Natural Language Processing
CSE 325/425

Sihong Xie

Lecture 15:

- Context-free grammars

Grammar

Grammar: the whole system and structure of a language or of languages in general, usually taken as consisting of **syntax** and morphology (including inflections) and sometimes also phonology and semantics.

Syntax: the way words are arraged together. We would not say

** together words the way are arraged.*

since it is not how words are organized in English.

Language models and HMM are computational models that embody a syntax.

Context-free grammar

Context-free grammar is a formal mathematical system modeling constituent structures.

Constitute structures: a group of words that acts as a single grammatical unit.

Harry the Horse	a high-class spot such as Mindy's
the Broadway coppers	the reason he comes into the Hot Box
they	three parties from Brooklyn

three parties from Brooklyn *arrive…*
a high-class spot such as Mindy's *attracts…*
the Broadway coppers *love…*
they *sit*

- a unit can't be broken down while maintaining its semantics.

On September seventeenth, I'd like to fly from Atlanta to Denver
I'd like to fly *on September seventeenth* from Atlanta to Denver
I'd like to fly from Atlanta to Denver *on September seventeenth*

- a unit can appear in different places in a sentence without changing its semantics

*On September, I'd like to fly seventeenth from Atlanta to Denver
*On I'd like to fly September seventeenth from Atlanta to Denver
*I'd like to fly on September from Atlanta to Denver seventeenth

Context-free grammar

Formal definition of CFG. (4-tuple)

N	a set of **non-terminal symbols** (or **variables**)
Σ	a set of **terminal symbols** (disjoint from N)
R	a set of **rules** or productions, each of the form $A \rightarrow \beta$,
	where A is a non-terminal,
	β is a string of symbols from the infinite set of strings $(\Sigma \cup N)^*$
S	a designated **start symbol** and a member of N

Compared to Hmm:

- Hmm 'Generate language
- Hmm defines [a seq. of pos-tags] for a sentence.

Why CFG: generative in ML

- define a language (generation).
- find structures for a sentence, which would have been unstructured.

Derivations: if $A \rightarrow \beta$ is a production of R and α and γ are any strings in the set $(\Sigma \cup N)^*$, then we say that $\alpha A \gamma$ **directly derives** $\alpha \beta \gamma$, or $\alpha A \gamma \Rightarrow \alpha \beta \gamma$.

$$\alpha_1 \Rightarrow \alpha_2, \alpha_2 \Rightarrow \alpha_3, \ldots, \alpha_{m-1} \Rightarrow \alpha_m \qquad \alpha_1 \overset{*}{\Rightarrow} \alpha_m$$

($\overset{*}{\Rightarrow}$ means one or multiple steps of derivations)

Language derived from CFG: $\mathcal{L}_G = \{w | w \text{ is in } \Sigma* \text{ and } S \overset{*}{\Rightarrow} w\}$

Context-free grammar

Example CFG.

Grammar Rules			Examples
S	→	*NP VP*	I + want a morning flight
NP	→	*Pronoun*	I
	|	*Proper-Noun*	Los Angeles
	|	*Det Nominal*	a + flight
Nominal	→	*Nominal Noun*	morning + flight
	|	*Noun*	flights
VP	→	*Verb*	do
	|	*Verb NP*	want + a flight
	|	*Verb NP PP*	leave + Boston + in the morning
	|	*Verb PP*	leaving + on Thursday
PP	→	*Preposition NP*	from + Los Angeles

Non-terminals

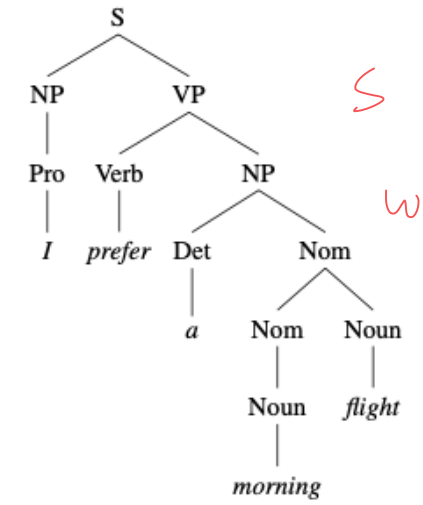
(POS-tags)
Non-terminals | Terminals

Noun	→	*flights* | *breeze* | *trip* | *morning*
Verb	→	*is* | *prefer* | *like* | *need* | *want* | *fly*
Adjective	→	*cheapest* | *non-stop* | *first* | *latest* | *other* | *direct*
Pronoun	→	*me* | *I* | *you* | *it*
Proper-Noun	→	*Alaska* | *Baltimore* | *Los Angeles* | *Chicago* | *United* | *American*
Determiner	→	*the* | *a* | *an* | *this* | *these* | *that*
Preposition	→	*from* | *to* | *on* | *near*
Conjunction	→	*and* | *or* | *but*

A parsing tree based on derivations using the CFG

level / layer 0

level / layer 1



level / layer n

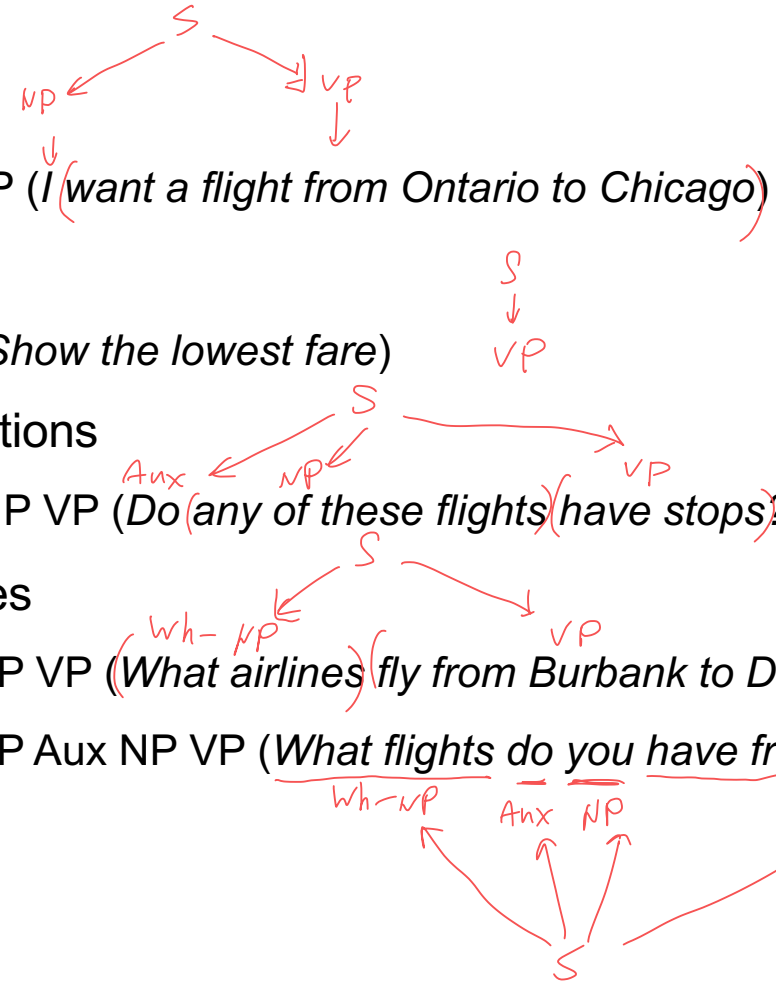
$S \overset{*}{\Rightarrow} w \in \Sigma^*$

$w = (I, prefer, a, morning, flight) \in \Sigma^*$

Recursive NN defined on a parsing tree.

Sentence structures

- Declarative
  - S→NP VP (*I want a flight from Ontario to Chicago*)
- Imperative
  - S→ VP (*Show the lowest fare*)
- Yes-no questions
  - S→Aux NP VP (*Do any of these flights have stops?*)
- Wh-structures
  - S→Wh-NP VP (*What airlines fly from Burbank to Denver?*)
  - S→Wh-NP Aux NP VP (*What flights do you have from Burbank to Denver?*)



Noun phrase (NP)

- NP→(Det) (Card) (Ord) (Quant) (AP) Nominal
  - () means the enclosed non-terminal is optional
  - Card→*one, two*, …; Ord→*first, second*, …; Quant→*many, (a) few, several*
  - AP→(RB) JJ: *least expensive*
- Determiners (Det)
  - simple lexical determiners: a, an, the
  - possessive (Det→NP's): *United's flight*
    - Together with NP→Det NP
    - *Det* ⇒ NP's ⇒Det NP's ⇒NP's NP's ⇒Det NP's NP's⇒NP's NP's NP's
    - *Denver's mayor's mother's canceled flight*
  - From the above example, we can see the derivation can be recursive.

Noun phrase (NP)

- Nomials Nominals
  - simple nomials: Nomial Nominal→Noun
    - this serves as the bottom of a parsing tree with no more recursion.
  - more complex nomials:
    - Nominal→Nomial Noun (*morning flight*)
    - Nominal→Nomial PP (*flight to Boston*)
    - Nominal→Nomial Gerundive-VP (*flight leaving before 10*) More on VP later. present tense
    - Nominal→Nomial ed-VP (*dinner served on the flight*)
    - Nominal→Nomial infinitive (*flight to arrive in Boston*)
    - Nominal →relative-clause (*flight that serves breakfast*)

Verb phrases (VP)

Examples of VP

- VP→Verb (*disappear*)
- VP→Verb NP (*prefer a morning flight*)
- VP→Verb NP PP (*leave Boston in the morning*)
- VP→Verb PP (*leaving on Thursday*)
- VP→Verb VP (*want to fly from Milwaukee to Orlando*)
- VP →Verb S (You said you had a two hundred sixty six dollar fare)
  - S is called “sentential complement”

Penn treebank

A corpus with sentences annotated with parsing trees.

- derive CFG grammars from the data;
- serve as training data for syntactic parsers (next few lectures).

```
((S
  (NP-SBJ (DT That)
    (JJ cold) (, ,)
    (JJ empty) (NN sky) )
  (VP (VBD was)
    (ADJP-PRD (JJ full)
      (PP (IN of)
        (NP (NN fire)
          (CC and)
          (NN light) ))))
  (. .) ))
```

(a)

```
((S
   (NP-SBJ The/DT flight/NN )
   (VP should/MD
     (VP arrive/VB
       (PP-TMP at/IN
         (NP eleven/CD a.m/RB ))
       (NP-TMP tomorrow/NN )))))
```

(b)