SYNTAX

Sudeshna Sarkar

14 AUG 2019

Syntax

- Constituency
- Ordering
- Grammatical relations and dependency
  - Heads, agreement, grammatical function
- Key formalisms
  - Context-free grammars
  - Dependency grammars
- Resources
  - Treebanks

Constituency

- Basic idea: groups of words act as a single unit
- Constituents form coherent classes that behave similarly
  - With respect to their internal structure: e.g., at the core of a noun phrase is a noun
  - With respect to other constituents: e.g., noun phrases generally occur before verbs

Constituency: Example

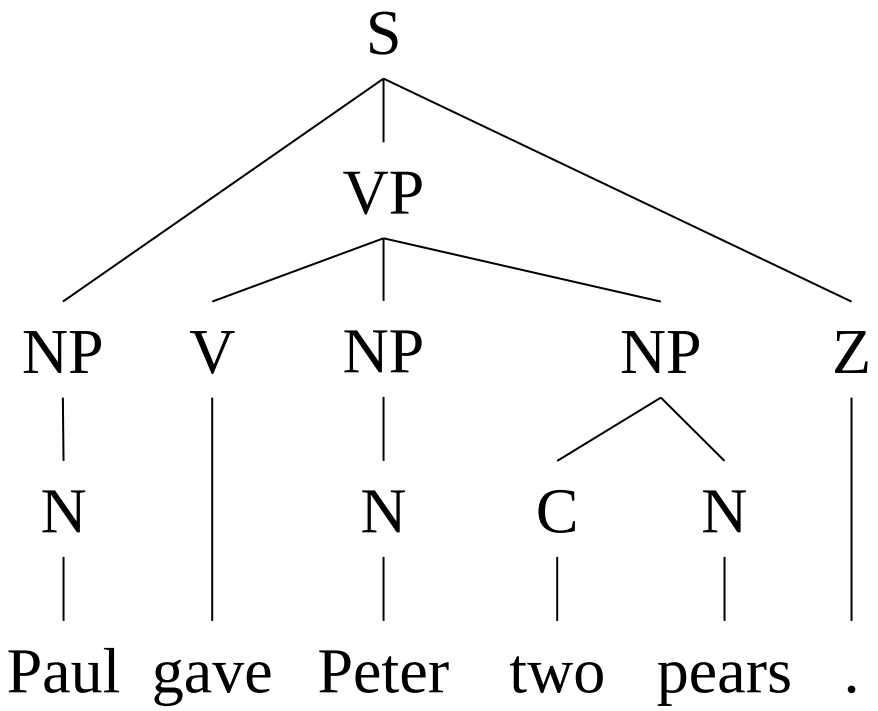
- Noun phrases in English...

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

- They can all precede verbs
- They can all be preposed/postposed

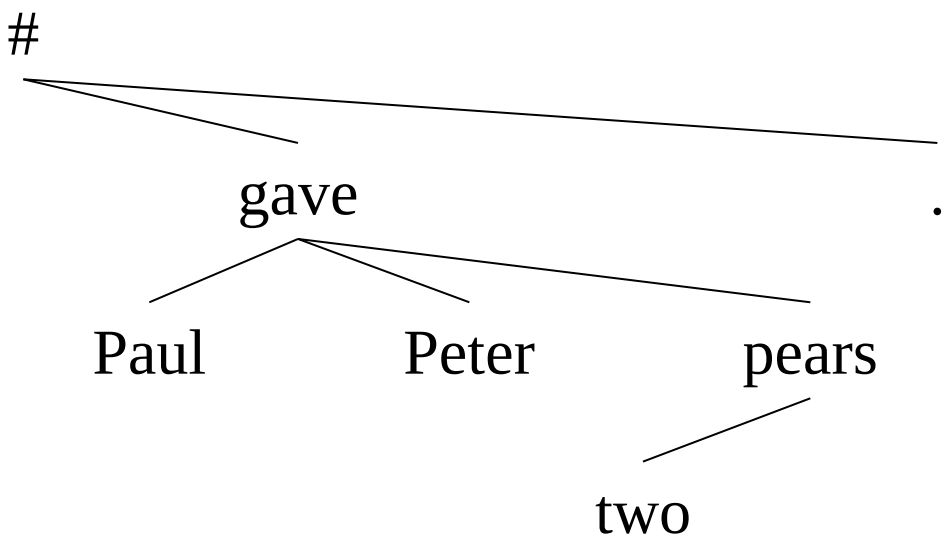
Example of Constituent Tree

((Paul (gave Peter (two pears))) .)



Example of Dependency Tree

[#,0] ([gave,2] ([Paul,1], [Peter,3], [pears,5] ([two,4])), [.,6])



Words and Phrases

- Word (token): smallest unit of the syntactic layer
  - grammatical (function) words
  - lexical (content) words
- Phrase
  - Sequence of immediate constituents (words or phrases).
- Phrase types by their main word—head
  - Noun phrase: *the new book of my grandpa*
  - Adjectival phrase: *brand new*
  - Adverbial phrase: *very well*
  - Prepositional phrase: *in the classroom*
  - Verb phrase: *to catch a ball*

Noun Phrase

- A noun or a (substantive) pronoun is the head.
  - *water*
  - *the book*
  - *new ideas*
  - *two millions of inhabitants*
  - *one small village*
  - *the greatest price movement in one year since the World War II*
  - *operating system that, regardless of all efforts by our admin, crashes just too often*
  - *he*
  - *whoever*

Evidence of Constituency

1. They can all appear in similar syntactic environments
   - NP before a verb

2. Preposed or Postposed constructions
   - The prepositional phrase can be placed in a number of different locations in the sentence
   - But the individual words in the phrase cannot.

Adjective Phrase

- An adjective or a determiner (attributive pronoun) is the head.
- Simple ADJPs are very frequent, complex ones are rare.
  - *old*
  - *very old*
  - *really very old*
  - *five times older than the oldest elephant in our ZOO*
  - *sure that he will arrive first*

Adverbial Phrases

- An adverb is the head.
  - *quickly*
  - *much more*
  - *how*
  - *louder than you can imagine*
  - *yesterday*

Prepositional (Postpositional) Phrase

- The preposition serves as head (because it determines the case of the rest of the phrase).
- Often have a function similar to adverbial phrases or noun phrases (object of a verb).
  - *in the city center*
  - *in God*
  - *around five o'clock*
  - *to a better future*
  - *up to a situation where neither of them could back out*
  - *with respect to his nonage*

Clause and Sentence

- Group of words with 1 predicate, e.g.:
  - *John loves Mary.*
  - *…that you are right.*
  - simple sentence or part of compound sentence
- Sentence
  - simple sentence or compound sentence
  - consists of one or more clauses
  - e.g. *John loves Mary.* or *“I realized that you were right.”*

Clause and Sentence

- Main clause
  - Independent of other clauses in the sentence
- Nested clause, relative clause
  - Depends on another clause, carries out a function in that clause (as a dependent phrase)
  - This is the man [that] I saw

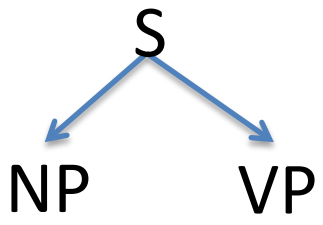
Sentence

- Consists of one or more main clauses.
- If there are more than one main clause then they are usually coordinated.

Formal Grammars of English

Context-free grammars (CFGs)

- Consist of
  - Rules
  - Terminals
  - Non-terminals
  - Start Symbol



- Specifies a set of tree structures that capture constituency and ordering in language

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

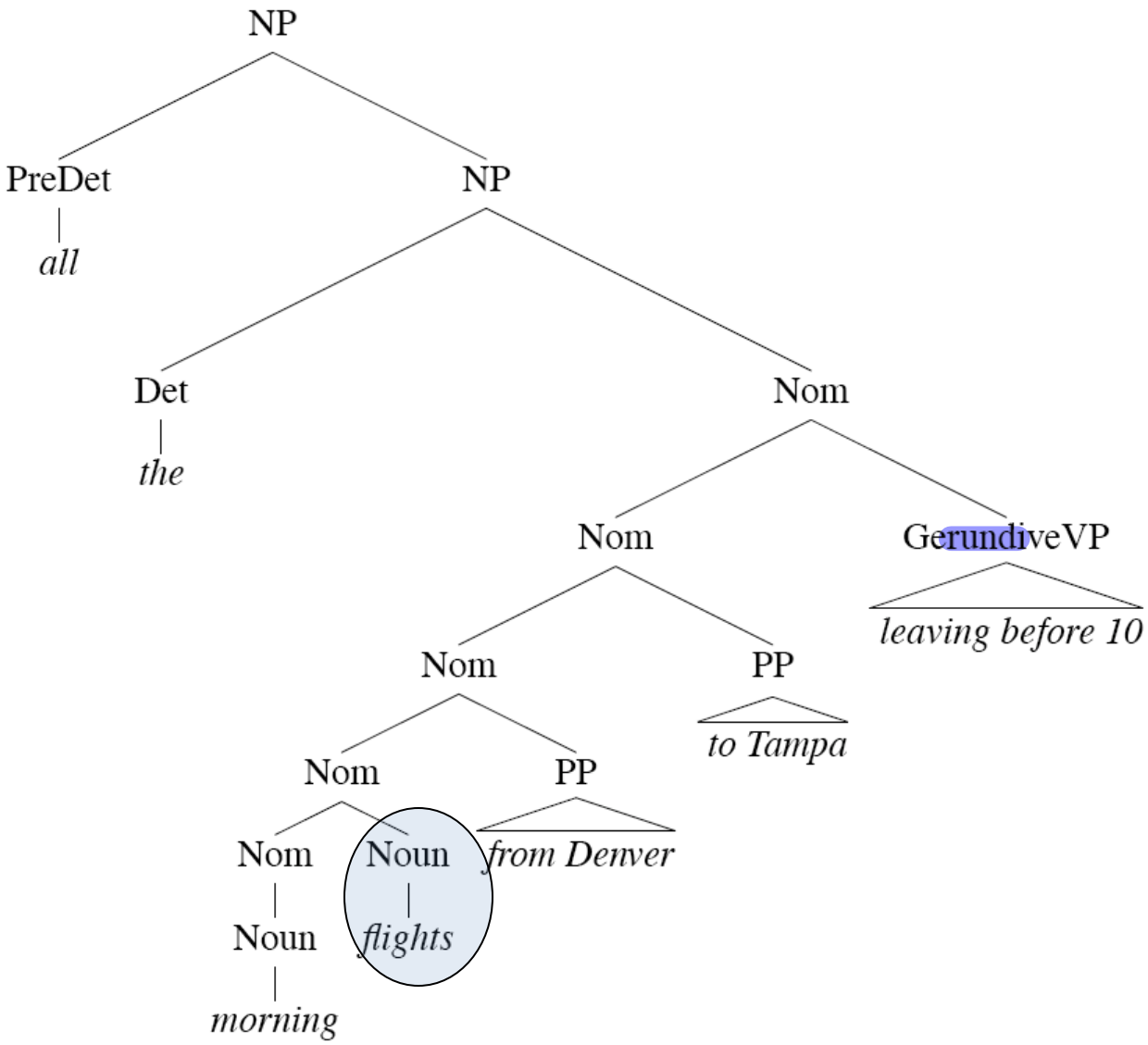
Productions of CFG

- A CFG can be thought of in two ways:
  - a device for generating sentences
    (Derivation)
  - a device for assigning a structure to a given sentence.

- Some rules for noun phrases:

$$
\begin{aligned}
NP &\rightarrow Det\ Nominal \\
NP &\rightarrow ProperNoun \\
Nominal &\rightarrow Noun \mid Nominal\ Noun
\end{aligned}
$$

Noun Phrases



Nominals

- Contain the head and any pre- and post- modifiers of the head.
  - Pre-
    - Quantifiers, cardinals, ordinals...
      - *Three cars*
    - Adjectives
      - *large cars*

Postmodifiers

- Three kinds
  - Prepositional phrases
    - *From Seattle*
  - Non-finite clauses
    - *Arriving before noon*
  - Relative clauses
    - *That serve breakfast*
- Same general (recursive) rules to handle these
  - *Nominal → Nominal PP*
  - *Nominal → Nominal GerundVP*
  - *Nominal → Nominal RelClause*

Verb Phrases

- English *VP*s consist of a verb (the head) along with 0 or more *following* constituents which we'll call *arguments*.

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

Subcategorization

- Even though there are many valid VP rules in English, not all verbs are allowed to participate in all those VP rules.
- We can *subcategorize* the verbs in a language according to the sets of VP rules that they participate in.
- This is just an elaboration on the traditional notion of transitive/intransitive.
- Modern grammars have many such classes

Subcategorization

- Sneeze: John sneezed
- Find: Please find [a flight to NY]$_{NP}$
- Give: Give [me]$_{NP}$[a cheaper fare]$_{NP}$
- Help: Can you help [me]$_{NP}$[with a flight]$_{PP}$
- Prefer: I prefer [to leave earlier]$_{TO\text{-}VP}$
- Told: I was told [United has a flight]$_{S}$
- …

Generative Grammar

- The use of formal languages to model Generative natural languages is called ***generative grammar*** since the language is defined by the set of possible sentences "generated" by the grammar.
- You can view these rules as either analysis or synthesis engines
  - Generate strings in the language
  - Reject strings not in the language
  - Assign structures (trees) to strings in the language

L0 Grammar

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

Sentence Types

- Declaratives: *A plane left.*

  $S \rightarrow NP\ VP$

- Imperatives: *Leave!*

  $S \rightarrow VP$

- Yes-No Questions: *Did the plane leave?*

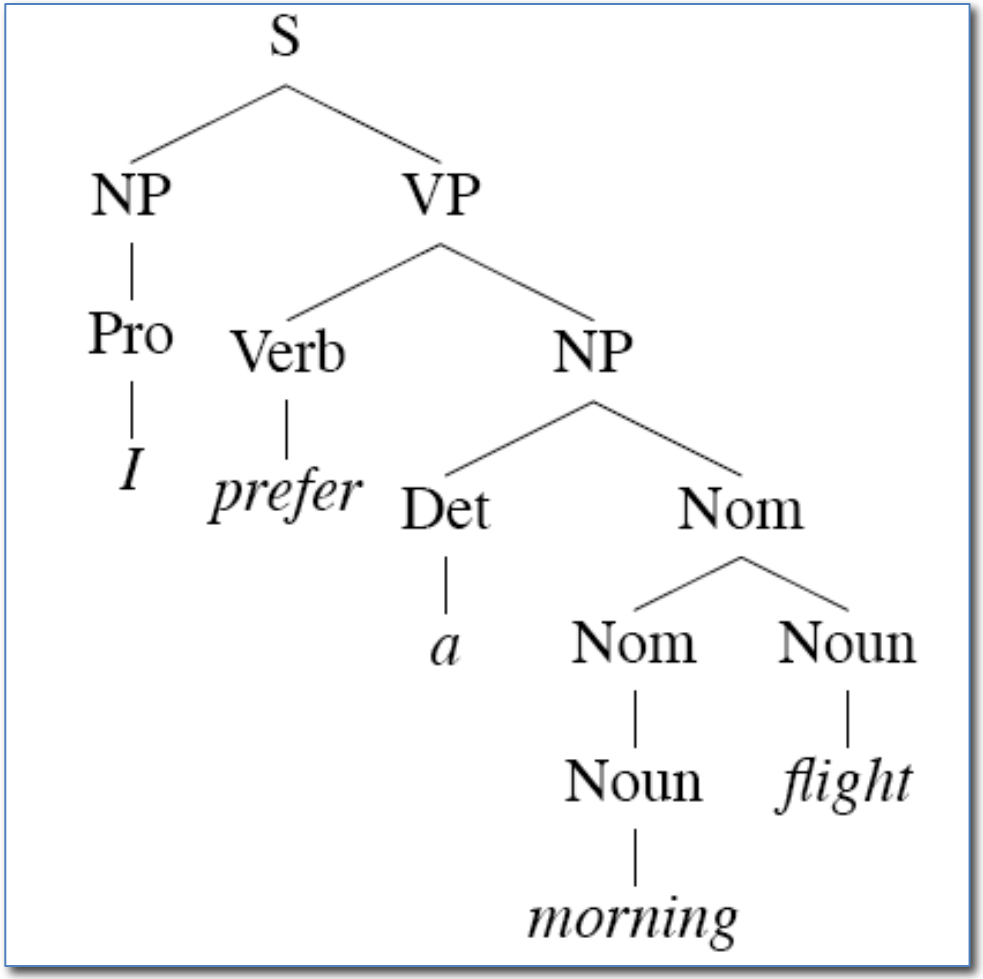
  $S \rightarrow Aux\ NP\ VP$

- WH Questions: *When did the plane leave?*

  $S \rightarrow WH\text{-}NP\ Aux\ NP\ VP$

Derivations

- A *derivation* is a sequence of rules applied to a string that *accounts* for that string
  - Covers all the elements in the string
  - Covers only the elements in the string



Parsing

- Parsing is the process of taking a string and a grammar and returning parse tree(s) for that string