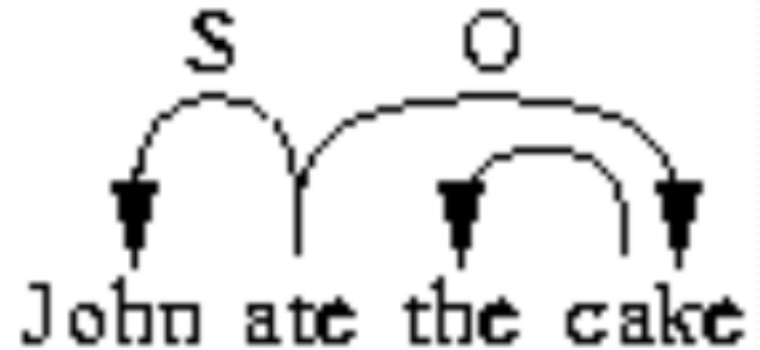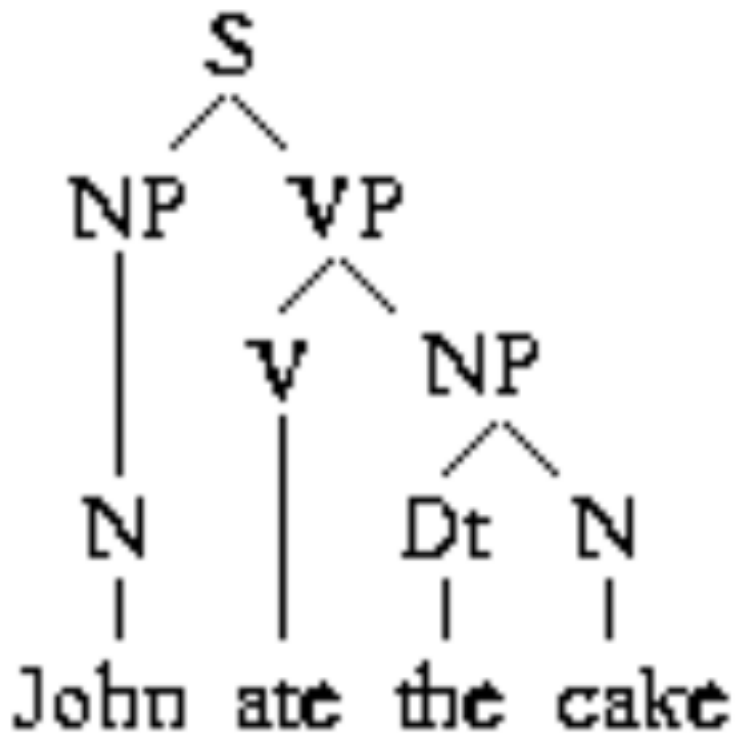# Computational Grammar

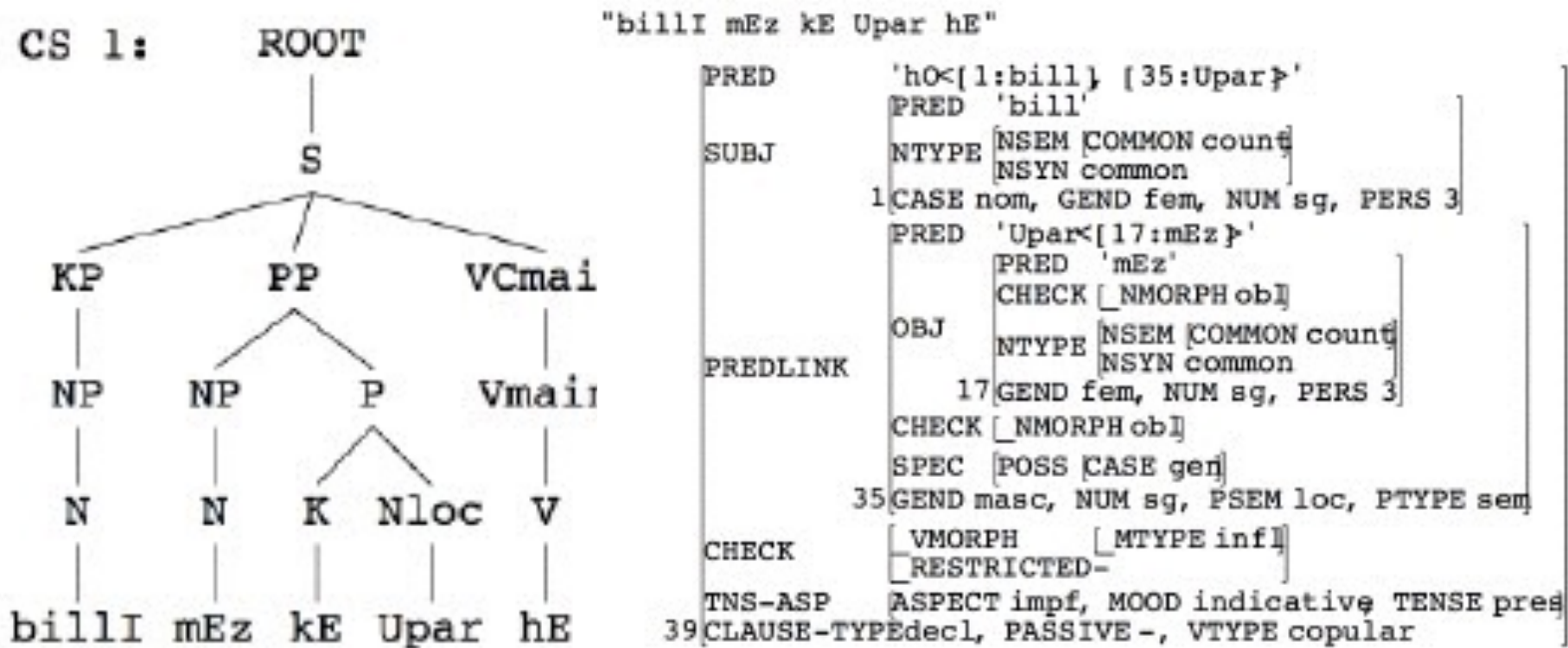## - Introduction

Tafseer Ahmed

# Syntax Representation

- Phrase Structure

- Dependency Structure

- Lexical Functional Grammar

# Phrase vs Dependency Structures

# Constituent and Functional Structures

- Lexical Functional Grammar (LFG)

# Phrase Structures

# Constituency

- Words normally group together into phrases
- Phrases group together to give larger phrases and sentences
- How to identify similar phrases or constituents
    - Similar syntactic environment
    - Movement

6

# Similar Syntactic Environment

- He sits
- The boy sits
- The naughty boy sits
- The three young men sit
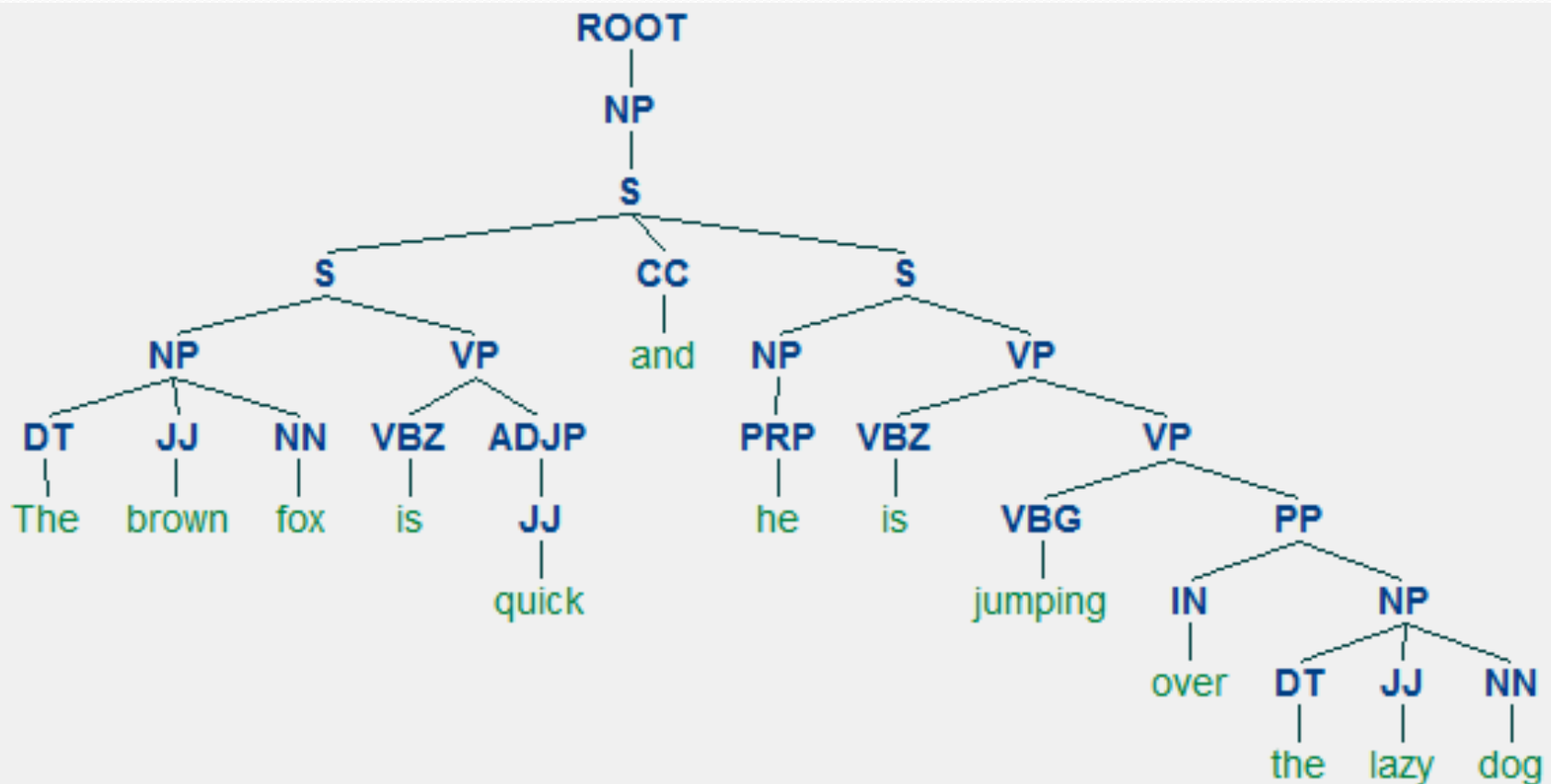- The sit
- Three sit
- Young sit

# Movement

- I am meeting with the nice old man
- With the nice old man, I am meeting
- With, I am meeting, the nice old man
- With the, I am meeting, nice old man
- With the nice, I am meeting, old man

# Phrase Structures

- Noun Phrase (NP)
- Verb Phrase (VP)
- Prepositional Phrase (PP)
- ….

# Phrase Structure - another example

# Dependency Structures

universaldependencies.org
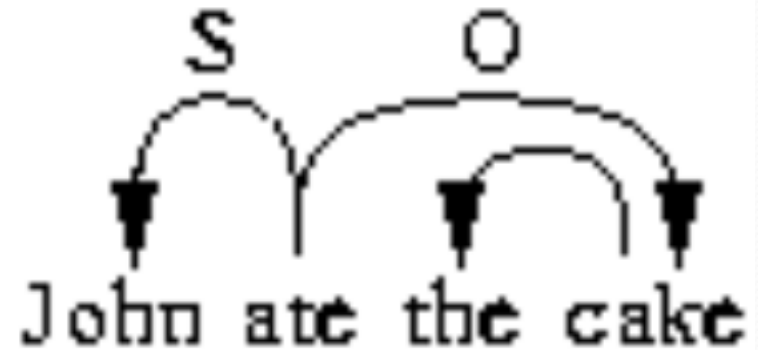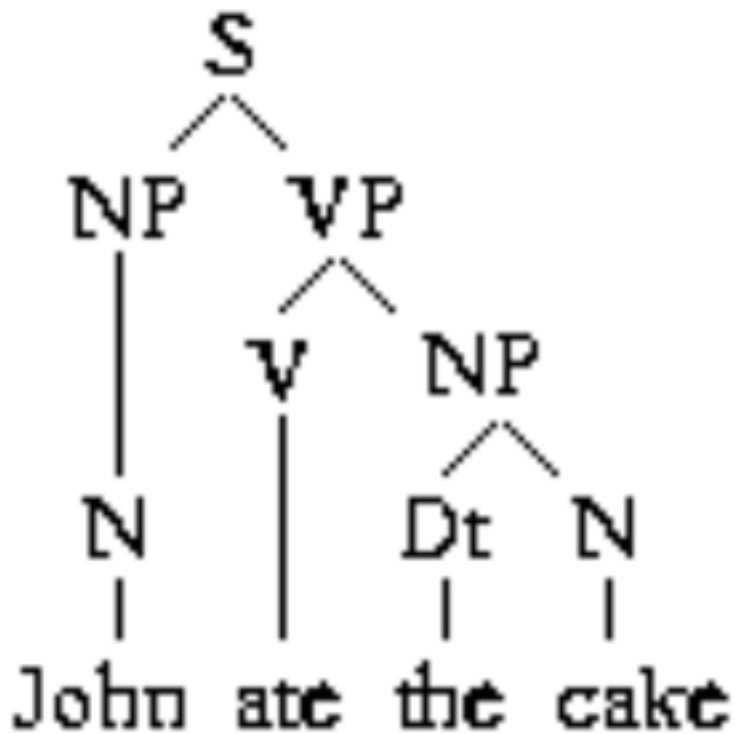
# Universal Dependencies

Universal Dependencies (UD) is a framework for consistent annotation of grammar (parts of speech, morphological features, and syntactic dependencies) across different human languages. UD is an open community effort with over 300 contributors producing more than 150 treebanks in 90 languages. If you're new to UD, you should

# Phrase vs Dependency Structures

# Universal Dependencies

|  | Nominals | Clauses | Modifier words | Function Words |
|---|---|---|---|---|
| Core arguments | nsubj<br>obj<br>iobj | csubj<br>ccomp<br>xcomp | | |
| Non-core dependents | obl<br>vocative<br>expl<br>dislocated | advcl | advmod*<br>discourse | aux<br>cop<br>mark |
| Nominal dependents | nmod<br>appos<br>nummod | acl | amod | det<br>clf<br>case |
| Coordination | MWE | Loose | Special | Other |
| conj<br>cc | fixed<br>flat<br>compound | list<br>parataxis | orphan<br>goeswith<br>reparandum | punct<br>root<br>dep |

# Urdu Examples
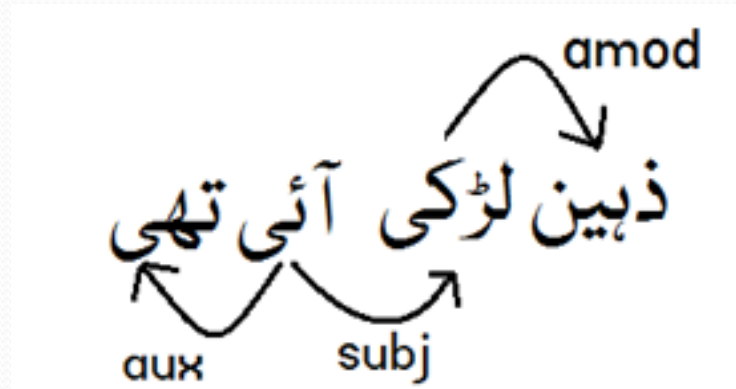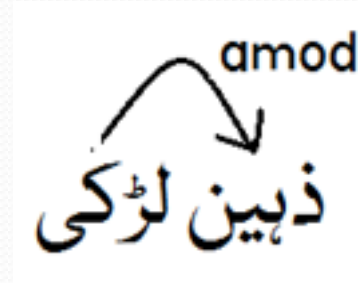
# Urdu Examples
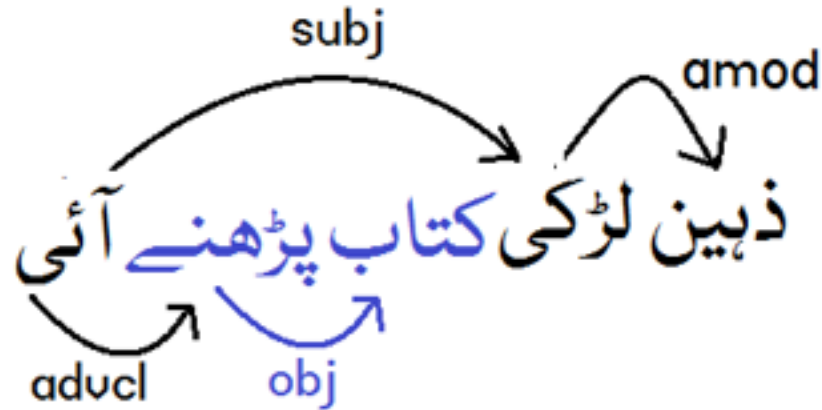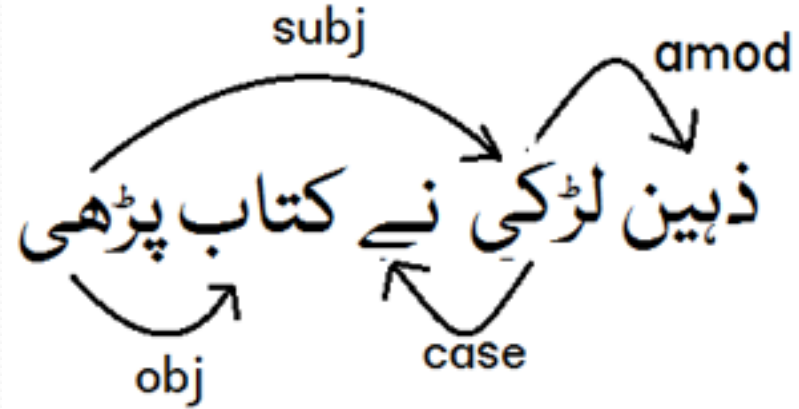
# Annotation

# Annotation - Urdu Example

| | ذہین | لڑکیوں | نے | اچھی | کتابیں | پڑھی | تھیں |
|---|---|---|---|---|---|---|---|

# Annotation - Urdu Example

| تھیں | پڑھی | کتابیں | اچھی | نے | لڑکیوں | ذہین | |
|------|------|--------|------|-----|--------|------|------|
| ہے | پڑھ | کتاب | اچھا | نے | لڑکی | ذہین | **Lemma** |

# Annotation - Urdu Example

| تھیں | پڑھی | کتابیں | اچھی | نے | لڑکیوں | ذہین | |
|---|---|---|---|---|---|---|---|
| ہے | پڑھ | کتاب | اچھا | نے | لڑکی | ذہین | **Lemma** |
| Aux | Verb | NN | Adj | AdP | Noun | Adj | **POS** |

# Annotation - Urdu Example

| تھیں | پڑھی | کتابیں | اچھی | نے | لڑکیوں | ذہین | |
|---|---|---|---|---|---|---|---|
| ہے | پڑھ | کتاب | اچھا | نے | لڑکی | ذہین | **Lemma** |
| Aux | Verb | NN | Adj | AdP | Noun | Adj | **POS** |
| | Form=Perf Gend=Fem Pers=3 Num=Pl | Gend=Fem Num=Pl Form=Nom | Gend=Fem Num=Pl Form=Nom | | Gend=Fem Num=Pl Form=Obl | Gend=Fem Num=Pl Form=Obl | **Features** |

# Annotation - Urdu Example



| تھیں | پڑھیں | کتابیں | اچھی | نے | لڑکیوں | ذہین | |
|---|---|---|---|---|---|---|---|
| ہے | پڑھ | کتاب | اچھا | نے | لڑکی | ذہین | **Lemma** |
| Aux | Verb | NN | Adj | AdP | Noun | Adj | **POS** |
| | Form= Perf Gend= Fem Pers=3 Num=Pl | Gend=Fem Num=Pl Form=Nom | Gend= Fem Num= Pl Form= Nom | | Gend= Fem Num=Pl Form=Obl | Gend= Fem Num=Pl Form= Obl | **Features** |

# CoNLL Format

- CoNLL (Conference on Natural Language Learning) format
- Representing graph (and other tags) in text file

Id

Word

Lemma

Coarse Grained POS

Fine Grained POS

Features

Host

Dependency Type

# CoNLL Format

File  Edit  Format  View  Help

| 1 | ذبین | ذبین | Adj | Adj | - | 2 | amod |
| 2 | لڑکیاں | لڑکی | Noun | NN | - | 6 | subj |
| 3 | نے | نے | Adp | PP | - | 2 | case |
| 4 | اچھی | اچھا | Adj | Adj | - | 5 | amod |
| 5 | کتابیں | کتاب | Noun | NN | - | 6 | obj |
| 6 | پڑھیں | پڑھ | Verb | VB | - | 0 | ROOT |
| 7 | تھیں | ے | Aux | Aux | - | 6 | aux |

# Tools for Annotation

- **GATE** (General Architecture for Text Engineering), Sheffield

- **Brat,** Manchester, Tokoyo,......

- **Webanno,** Darmstadt

- ...

# Lexical Functional Grammar
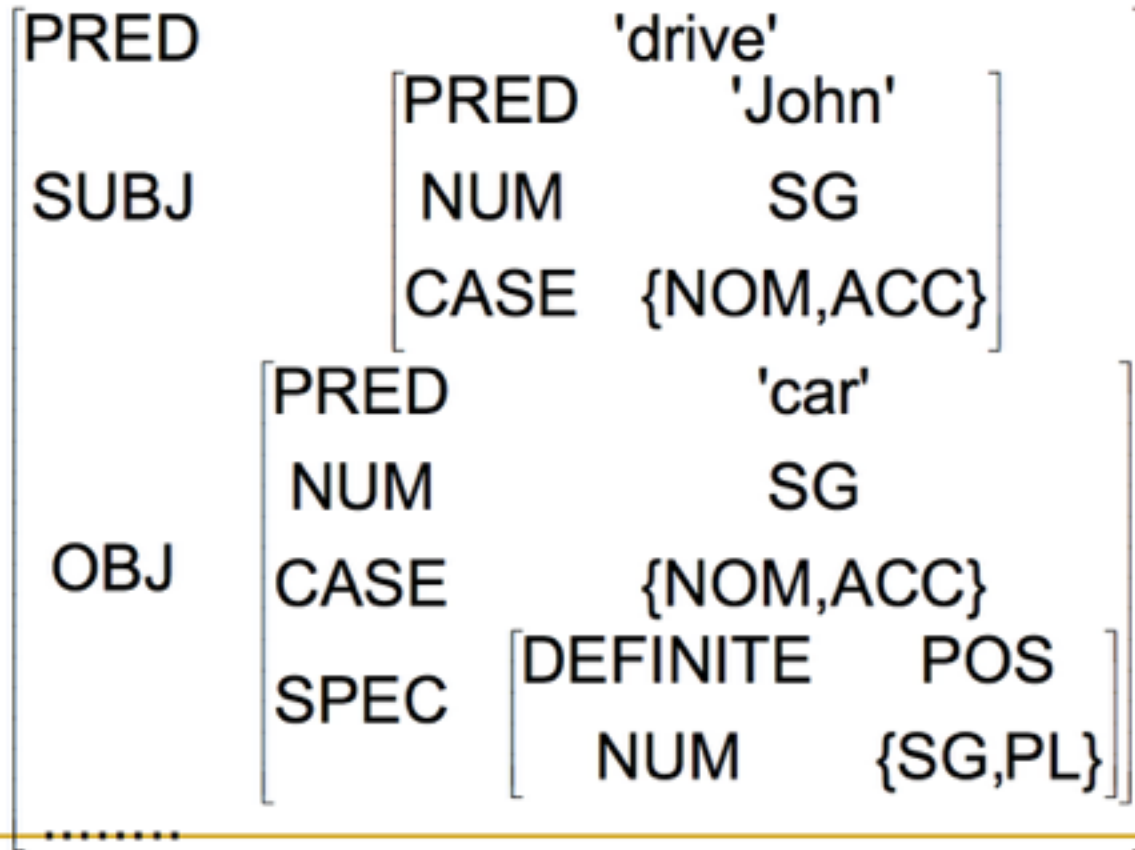
# Lexical Functional Grammar (CFG)

- Defines two different syntactic structures
  - **C(Constituent) Structure** represents linear and hierarchical organization of words into phrases
  - **F(Functional) Structure** represents abstract functional syntactic organization of the sentence, representing syntactic predicate-argument structure and functional relations like subject and object

# Functional Structure

John drives the car.

$$
\begin{bmatrix}
\text{PRED} & \text{'drive'} \\
\text{SUBJ} & \begin{bmatrix} \text{PRED} & \text{'John'} \\ \text{NUM} & \text{SG} \\ \text{CASE} & \{\text{NOM,ACC}\} \end{bmatrix} \\
\text{OBJ} & \begin{bmatrix} \text{PRED} & \text{'car'} \\ \text{NUM} & \text{SG} \\ \text{CASE} & \{\text{NOM,ACC}\} \\ \text{SPEC} & \begin{bmatrix} \text{DEFINITE} & \text{POS} \\ \text{NUM} & \{\text{SG,PL}\} \end{bmatrix} \end{bmatrix} \\
\dots\dots
\end{bmatrix}
$$

زبان / Language : اردو

جملہ لکھیں / Enter Sentence : وہ سیب کھاتا ہے

ترکیب / Parse

Funtional Structure [F-Structure]    تفاعلی ساخت    Constituent Structure [C-Structure]    نحوی ساخت

**F-Structure**

- F-Structure # 1
  - Root
    - SUBJ
      - PRED —— 'pro'
      - NUM —— SG
      - GEND —— M
      - PRONTYPE —— PERS
      - PERS —— 3
      - FORM —— NOM
      - CASE —— NOM
      - ANIM —— POS
      - RESPECT —— NORESPECT
      - PRONFORM —— 'WOH'
    - OBJ
      - NUM —— SG
      - RESPECT —— NORESPECT
      - GEND —— M
      - FORM —— NOM OBL
      - PRED —— 'سیب'
      - CTYPE —— COUNT
      - NTYPE —— COMMON
      - ANIM —— NEG
      - PERS —— 3
      - CASE —— NOM
    - PRED —— 'کھا<SUBJ OBJ>'
    - _VERB_FORM —— 'kha'
    - _SUBJ_CASE —— NOM

**C-Structure**

- C-Structure # 1
  - Sdec
    - KP
      - NP
        - NPpron
          - 'وہ' —— pro
    - KP
      - NP
        - NPcommon
          - AP
          - 'سیب' —— n
    - VP
      - ComplexVP
        - MainVP
          - 'کھاتا' —— v
      - TenseAuxP
        - 'ہے' —— tense_aux