UNIB20005

Language and Computation

Analyzing Sentence Structure

Steven Bird, Department of Computing and Information Systems

Review: Syntactic Structure
(from Lesley Stirling's lectures)

- All languages structure information in terms of: kinds of things; their properties; the actions and relationships they are involved in
- they express these relationships formally in a great variety of ways
- English: a "word-order" scheme: the words that constitute a description of an entity appear as a block together; relative order of these blocks with respect to words denoting actions and relationships (verbs) indicates the role of the entities within the event / situation
- A syntactic constituent is a word or a group of words that functions as a single unit within a syntactic structure
    - Evidence: English question formation; Fodor, Bever & Garrett click experiments
- Tests for constituent structure

Syntactic categories

Symbol	Meaning	Example
S	sentence	*the man walked*
NP	noun phrase	*a dog*
VP	verb phrase	*saw a park*
PP	prepositional phrase	*with a telescope*
Det	determiner	*the*
N	noun	*dog*
V	verb	*walked*
P	preposition	*in*

- each syntactically constituent has a type
- based on its *context* and its *content*
- see Syntax 6.3

The Infiniteness of Language

Recursion in Syntactic Structure

(1) a. Usain Bolt broke the 100m record.

b. The Jamaica Observer reported that Usain Bolt broke the 100m record.

c. Andre said The Jamaica Observer reported that Usain Bolt broke the 100m record.

d. I think Andre said the Jamaica Observer reported that Usain Bolt broke the 100m record.

- The Jamaica Observer reported that *S*
- Andre said *S*
- I think *S*

The Infiniteness of Language

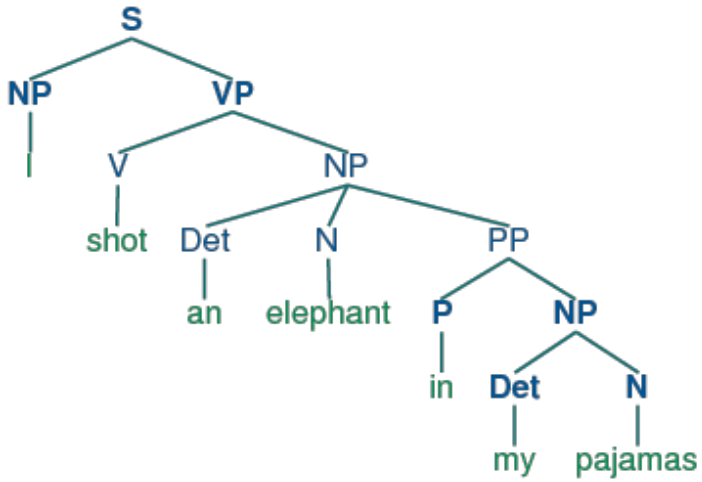
Previously unseen sentences

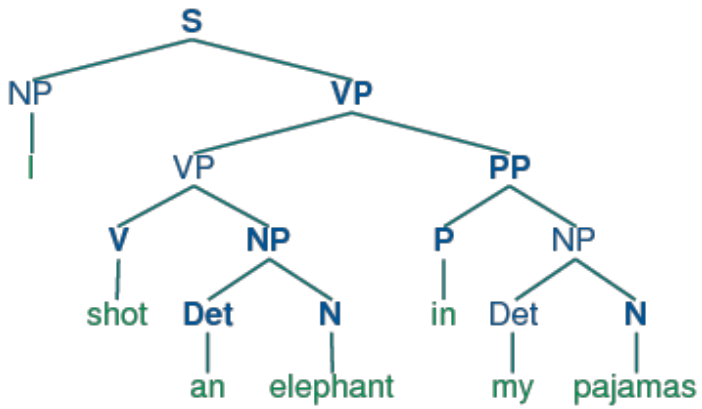
[You can imagine Piglet's joy when at last the ship came in sight of him.]
In after-years he liked to think that he had been in Very Great Danger during the Terrible Flood, but the only danger he had really been in was the last half-hour of his imprisonment, when Owl, who had just flown up, sat on a branch of his tree to comfort him, and told him a very long story about an aunt who had once laid a seagull's egg by mistake, and the story went on and on, rather like this sentence, until Piglet who was listening out of his window without much hope, went to sleep quietly and naturally, slipping slowly out of the window towards the water until he was only hanging on by his toes, at which moment, luckily, a sudden loud squawk from Owl, which was really part of the story, being what his aunt said, woke the Piglet up and just gave him time to jerk himself back into safety and say, "How interesting, and did she?" when — well, you can imagine his joy when at last he saw the good ship, Brain of Pooh (Captain, C. Robin; 1st Mate, P. Bear) coming over the sea to rescue him...

The Infiniteness of Language

Ubiquitous ambiguity

- While hunting in Africa, I shot an elephant in my pajamas. How an elephant got into my pajamas I'll never know.





Sequential Models of Language

The problem with ngram models

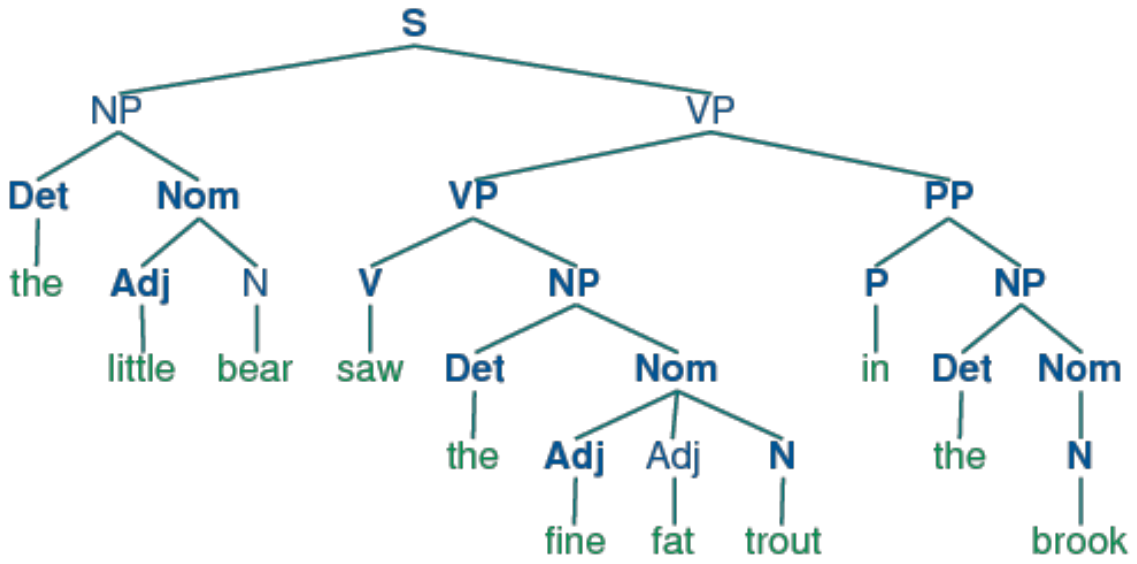
- random text generated by bigram model (aka conditional frequency distribution) -- cf `generate` function in textbook chapter 1.
- green pill vs green frog / w_s w_1 w_2 w_3 w_t
- NP V NP PP
- this is why we must consider hierarchical structure (not just sequential structure), when building computational models of language
- how do we work out the hierarchical structure?
  OR: what are the “constituents”?

Constituent structure: substitution tests

the	little	bear	saw	the	fine	fat	trout	in	the	brook
the	bear		saw	the	trout			in	it	
He			saw	it				there		
He			ran					there		
He			ran							

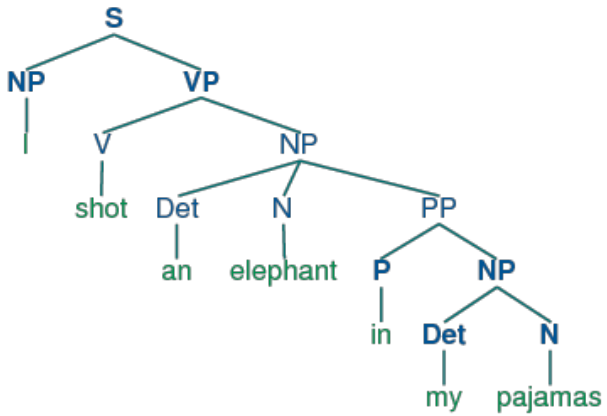
Phrase Structure Trees

Det the	Adj little	N bear	V saw	Det the	Adj fine	Adj fat	N trout	P in	Det the	N brook
Det the	Nom bear		V saw	Det the	Nom trout			P in	NP it	
NP He			V saw	NP it				PP there		
NP He			VP ran					PP there		
NP He			VP ran							

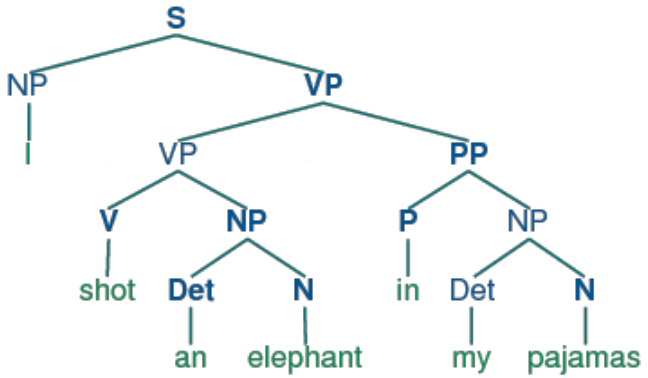


Text notations: $[_S\ [_{NP}\ \ldots\ ]_{NP}\ \ldots\ ]_S$ (S (NP ...) ...)

Phrase Structure Trees




```
(S
  (NP I)
  (VP
    (V shot)
    (NP (Det an) (N elephant) (PP (P in) (NP (Det my) (N pajamas))))))
```




```
(S
  (NP I)
  (VP
    (VP (V shot) (NP (Det an) (N elephant)))
    (PP (P in) (NP (Det my) (N pajamas)))))
```

Grammars

Textbook chapter 8.3

- a multiline string in a particular format
  start symbol, productions, non-terminals, terminals
  (see Syntax 6.5)

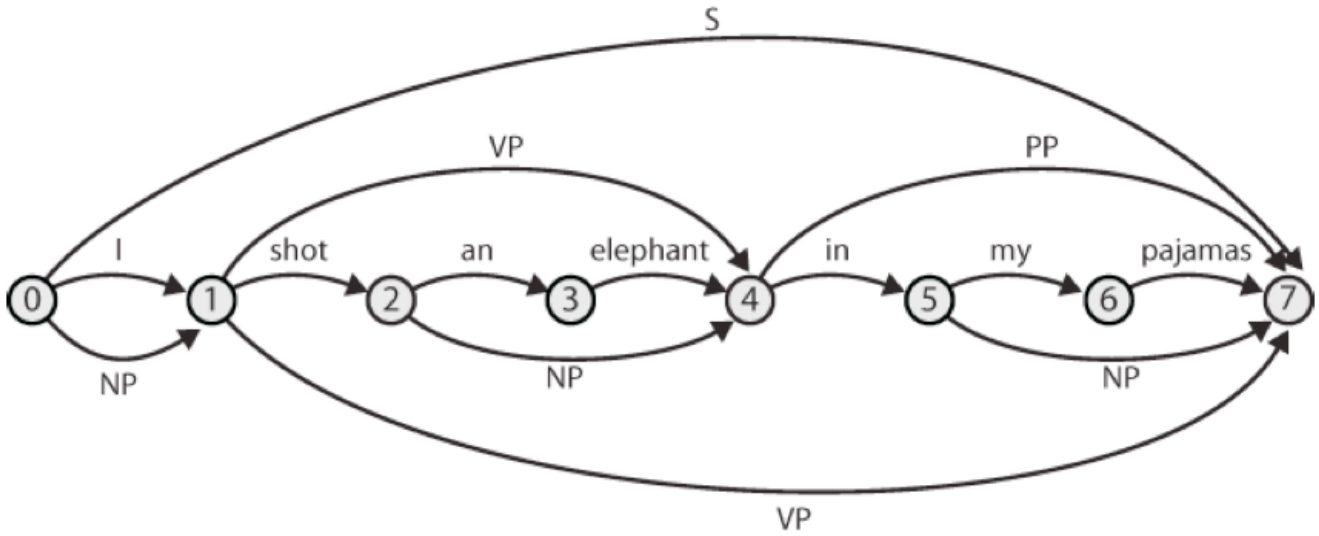
```
S -> NP VP
PP -> P NP
NP -> Det N | Det N PP | 'I'
VP -> V NP | VP PP
Det -> 'an' | 'my'
N -> 'elephant' | 'pajamas'
V -> 'shot'
P -> 'in'
```

- illustration: manual generation from a grammar

Simple Parsers, Demonstrations

Textbook chapter 8.4

- Top-Down Parsers
- Bottom-up Parsers
- Chart Parsers



Syntax: valency

- verb complements
- intransitive, transitive, ditransitive, sentential complements
- more specific verb classes (Levin classes)
- arguments vs adjuncts
- selectional restrictions
- count vs mass nouns
- stative vs nonstative adjectives
- see Syntax 6.4

VP -> V Adj	*was*
VP -> V NP	*saw*
VP -> V S	*thought*
VP -> V NP PP	*put*

Syntactic Agreement

Textbook chapter 9.1

(1) a. this dog
b. *these dog

(2) a. these dogs
b. *this dogs

(3) a. the dog runs
b. *the dog run

(4) a. the dogs run
b. *the dogs runs

(7)
```
S   ->   NP VP
NP  ->   Det N
VP  ->   V

Det  ->  'this'
N    ->  'dog'
V    ->  'runs'
```

Syntactic Agreement (cont)

	Singular	Plural
1st person	*I run*	*we run*
2nd person	*you run*	*you run*
3rd person	*he/she/it runs*	*they run*

(5)
```
the dog      run-s
    dog.3.SG run-3.SG
```

(6)
```
the dog-s    run
    dog.3.PL run-3.PL
```

(8)
```
S -> NP_SG VP_SG
S -> NP_PL VP_PL
NP_SG -> Det_SG N_SG
NP_PL -> Det_PL N_PL
VP_SG -> V_SG
VP_PL -> V_PL

Det_SG -> 'this'
Det_PL -> 'these'
N_SG -> 'dog'
N_PL -> 'dogs'
V_SG -> 'runs'
V_PL -> 'run'
```

(11)
```
S -> NP[NUM=?n] VP[NUM=?n]
NP[NUM=?n] -> Det[NUM=?n] N[NUM=?n]
VP[NUM=?n] -> V[NUM=?n]
```

(10)
```
Det[NUM=sg] -> 'this'
Det[NUM=pl] -> 'these'

N[NUM=sg] -> 'dog'
N[NUM=pl] -> 'dogs'
V[NUM=sg] -> 'runs'
V[NUM=pl] -> 'run'
```

Syntax summary

- constituency and trees
- syntactic ambiguity and lexical ambiguity
- grammars and parsers
- top-down and bottom-up parsers
- corpus-based and grammar-based models of language
- Homework: Read textbook chapter 8;
  Do exercises in syntax handout