature specification:
$$\begin{bmatrix} \text{spec} & \text{definite} \\ \text{root} & \text{mother} \\ \text{agr} & \begin{bmatrix} \text{number} & \text{singular} \\ \text{gender} & \text{female} \end{bmatrix} \end{bmatrix}$$

Scanned by CamScanner

Word Form : goes

Feature Specification :

$$\begin{bmatrix} \text{root} & \text{go} \\ \text{agr} & \begin{bmatrix} \text{tense} & \text{Present} \\ \text{aspect} & \text{Simple} \end{bmatrix} \\ \text{number} & \text{singular} \end{bmatrix}$$

Now, while finding the specification of higher forms → 'The mother goes' : 'number' is seen as a common feature. Hence, an agreement check is conducted. Since : values of both specifications have 'singular' for number, the <u>unification</u> is accepted.

Scanned by CamScanner