Introduction to Syntax and Context-Free Grammars

http://www1.cs.columbia.edu/~rambow/teaching/lecture-2009-09-22.ppt

Owen Rambow

rambow@ccls.columbia.edu

Slides with contributions from Kathy McKeown, Dan Jurafsky and James Martin

Announcements

- Talks
  - Information Extraction, Data Mining and Joint Inference, Prof. Andrew McCallum, Univ. of Massachusetts, 11AM Wed. Oct. 1st, Davis Auditorium, Schapiro
  - Integrity of Elections, Dr. Peter G. Neumann, SRI International, 11 AM Mon. Oct. 6th, Davis Auditorium, Schapiro

What is Syntax?

- Study of structure of language
- Refers to the way words are arranged together, and the relationship between them.
- Roughly, goal is to relate surface form (what we perceive when someone says something) to semantics (what that utterance means)

What is Syntax Not?

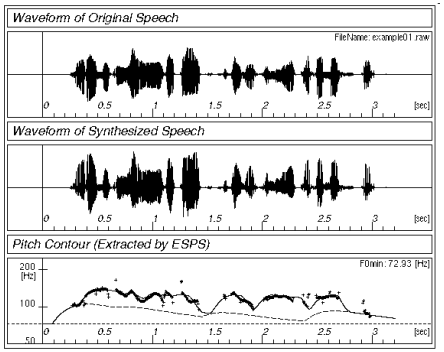
- Phonology: study of sound systems and how sounds combine
- Morphology: study of how words are formed from smaller parts (morphemes)
- Semantics: study of meaning of language

What is Syntax? (2)

- Study of structure of language
- Specifically, goal is to relate an interface to morphological component to an interface to a semantic component
- Note: interface to morphological component may look like written text
- Representational device is **tree structure**

Simplified View of Linguistics

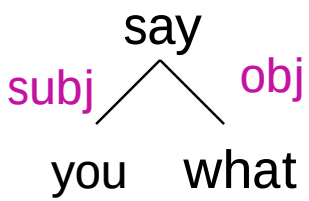
Phonology



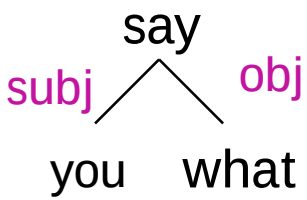
⇔ /waddyasai/

Morphology /waddyasai/ ⇔ what did you say

Syntax what did you say ⇔



Semantics



⇔ P[λx. say(you, x)]

The Big Picture

Empirical Matter

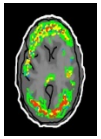
?

Formalisms

- Data structures
- Formalisms (e.g., CFG)
- Algorithms
- Distributional Models

?

Maud expects there to be a riot
*Teri promised there to be a riot
Maud expects the shit to hit the fan
*Teri promised the shit to hit the fan



?

?

Linguistic Theory

What About Chomsky?

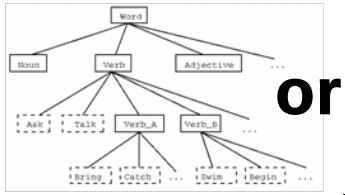
- At birth of formal language theory (comp sci) and formal linguistics
- Major contribution: syntax is **cognitive** reality
- Humans able to learn languages quickly, but not all languages $\Rightarrow$ **universal grammar** is biological
- Goal of syntactic study: find universal **principles and** language-specific **parameters**
- Specific Chomskyan theories change regularly
- General ideas adopted by almost all contemporary syntactic theories ("principles-and-parameters-type theories")

Types of Linguistic Theories

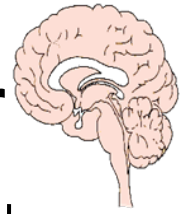
- **Prescriptive**: "prescriptive linguistics" is an oxymoron
  - **Prescriptive grammar:** how people ought to talk
- **Descriptive**: provide account of syntax of a language
  - **Descriptive grammar**: how people do talk
  - often appropriate for NLP engineering work
- **Explanatory**: provide principles-and-parameters style account of syntax of (preferably) several languages

The Big Picture

Empirical Matter



or

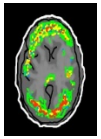


Formalisms

- Data structures
- Formalisms
- Algorithms
- Distributional Models

?

Maud expects there to be a riot
*Teri promised there to be a riot
Maud expects the shit to hit the fan
*Teri promised the shit to hit the



?

?

Linguistic Theory

Syntax: Why should we care?

- Grammar checkers
- Question answering
- Information extraction
- Machine translation

key ideas of syntax

- Constituency (we'll spend most of our time on this)
- Subcategorization
- Grammatical relations
- Movement/long-distance dependency

Structure in Strings

- Some words: *the a small nice big very boy girl sees likes*
- Some good sentences:
  - the boy likes a girl
  - the small girl likes the big girl
  - a very small nice boy sees a very nice boy
- Some bad sentences:
  - *the boy the girl
  - *small boy likes nice girl
- Can we find subsequences of words (**constituents**) which in some way behave alike?

Structure in Strings
Proposal 1

- Some words: *the a small nice big very boy girl sees likes*
- Some good sentences:
  - (the) boy (likes a girl)
  - (the small) girl (likes the big girl)
  - (a very small nice) boy (sees a very nice boy)
- Some bad sentences:
  - *(the) boy (the girl)
  - *(small) boy (likes the nice girl)

Structure in Strings
Proposal 2

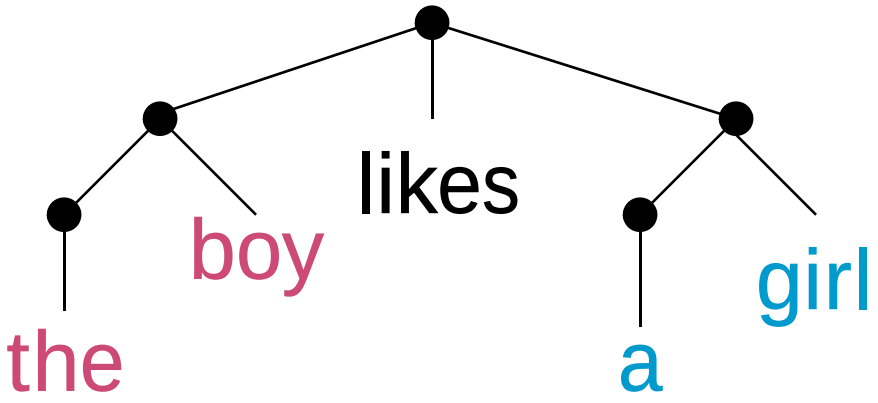
- Some words: *the a small nice big very boy girl sees likes*
- Some good sentences:
  - (the boy) likes (a girl)
  - (the small girl) likes (the big girl)
  - (a very small nice boy) sees (a very nice boy)
- Some bad sentences:
  - *(the boy) (the girl)
  - *(small boy) likes (the nice girl)
- This is better proposal: fewer types of constituents
  (blue and red are of same type)

More Structure in Strings
Proposal 2 -- ctd

- Some words: *the a small nice big very boy girl sees likes*
- Some good sentences:
  - ((the) boy) likes ((a) girl)
  - ((the) (small) girl) likes ((the) (big) girl)
  - ((a) ((very) small) (nice) boy) sees ((a) ((very) nice) girl)
- Some bad sentences:
  - *((the) boy) ((the) girl)
  - *((small) boy) likes ((the) (nice) girl)

From Substrings to Trees

- (((the) boy) likes ((a) girl))

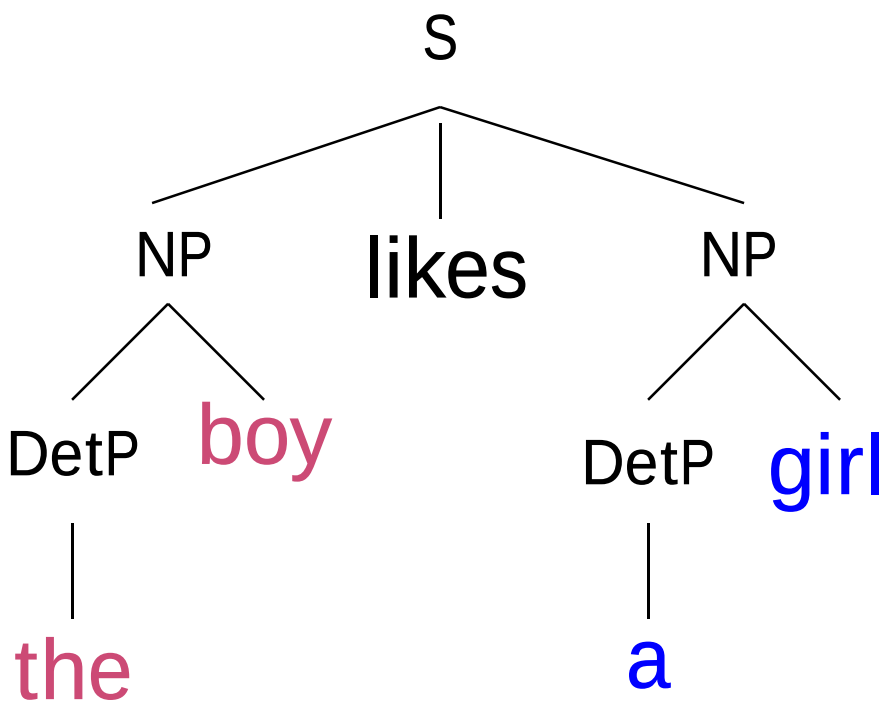


Node Labels?

- (((the) boy) likes ((a) girl))
- Choose constituents so each one has one non-bracketed word: the **head**
- Group words by distribution of constituents they head (part-of-speech, POS):
  - Noun (N), verb (V), adjective (Adj), adverb (Adv), determiner (Det)
- Category of constituent: XP, where X is POS
  - NP, S, AdjP, AdvP, DetP

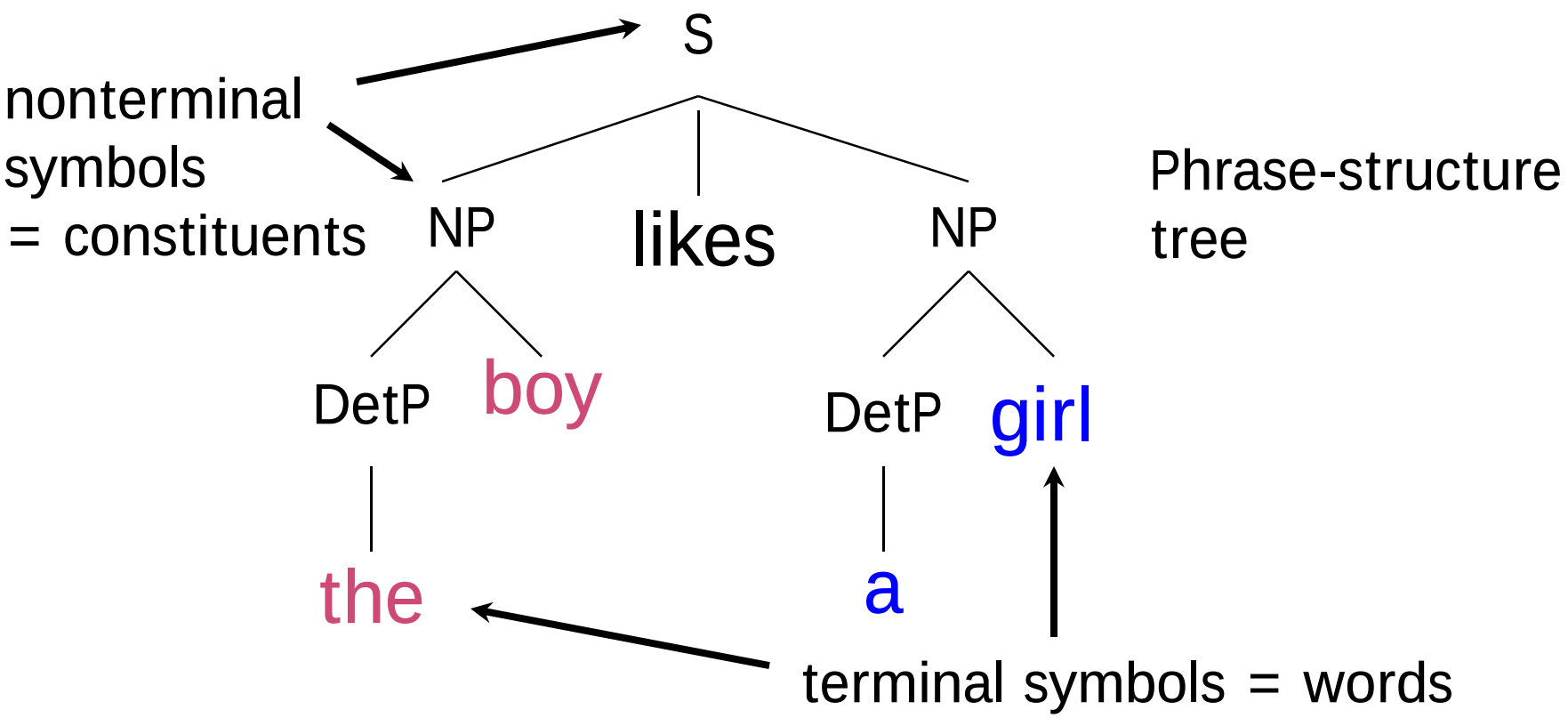
Node Labels

- (((the/Det) boy/N) likes/V ((a/Det) girl/N))



Types of Nodes

- (((the/Det) boy/N) likes/V ((a/Det) girl/N))



Determining Part-of-Speech

*A **blue** seat/a **child** seat:* noun or adjective?

– Syntax:
  - a **blue** seat a **child** seat
  - a very **blue** seat *a very **child** seat
  - this seat is **blue** *this seat is **child**

– Morphology:
  - bluer *childer

– **blue** and **child** are not the same POS

– **blue** is Adj, **child** is Noun

Determining Part-of-Speech (2)

– preposition or particle?

- A he threw **out** the garbage
- B he threw the garbage **out** the door

- A he threw the garbage **out**
- B *he threw the garbage the door **out**

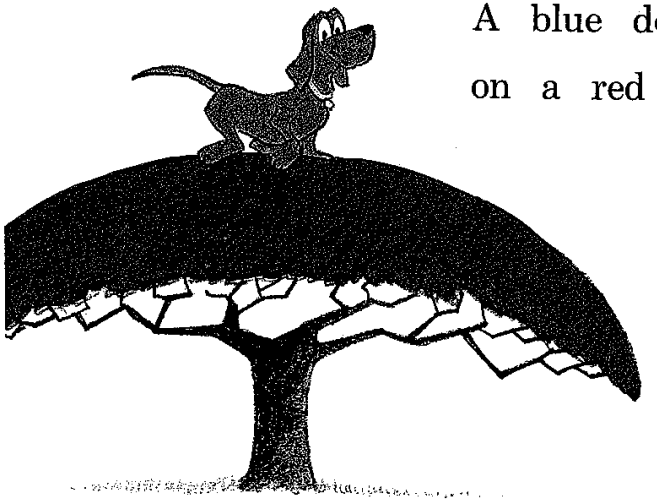
- The two **out** are not same POS; A is particle, B is Preposition

Constituency (Review)

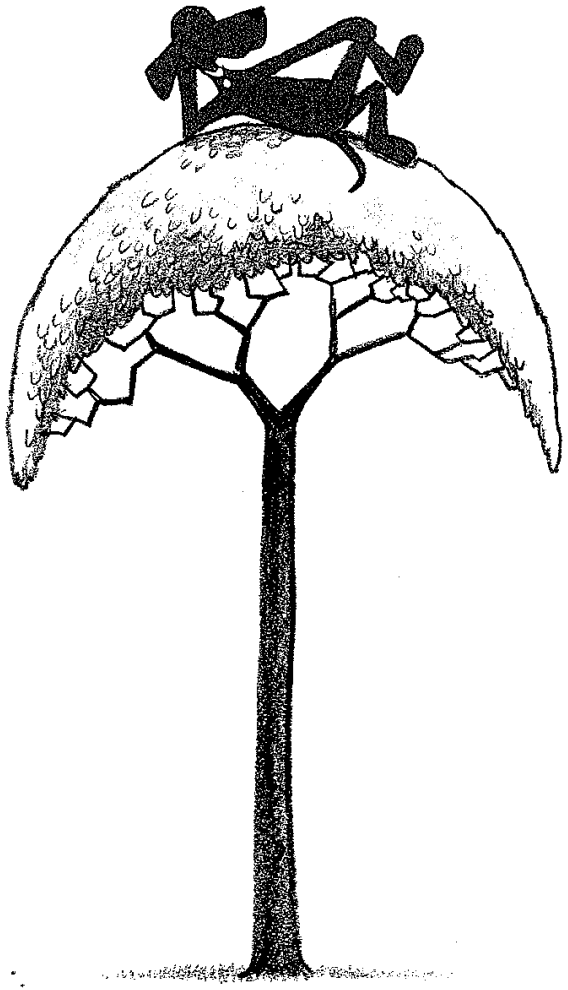
- E.g., Noun phrases (NPs)
  - A red dog on a blue tree
  - A blue dog on a red tree
  - Some big dogs and some little dogs
  - A dog
  - I
  - Big dogs, little dogs, red dogs, blue dogs, yellow dogs, green dogs, black dogs, and white dogs
- How do we know these form a constituent?

A red dog
on a blue tree.

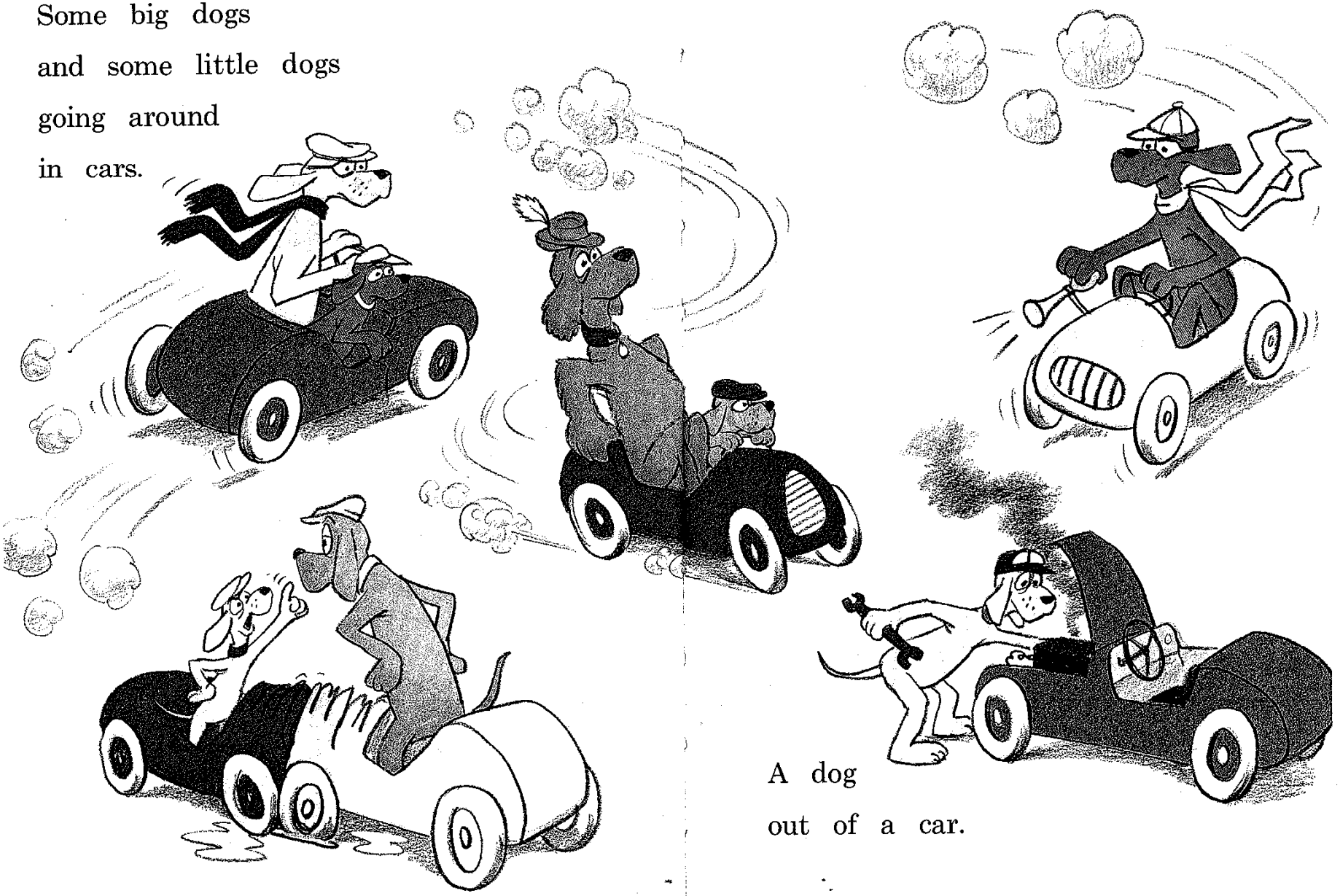
A blue dog
on a red tree.



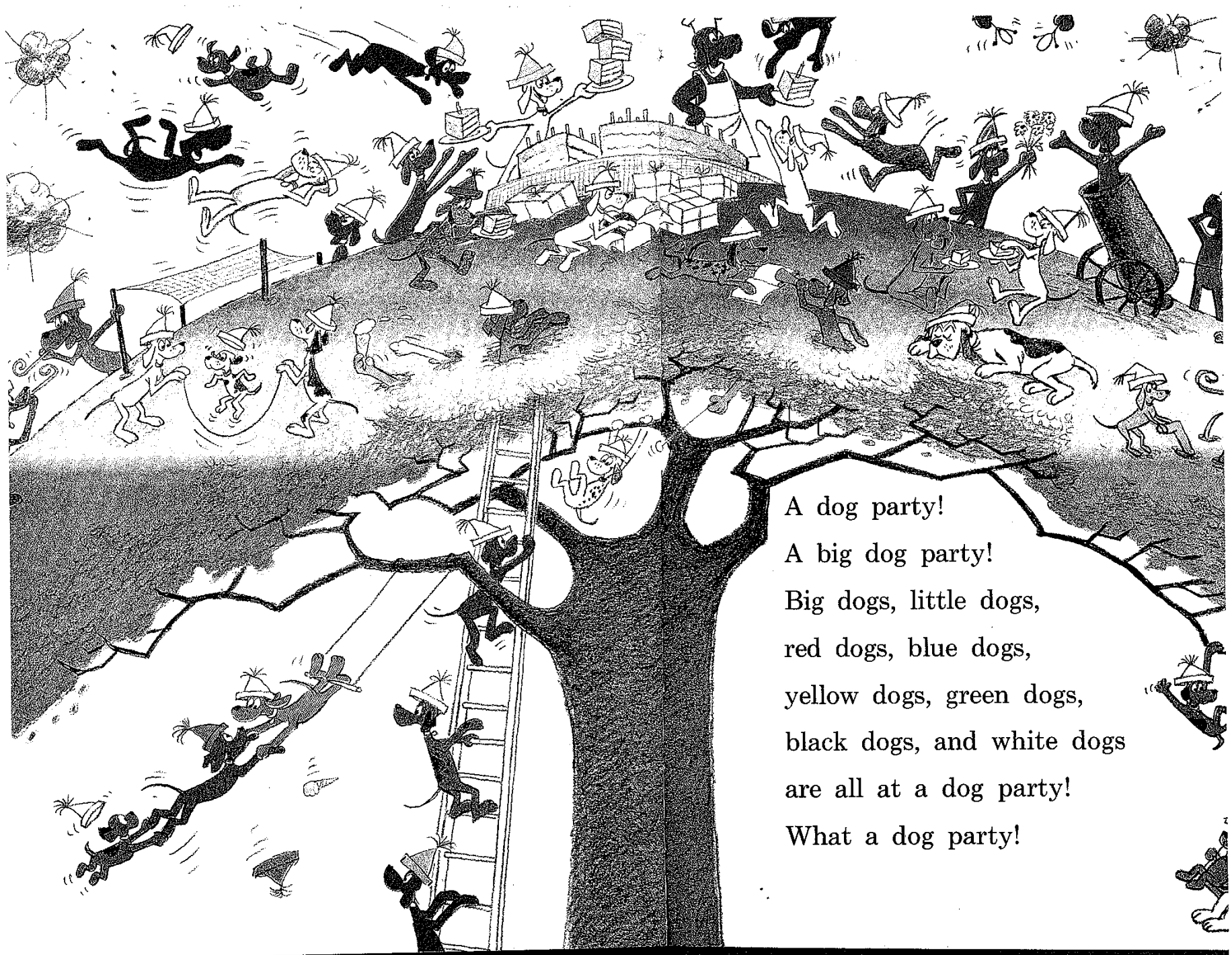
A green dog
on a yellow tree.



Some big dogs
and some little dogs
going around
in cars.



A dog
out of a car.



A dog party!
A big dog party!
Big dogs, little dogs,
red dogs, blue dogs,
yellow dogs, green dogs,
black dogs, and white dogs
are all at a dog party!
What a dog party!

Constituency (II)

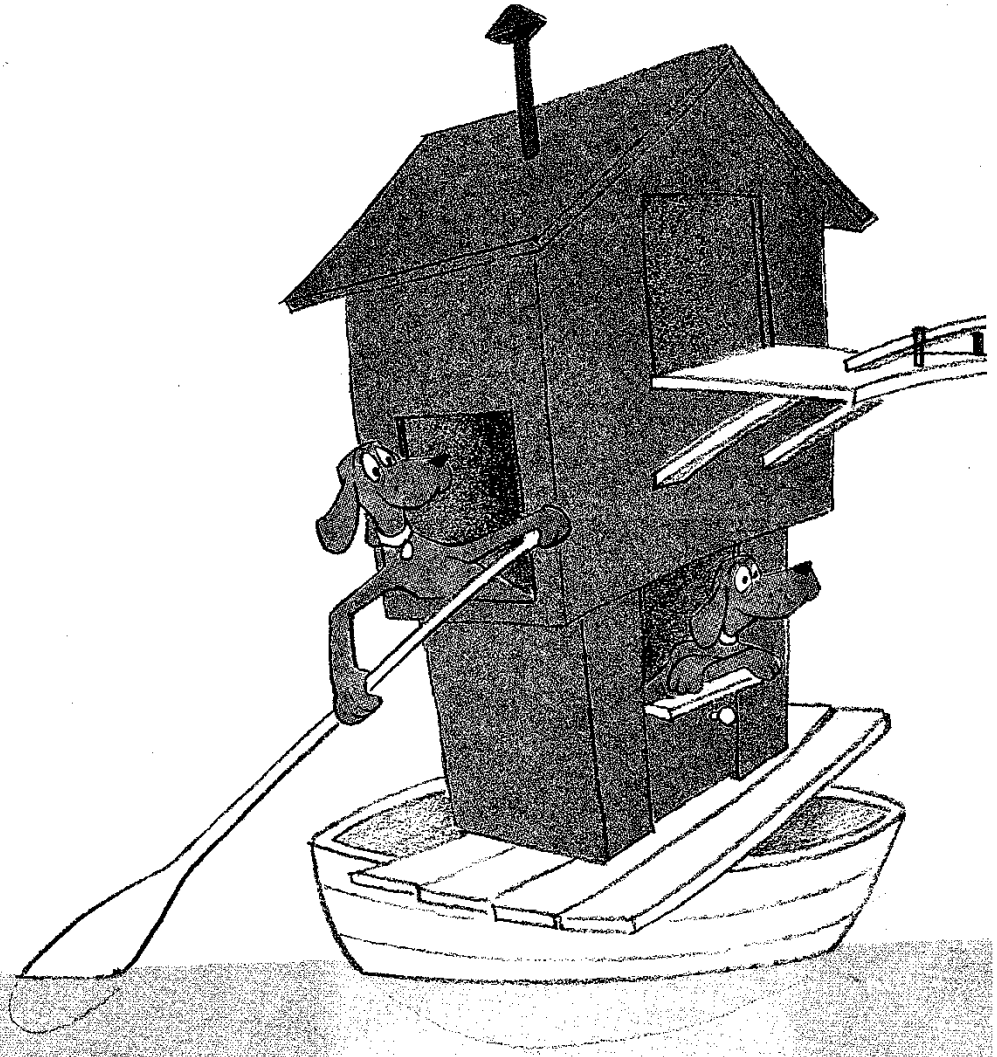
- They can all appear before a verb:
  - Some big dogs and some little dogs are going around in cars…
  - Big dogs, little dogs, red dogs, blue dogs, yellow dogs, green dogs, black dogs, and white dogs are all at a dog party!
  - I do not
- But individual words can't always appear before verbs:
  - *little are going…
  - *blue are…
  - *and are
- Must be able to state generalizations like:
  - Noun phrases occur before verbs

Constituency (III)

- Preposing and postposing:
  - Under a tree is a yellow dog.
  - A yellow dog is under a tree.

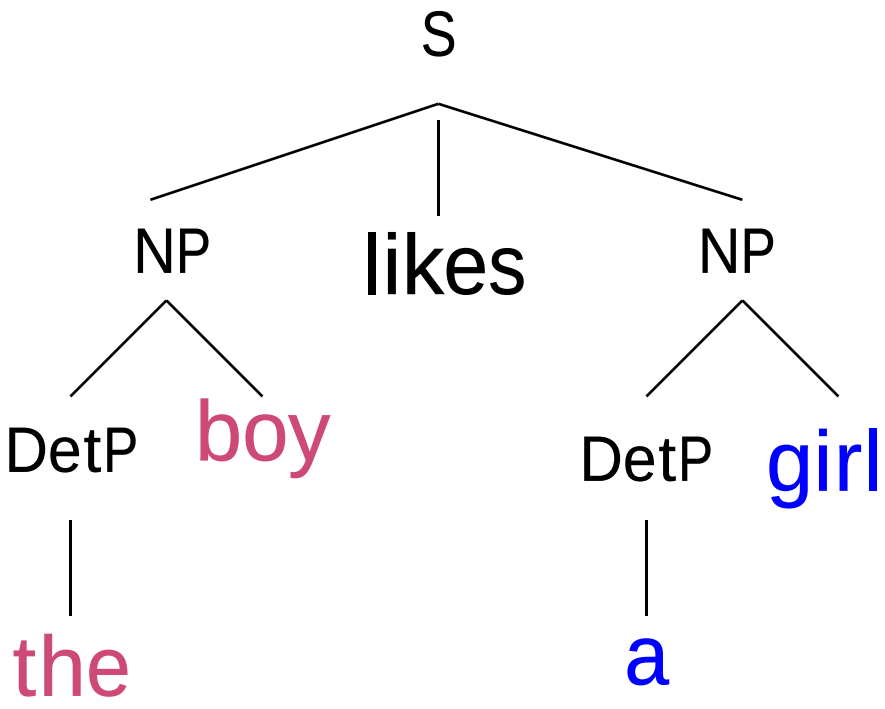
- But not:
  - *Under, is a yellow dog a tree.
  - *Under a is a yellow dog tree.

- Prepositional phrases notable for ambiguity in attachment

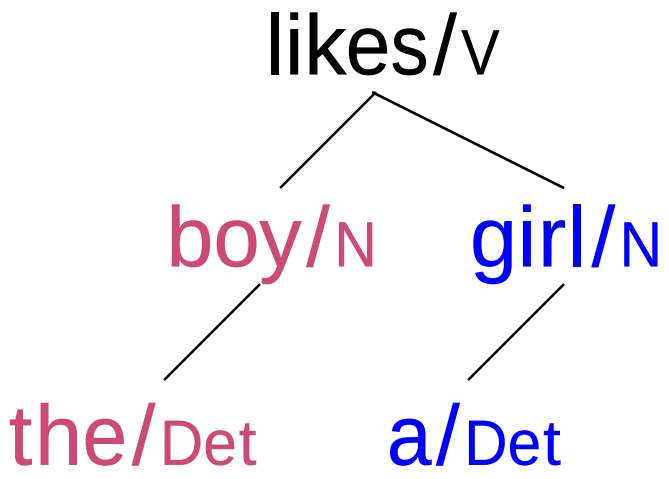


Two dogs
in a house
on a boat
in the water.

Phrase Structure and Dependency Structure

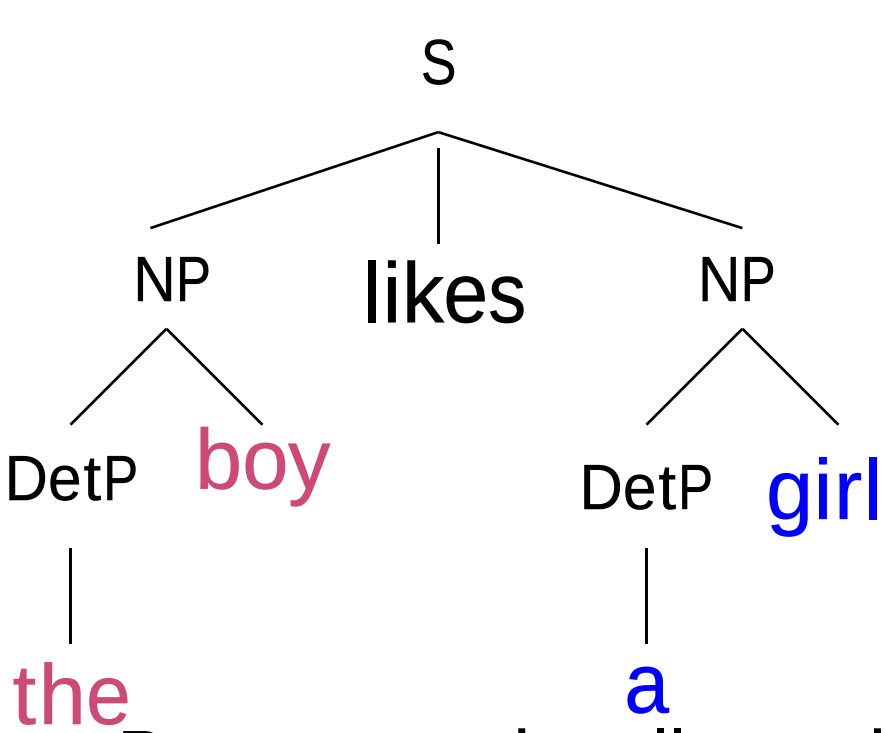


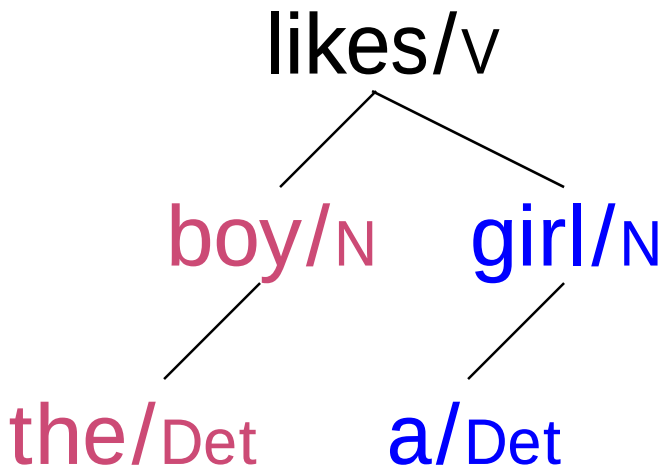
Only leaf nodes labeled with words!



All nodes are labeled with words!

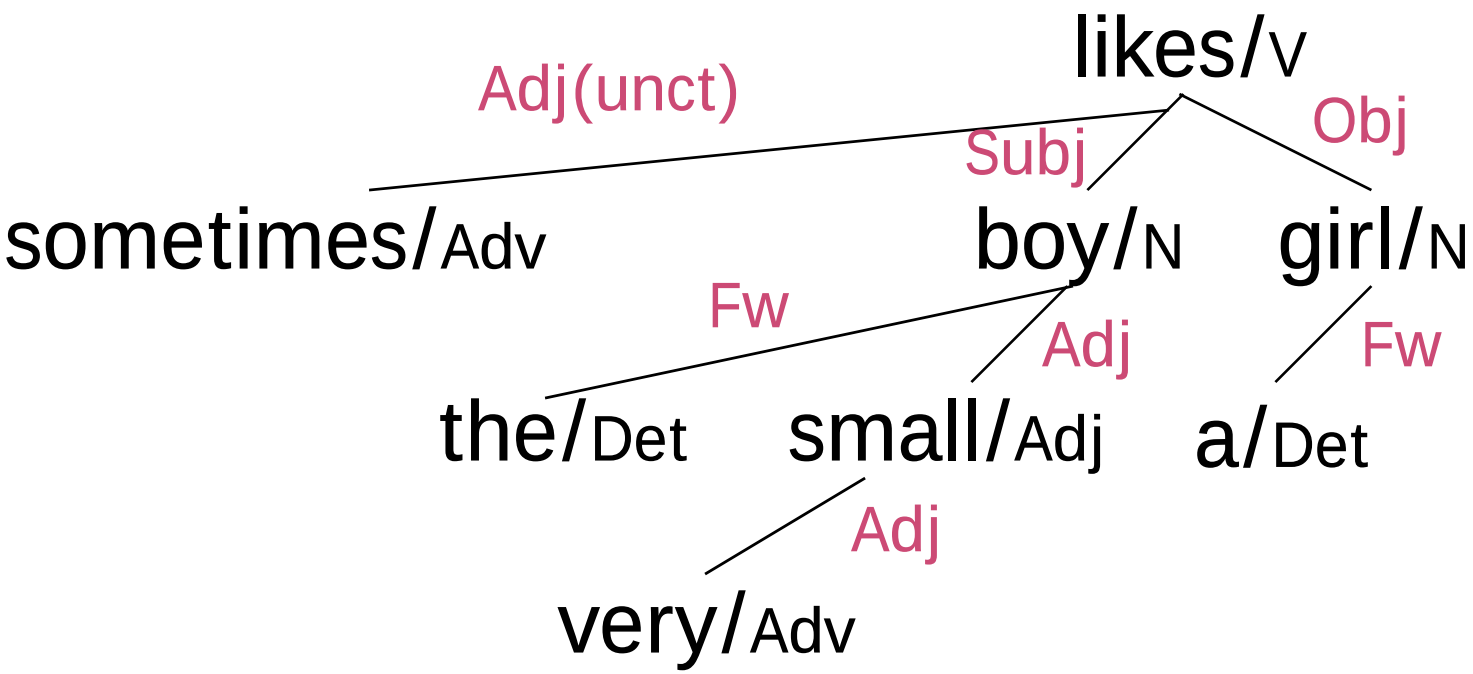
Phrase Structure and Dependency Structure (ctd)





Representationally equivalent if each nonterminal node has one lexical daughter (its head)

Types of Dependency



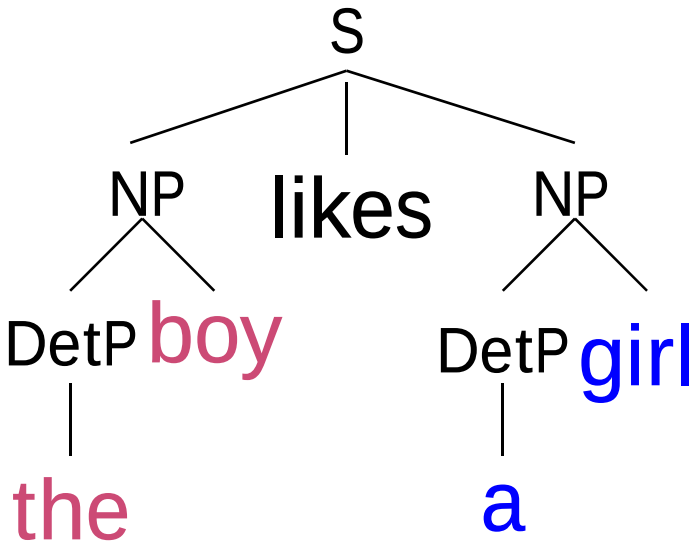
Grammatical Relations

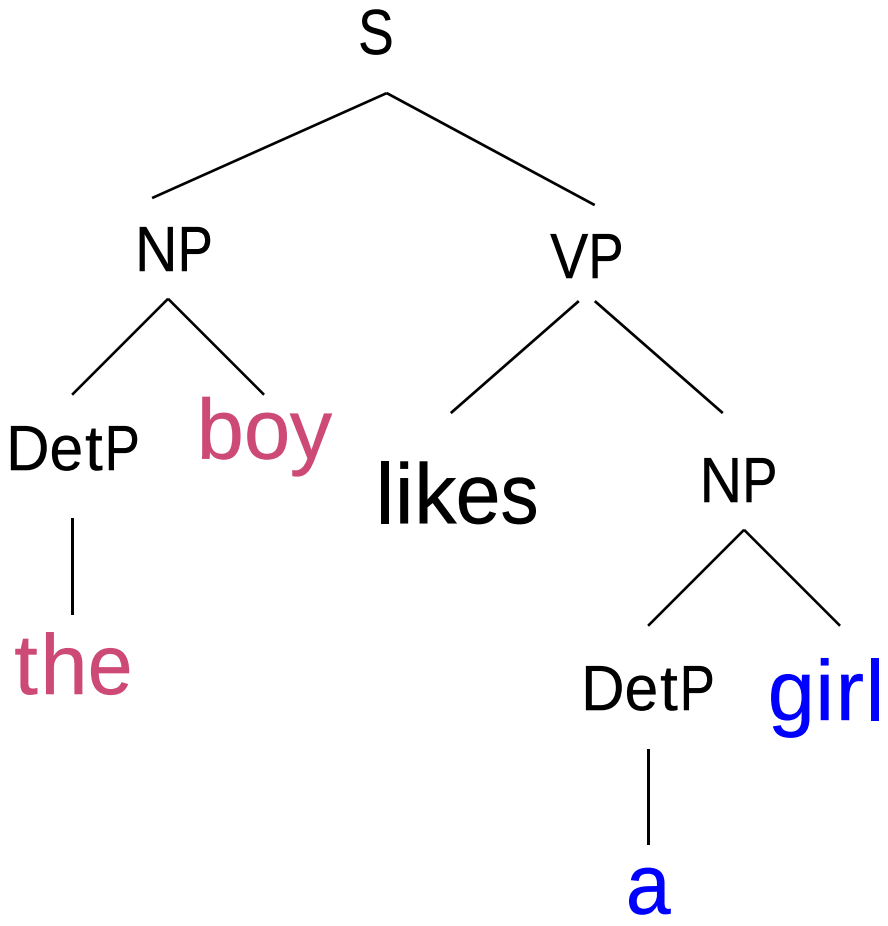
- Types of relations between words
  - Arguments: subject, object, indirect object, prepositional object
  - Adjuncts: temporal, locative, causal, manner, …
  - Function Words

Subcategorization

- List of arguments of a word (typically, a verb), with features about realization (POS, perhaps case, verb form etc)
- In canonical order Subject-Object-IndObj
- Example:
  - like: N-N, N-V(to-inf)
  - see: N, N-N, N-N-V(inf)
- Note: J&M talk about subcategorization only within VP

What About the VP?





What About the VP?

- Existence of VP is a linguistic (i.e., empirical) claim, not a methodological claim
- Semantic evidence???
- Syntactic evidence
  - VP-fronting (*and quickly clean the carpet he did!*)
  - VP-ellipsis (*He cleaned the carpets quickly, and so did she*)
  - Can have adjuncts before and after VP, but not in VP (*He often eats beans, *he eats often beans*)
- Note: VP cannot be represented in a dependency representation

Context-Free Grammars

- Defined in formal language theory (comp sci)
- Terminals, nonterminals, start symbol, rules
- String-rewriting system
- Start with start symbol, rewrite using rules, done when only terminals left
- NOT A LINGUISTIC THEORY, just a formal device

CFG: Example

- Many possible CFGs for English, here is an example (fragment):
  - S $\rightarrow$ NP VP
  - VP $\rightarrow$ V NP
  - NP $\rightarrow$ DetP N | AdjP NP
  - AdjP $\rightarrow$ Adj | Adv AdjP
  - N $\rightarrow$ boy | girl
  - V $\rightarrow$ sees | likes
  - Adj $\rightarrow$ big | small
  - Adv $\rightarrow$ very
  - DetP $\rightarrow$ a | the

the very small boy likes a girl

Derivations in a CFG

S

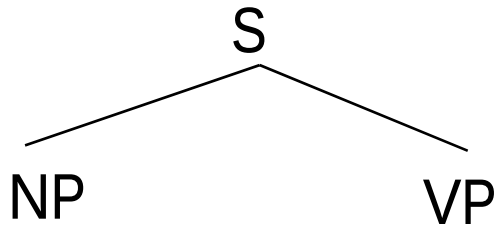
S → NP VP
VP → V NP
NP → DetP N | AdjP NP
AdjP → Adj | Adv AdjP
N → boy | girl
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the

S

Derivations in a CFG

NP VP

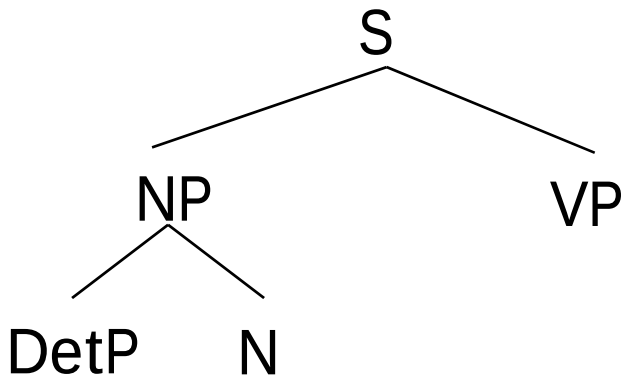
S → NP VP
VP → V NP
NP → DetP N | AdjP NP
AdjP → Adj | Adv AdjP
N → boy | girl
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the



Derivations in a CFG

DetP N VP

S → NP VP
VP → V NP
NP → DetP N | AdjP NP
AdjP → Adj | Adv AdjP
N → boy | girl
V → sees | likes
Adj → big | small
Adv → very
DetP → a | **the**



Derivations in a CFG

the boy VP

S → NP VP

VP → V NP

NP → DetP N | AdjP NP

AdjP → Adj | Adv AdjP

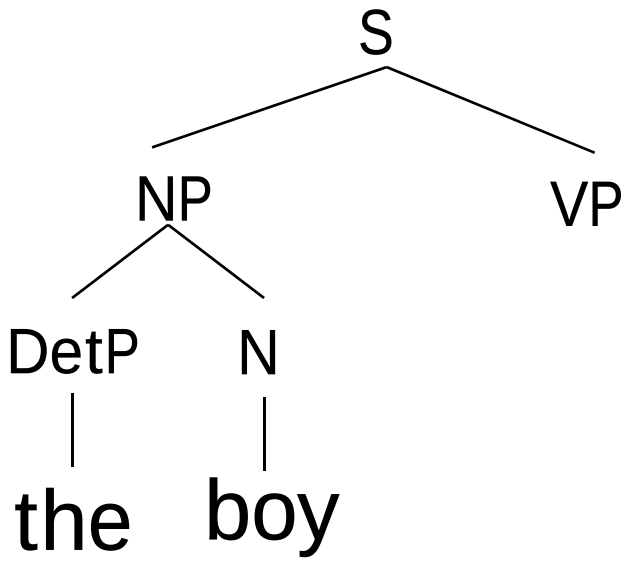
N → boy | girl

V → sees | **likes**

Adj → big | small

Adv → very

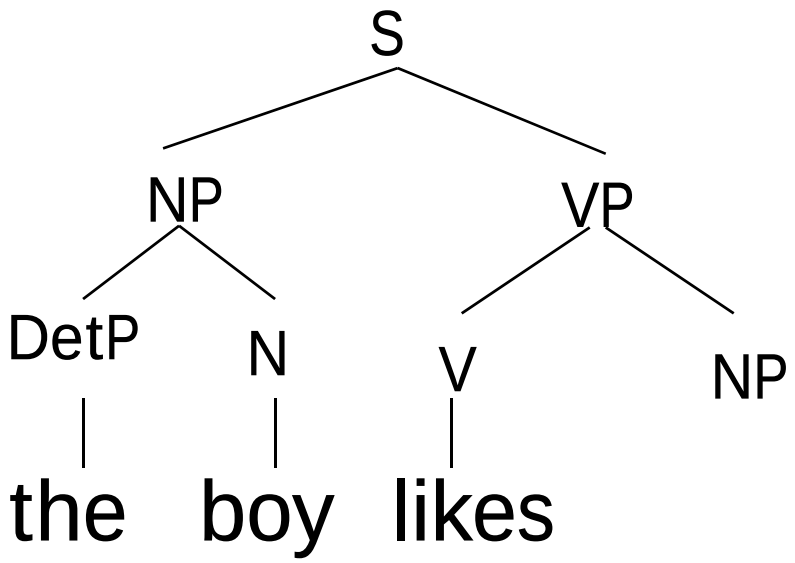
DetP → a | the



Derivations in a CFG

the boy likes NP

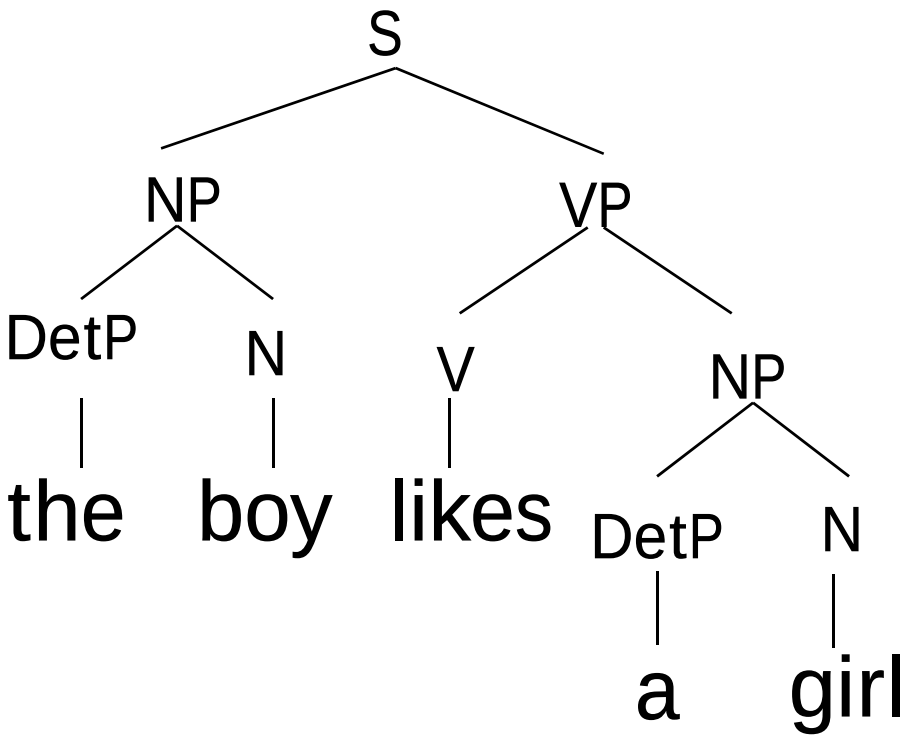
S → NP VP
VP → V NP
NP → DetP N | AdjP NP
AdjP → Adj | Adv AdjP
N → boy | **girl**
V → sees | likes
Adj → big | small
Adv → very
DetP → a | the



Derivations in a CFG

the boy likes a girl

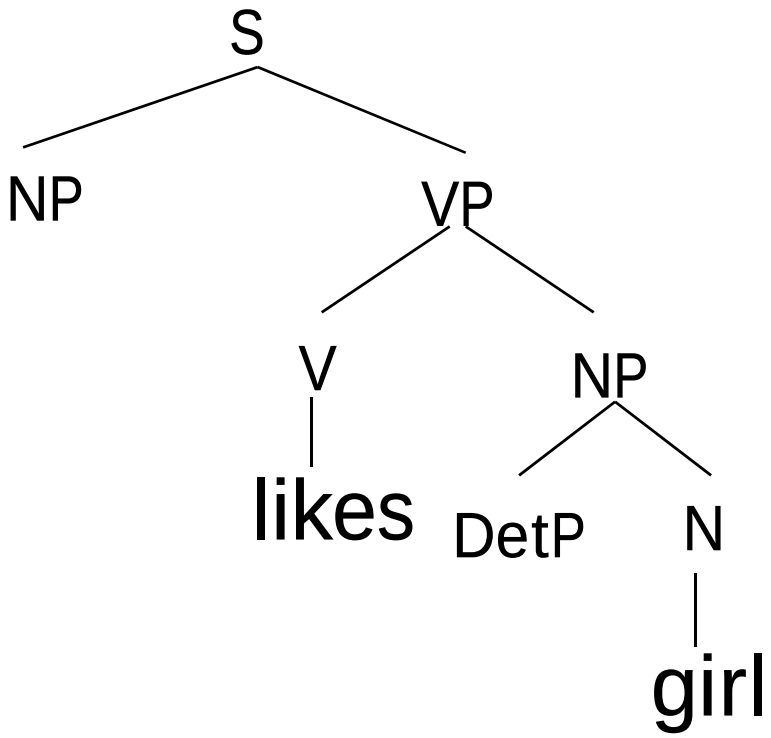
$S \rightarrow NP\ VP$
$VP \rightarrow V\ NP$
$NP \rightarrow DetP\ N \mid AdjP\ NP$
$AdjP \rightarrow Adj \mid Adv\ AdjP$
$N \rightarrow boy \mid girl$
$V \rightarrow sees \mid likes$
$Adj \rightarrow big \mid small$
$Adv \rightarrow very$
$DetP \rightarrow a \mid the$



Derivations in a CFG; Order of Derivation Irrelevant

NP likes DetP girl

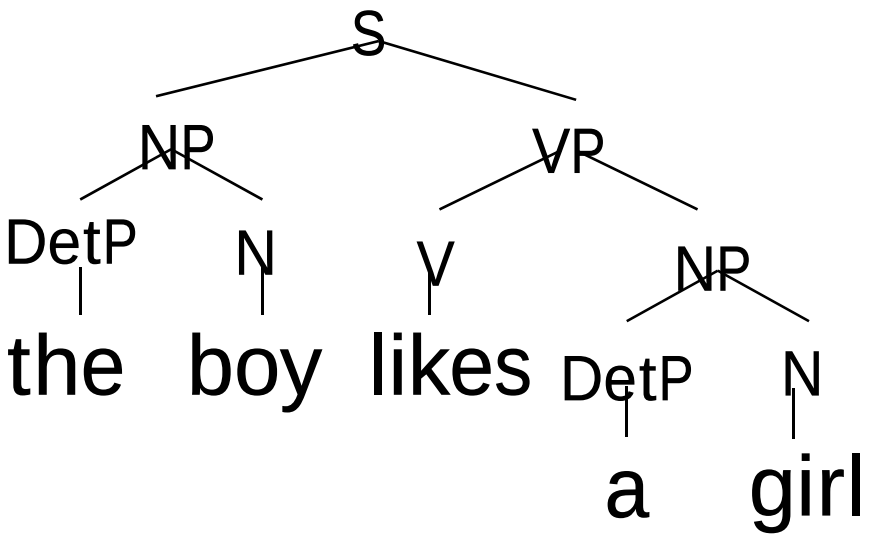
$S \rightarrow NP\ VP$
$VP \rightarrow V\ NP$
$NP \rightarrow DetP\ N \mid AdjP\ NP$
$AdjP \rightarrow Adj \mid Adv\ AdjP$
$N \rightarrow boy \mid girl$
$V \rightarrow sees \mid likes$
$Adj \rightarrow big \mid small$
$Adv \rightarrow very$
$DetP \rightarrow a \mid the$



Derivations of CFGs

- String rewriting system: we derive a string (=**derived** structure)
- But derivation history represented by phrase-structure tree (=**derivation** structure)!

the boy likes a girl



Formal Definition of a CFG

G = (V,T,P,S)

- V: finite set of nonterminal symbols
- T: finite set of terminal symbols, V and T are disjoint
- P: finite set of productions of the form
  $A \rightarrow \alpha$, $A \in V$ and $\alpha \in (T \cup V)^*$
- $S \in V$: start symbol

Context?

- The notion of context in CFGs has nothing to do with the ordinary meaning of the word context in language

- All it really means is that the non-terminal on the left-hand side of a rule is out there all by itself (free of context)

  A -> B C

  Means that I can rewrite an A as a B followed by a C regardless of the context in which A is found

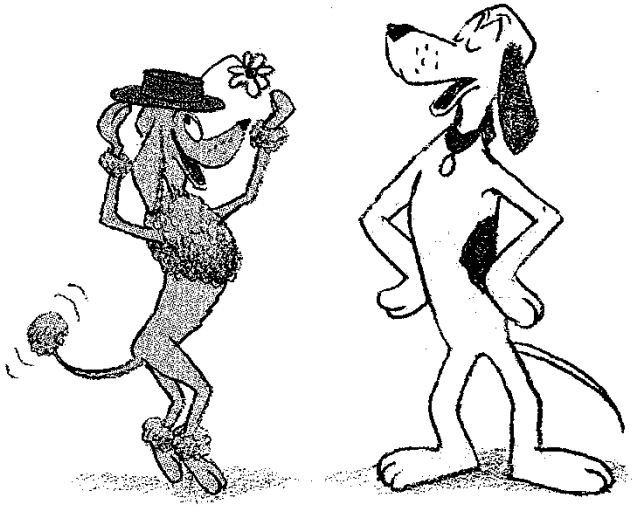
Key Constituents (English)

- Sentences
- Noun phrases
- Verb phrases
- Prepositional phrases

Sentence-Types

- Declaratives: I do not.
  S -> NP VP
- Imperatives: Go around again!
  S -> VP
- Yes-No Questions: Do you like my hat?
  S -> Aux NP VP
- WH Questions: What are they going to do?
  S -> WH Aux NP VP

"Do you like my hat?"

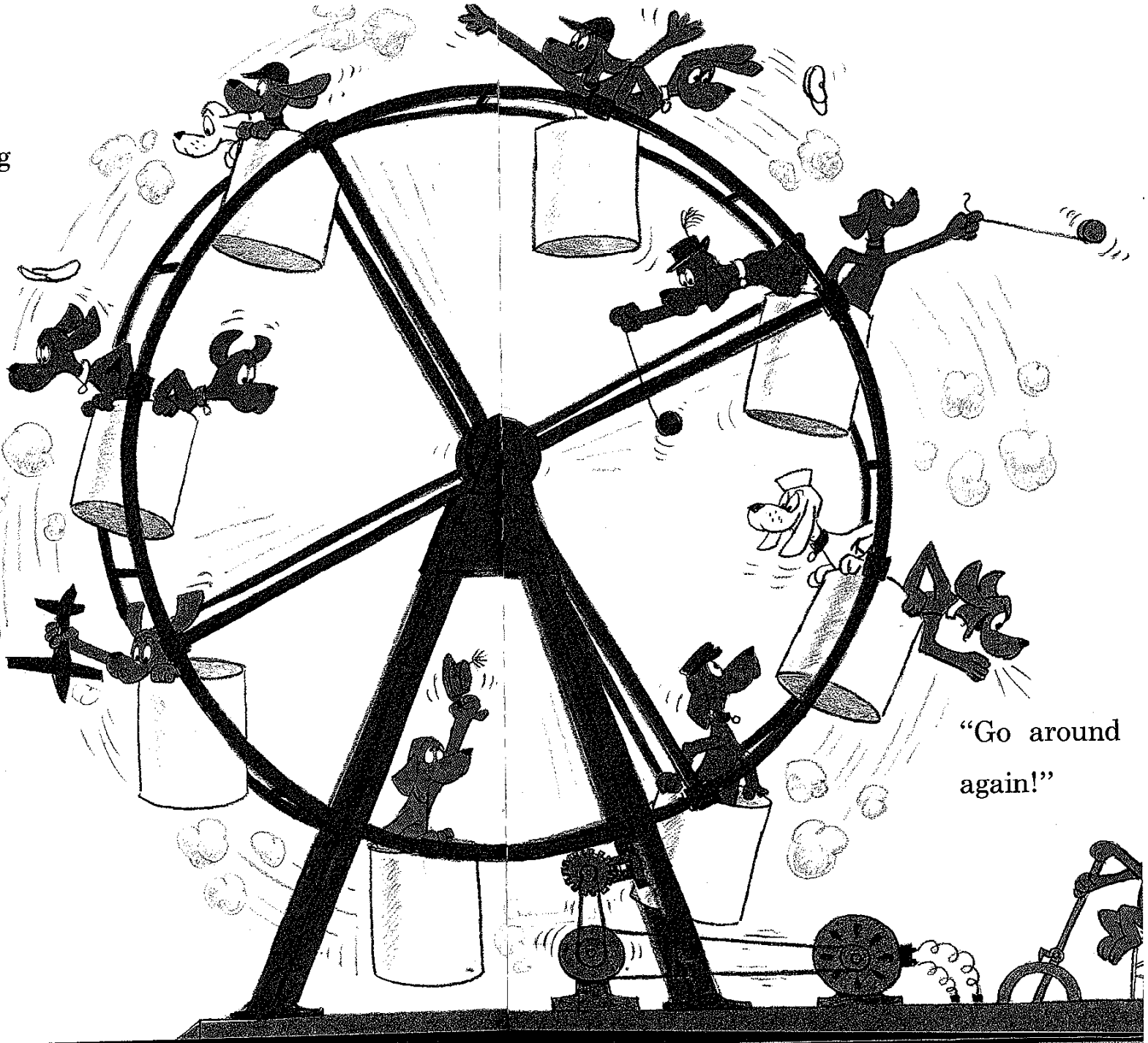


"I do not."

"Good-by!"

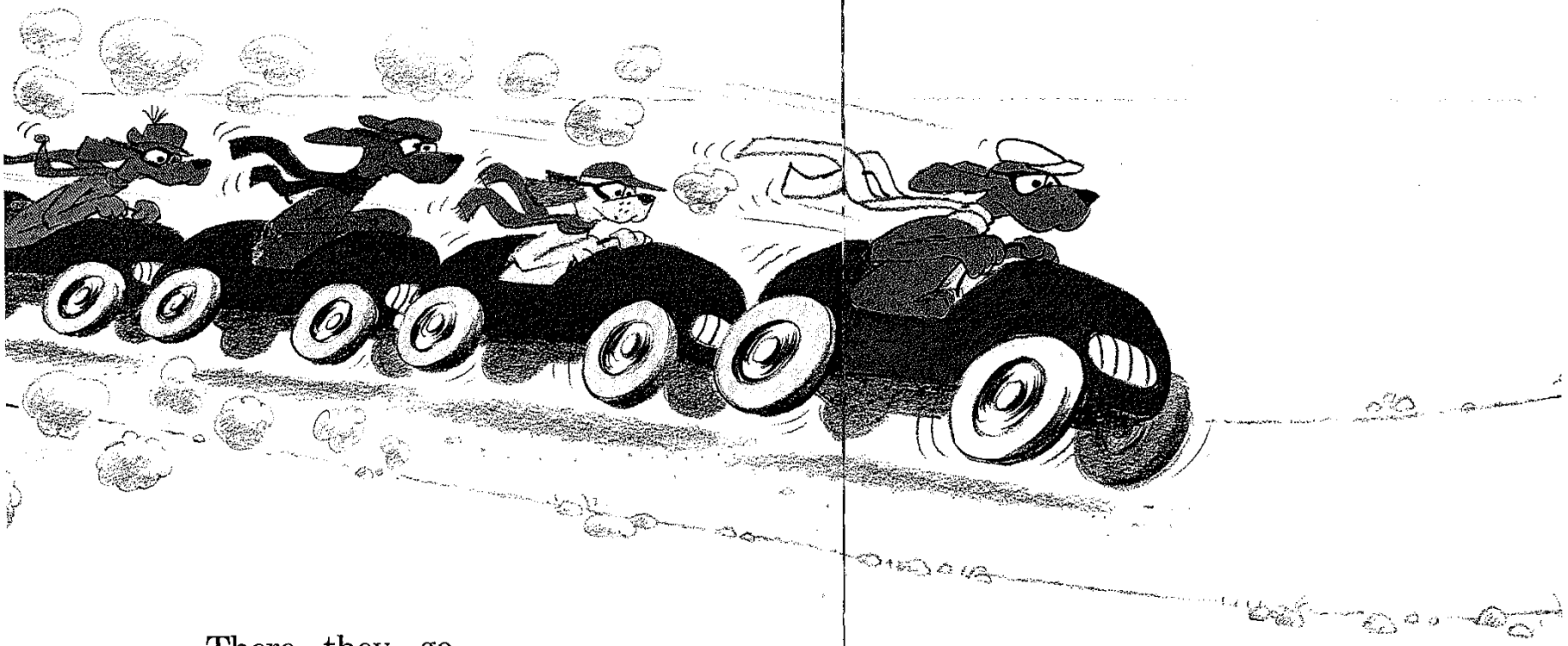
"Good-by!"

The dogs
are all going
around,
and around,
and around.



"Go around
again!"

Why are they going fast
in those cars?
What are they going to do?
Where are those dogs going?



There they go.
Look at those dogs go!

NPs

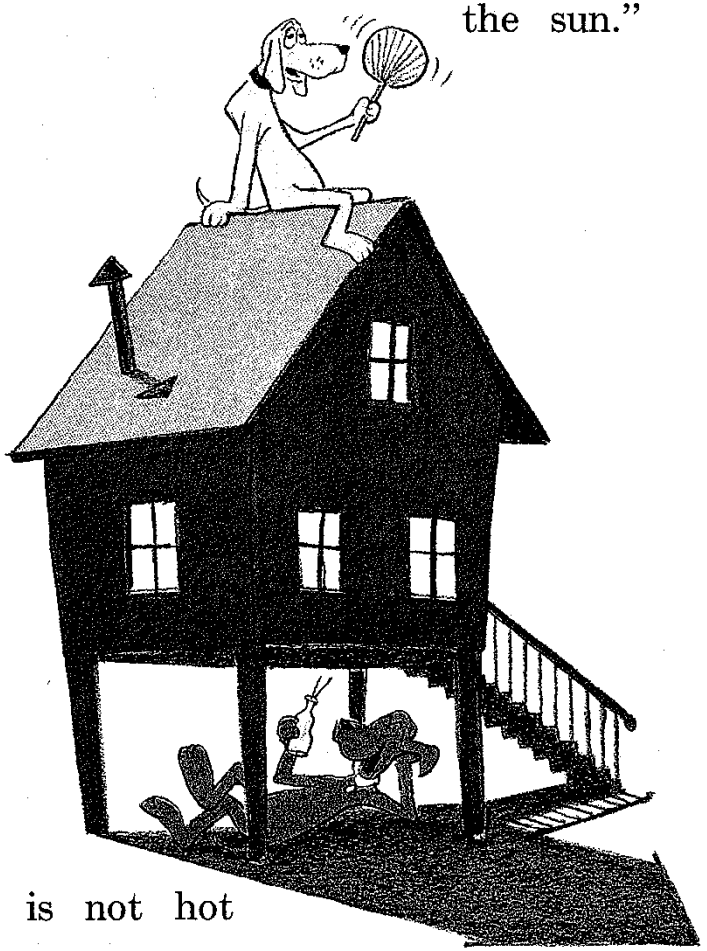
- NP -> Pronoun
  - I came, you saw it, they conquered
- NP -> Proper-Noun
  - New Jersey is west of New York City
  - Lee Bollinger is the president of Columbia
- NP -> Det Noun
  - The president
- NP -> Nominal
- Nominal -> Noun Noun
  - A morning flight to Denver

PPs

- PP -> Preposition NP
  - Over the house
  - Under the house
  - To the tree
  - At play
  - At a party on a boat at night

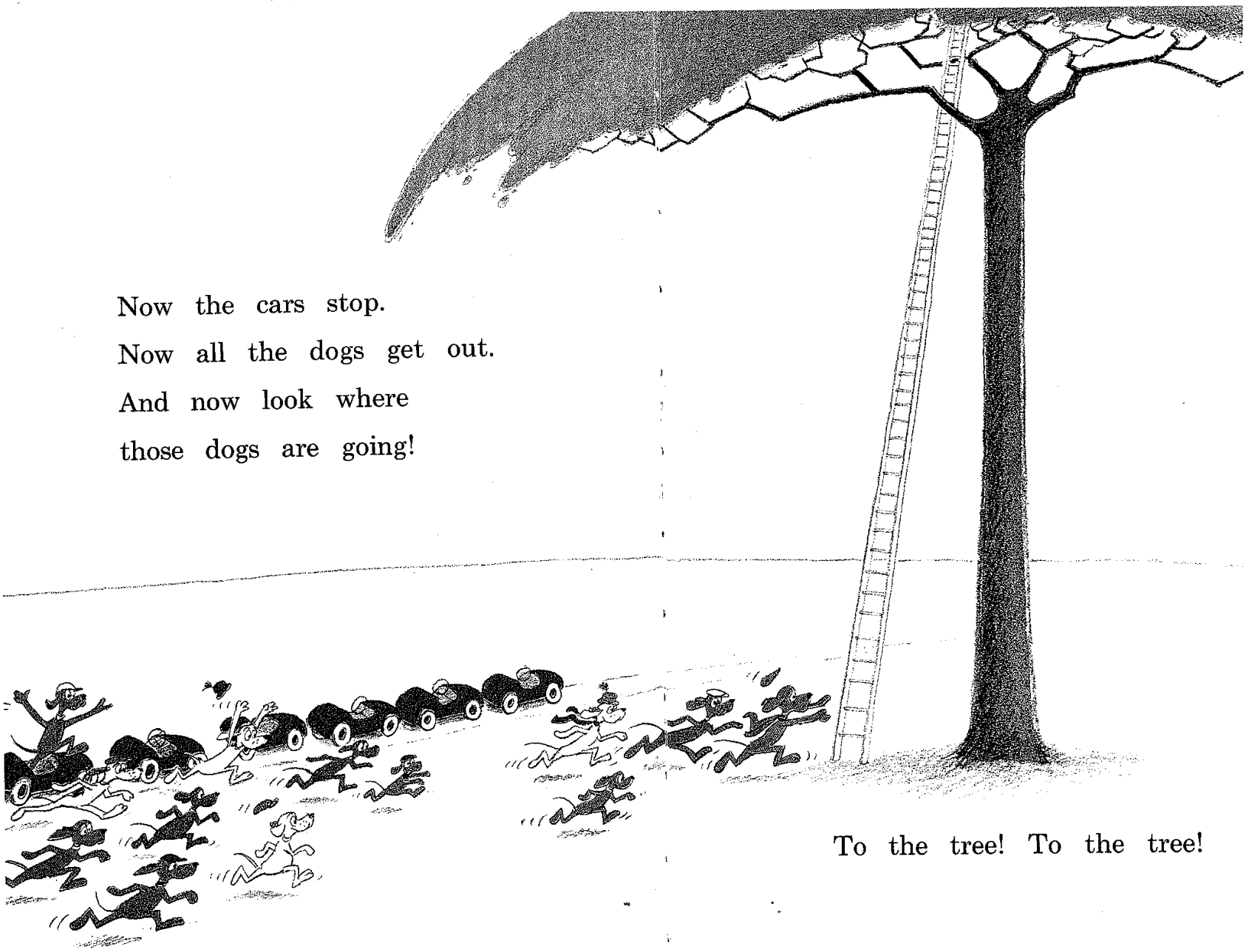
The sun is up.
The sun is yellow.
The yellow sun
is over the house.

"It is hot
out here in
the sun."



"It is not hot
here under the house."

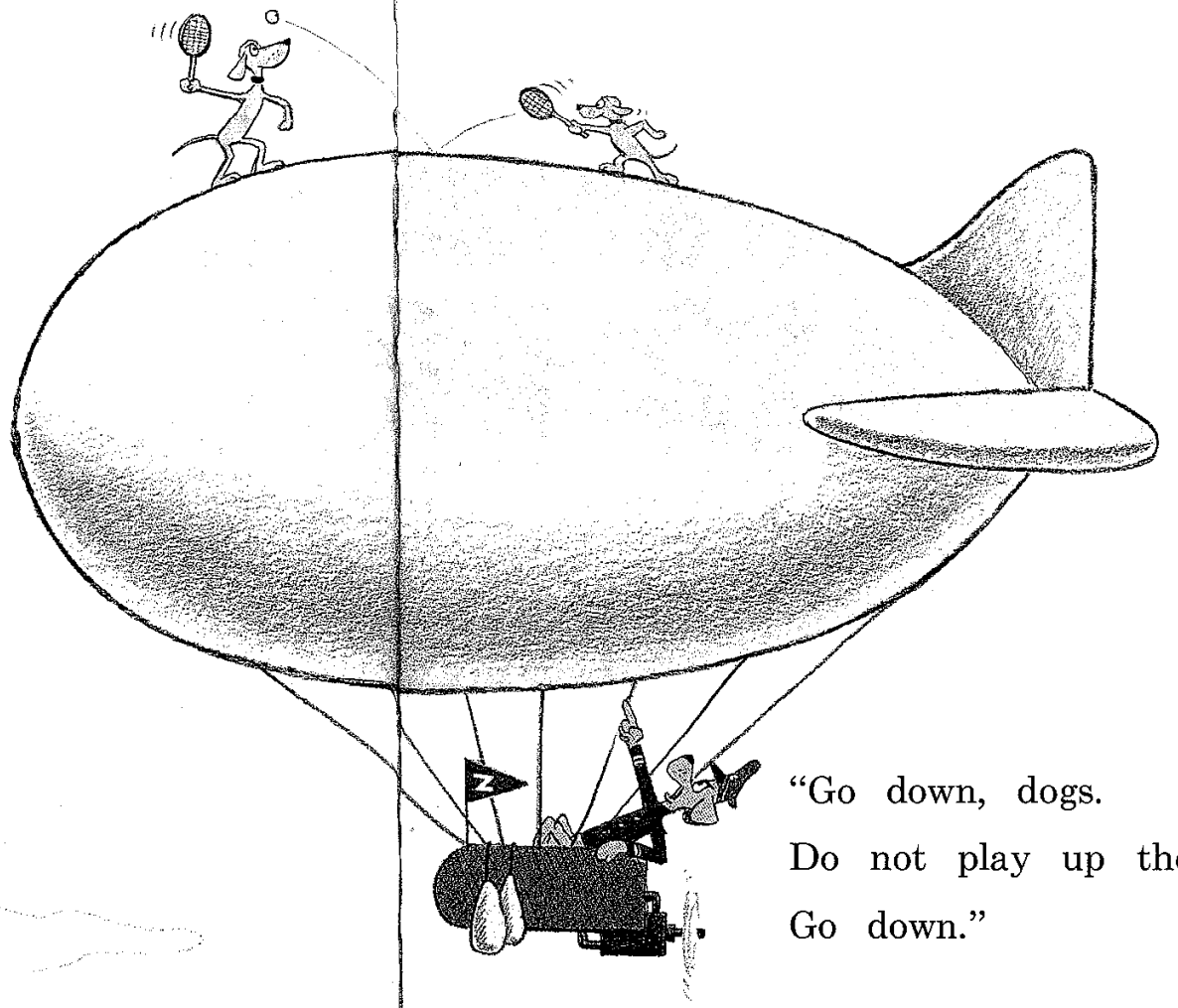
Now the cars stop.
Now all the dogs get out.
And now look where
those dogs are going!



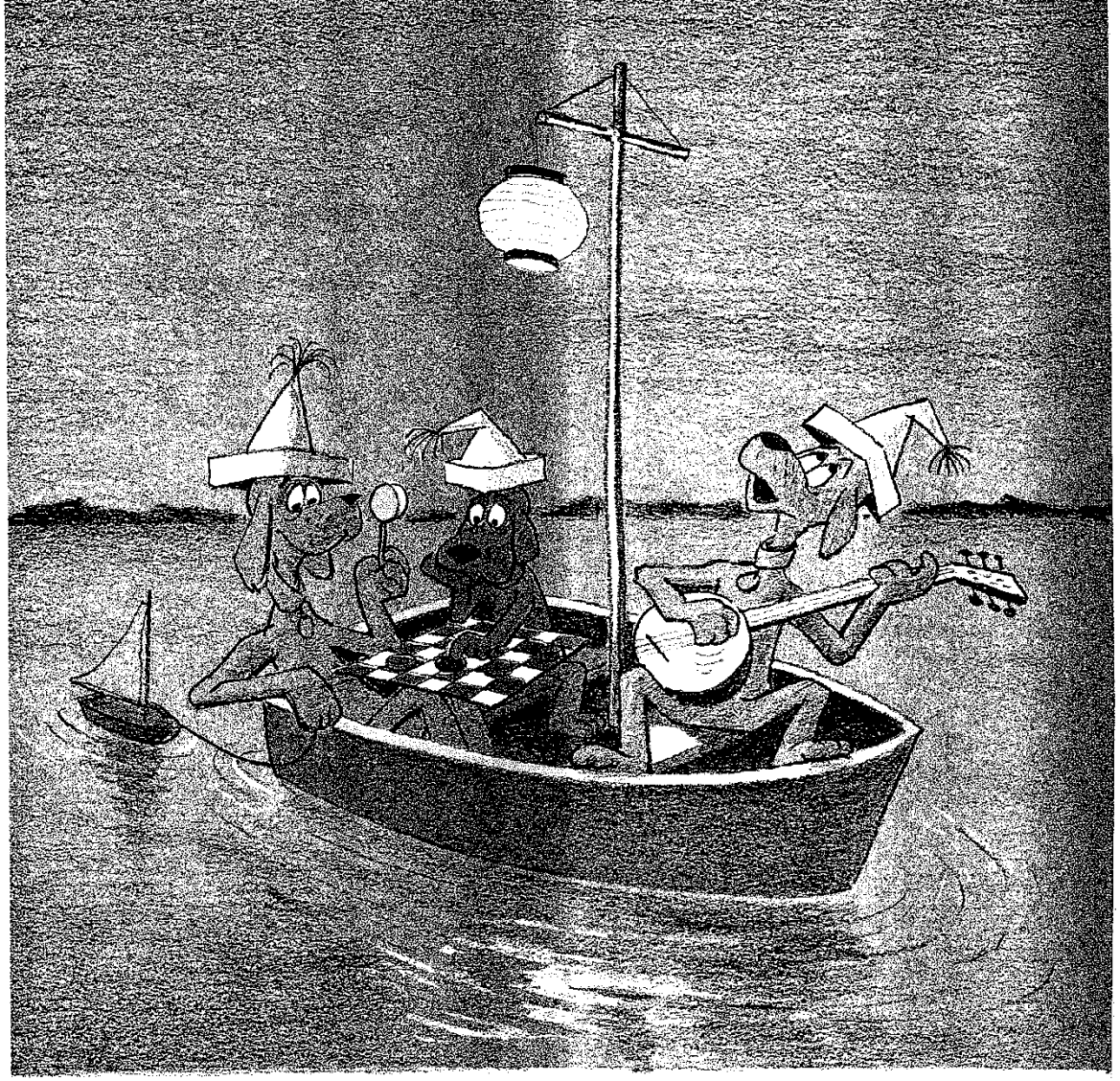
To the tree! To the tree!

Two dogs at play.

At play up on top.



"Go down, dogs.
Do not play up there.
Go down."



Three dogs
at a party
on a boat
at night.

Recursion

- We'll have to deal with rules such as the following where the non-terminal on the left also appears somewhere on the right (directly)

  NP -> NP PP [[The flight] [to Boston]]

  VP -> VP PP [[departed Miami] [at noon]]

  (indirectly)

  NP -> NP Srel

  Srel -> NP VP [[the dog] [[the cat] likes]]

Recursion

- Of course, this is what makes syntax interesting

The dog bites

The dog the mouse bit bites

The dog the mouse the cat ate bit bites

Recursion

[[Flights] [from Denver]]

[[[Flights] [from Denver]] [to Miami]]

[[[[Flights] [from Denver]] [to Miami]] [in February]]

[[[[[Flights] [from Denver]] [to Miami]] [in February]] [on a Friday]]

Etc.

NP -> NP PP

Implications of Recursion and Context-Freeness

- VP -> V NP
- (I) hate

  flights from Denver

  flights from Denver to Miami

  flights from Denver to Miami in February

  flights from Denver to Miami in February on a Friday

  flights from Denver to Miami in February on a Friday under $300

  flights from Denver to Miami in February on a Friday under $300 with lunch
- This is why context-free grammars are appealing! If you have a rule like

  VP -> V NP

  – It only cares that the thing after the verb is an NP

    It doesn't have to know about the internal affairs of that NP

Grammar Equivalence

- Can have different grammars that generate same set of strings (weak equivalence)
  - Grammar 1: NP $\rightarrow$ DetP N and DetP $\rightarrow$ a | the
  - Grammar 2: NP $\rightarrow$ a N | NP $\rightarrow$ the N
- Can have different grammars that have same set of derivation trees (strong equivalence)
  - With CFGs, possible only with useless rules
  - Grammar 2: NP $\rightarrow$ a N | NP $\rightarrow$ the N
  - Grammar 3: NP $\rightarrow$ a N | NP $\rightarrow$ the N, DetP $\rightarrow$ many
- Strong equivalence implies weak equivalence

Normal Forms &c

- There are weakly equivalent normal forms (Chomsky Normal Form, Greibach Normal Form)
- There are ways to eliminate useless productions and so on

Chomsky Normal Form

A CFG is in Chomsky Normal Form (CNF) if all productions are of one of two forms:

- $A \rightarrow BC$ with A, B, C nonterminals
- $A \rightarrow a$, with A a nonterminal and a a terminal

Every CFG has a weakly equivalent CFG in CNF

“Generative Grammar”

- Formal languages: formal device to generate a set of strings (such as a CFG)
- Linguistics (Chomskyan linguistics in particular): approach in which a linguistic theory enumerates all possible strings/structures in a language (=competence)
- Chomskyan theories do not really use formal devices – they use CFG + informally defined transformations

Nobody Uses Simple CFGs (Except Intro NLP Courses)

- All major syntactic theories (Chomsky, LFG, HPSG, TAG-based theories) represent both phrase structure and dependency, in one way or another
- All successful parsers currently use statistics about phrase structure and about dependency
- Derive dependency through “head percolation”: for each rule, say which daughter is head

Massive Ambiguity of Syntax

- For a standard sentence, and a grammar with wide coverage, there are 1000s of derivations!
- Example:
  - The large portrait painter told the delegation that he sent money orders in a letter on Wednesday

Penn Treebank (PTB)

- Syntactically annotated corpus of newspaper texts (phrase structure)
- The newspaper texts are naturally occurring data, but the PTB is not!
- PTB annotation represents a particular linguistic theory (but a fairly "vanilla" one)
- Particularities
  - Very indirect representation of grammatical relations (need for head percolation tables)
  - Completely flat structure in NP (*brown bag lunch, pink-and-yellow child seat*)
  - Has flat Ss, flat VPs

Example from PTB

```
( (S (NP-SBJ It)
    (VP 's
      (NP-PRD (NP (NP the latest investment craze)
                  (VP sweeping
                      (NP Wall Street)))
              :
              (NP (NP a rash)
                 (PP of
                          (NP (NP new closed-end country funds)
                             ,
                             (NP (NP those
                                      (ADJP publicly traded)
                                      portfolios)
                                   (SBAR (WHNP-37 that)
                                      (S (NP-SBJ *T*-37)
                                               (VP invest
                                                  (PP-CLR in
                                                           (NP (NP stocks)
                                                               (PP of
                                                                 (NP a single foreign country)))))))))))))
```

Types of syntactic constructions

- Is this the same construction?
  - An elf **decided** to clean the kitchen
  - An elf **seemed** to clean the kitchen

  An elf cleaned the kitchen
- Is this the same construction?
  - An elf **decided** to be in the kitchen
  - An elf **seemed** to be in the kitchen

  An elf was in the kitchen

Types of syntactic constructions (ctd)

- Is this the same construction?

  There is an elf in the kitchen

  – *There **decided** to be an elf in the kitchen

  – There **seemed** to be an elf in the kitchen

- Is this the same construction?

  It is raining/it rains

  – ??It **decided** to rain/be raining

  – It **seemed** to rain/be raining

Types of syntactic constructions (ctd)

- Is this the same construction?
  - An elf **decided** that he would clean the kitchen
  - * An elf **seemed** that he would clean the kitchen

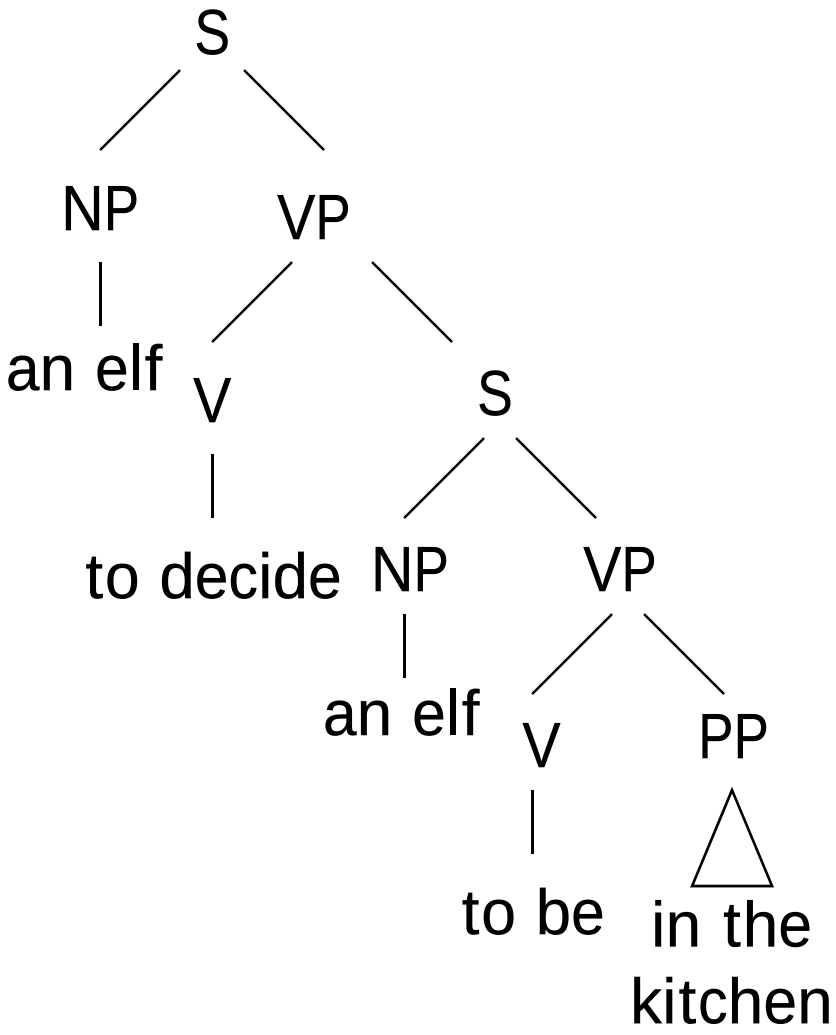
  An elf cleaned the kitchen

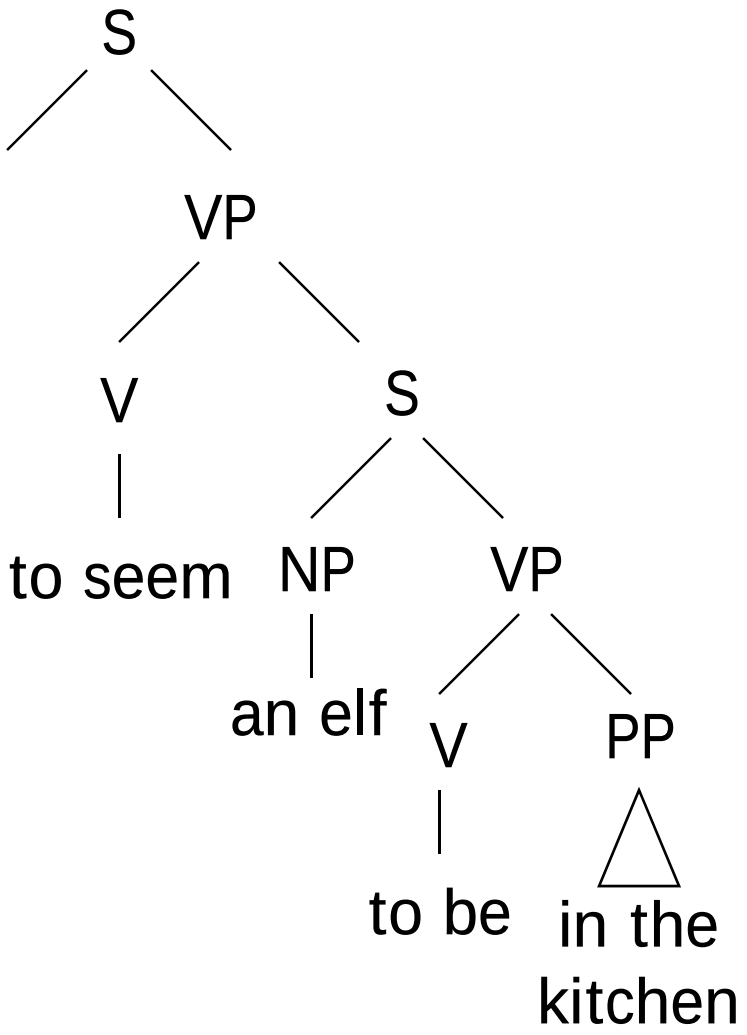
Types of syntactic constructions (ctd)

Conclusion:

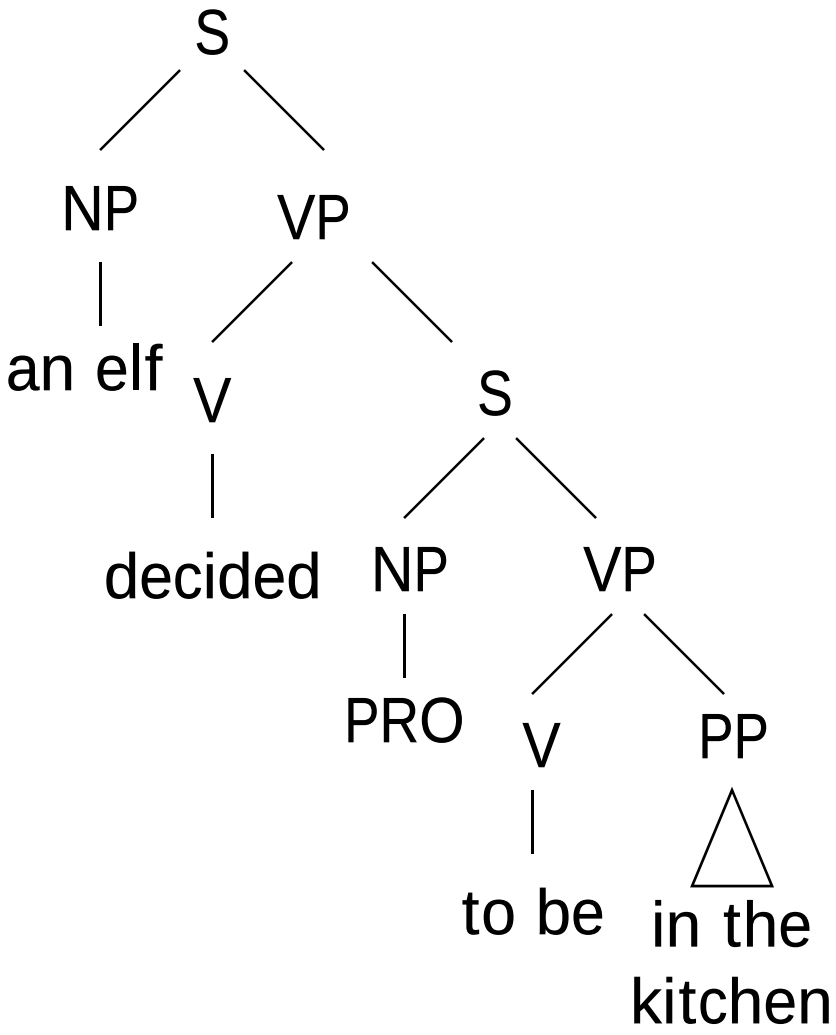
- *to seem:* whatever is embedded surface subject can appear in upper clause
- *to decide:* only full nouns that are referential can appear in upper clause
- Two types of verbs

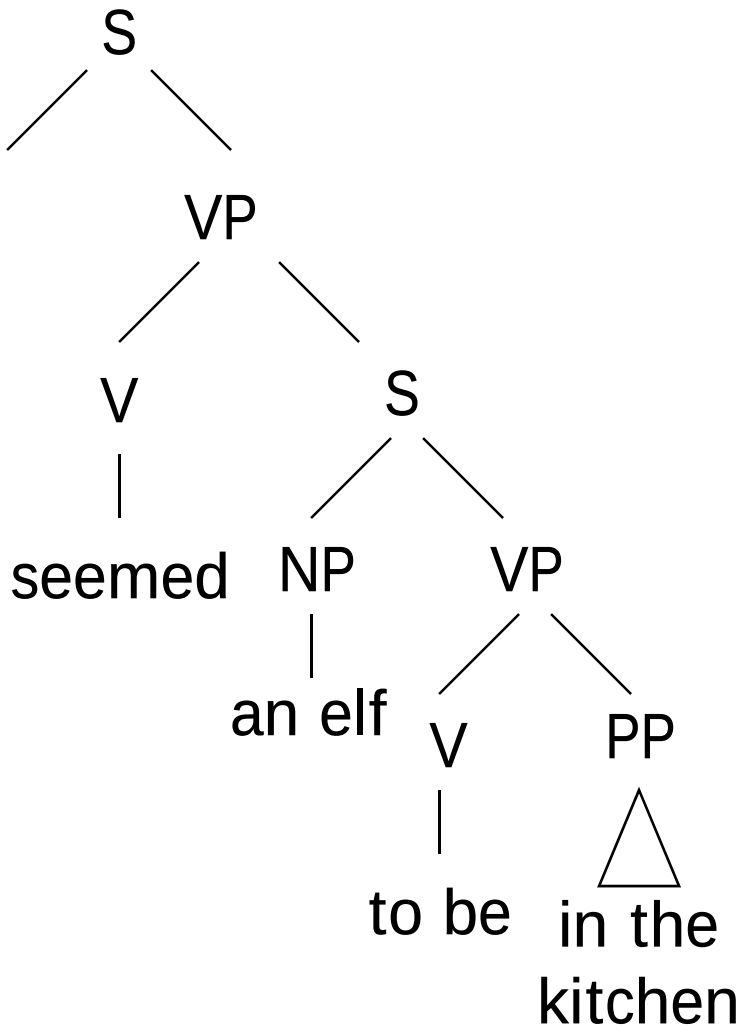
Types of syntactic constructions: Analysis



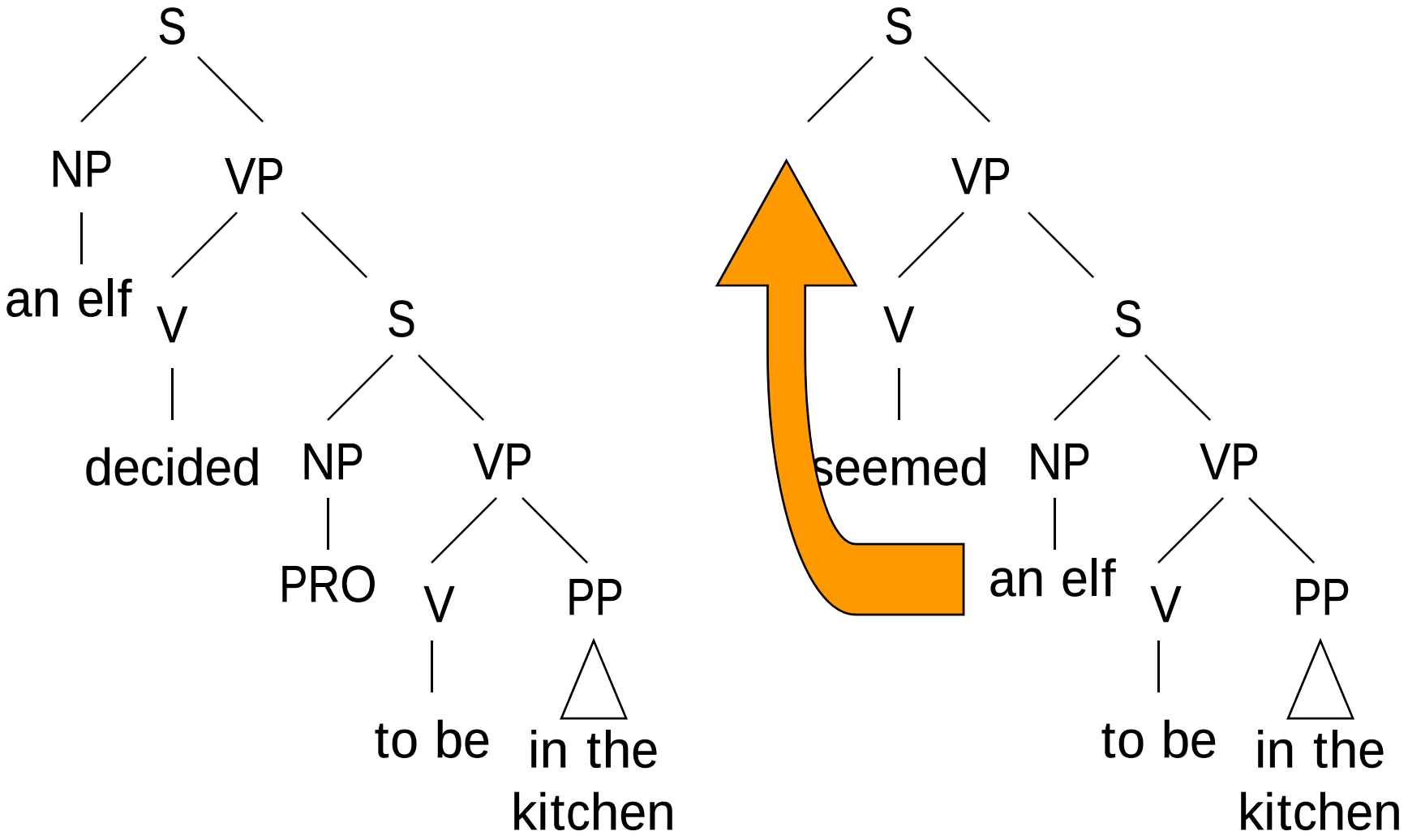


Types of syntactic constructions: Analysis

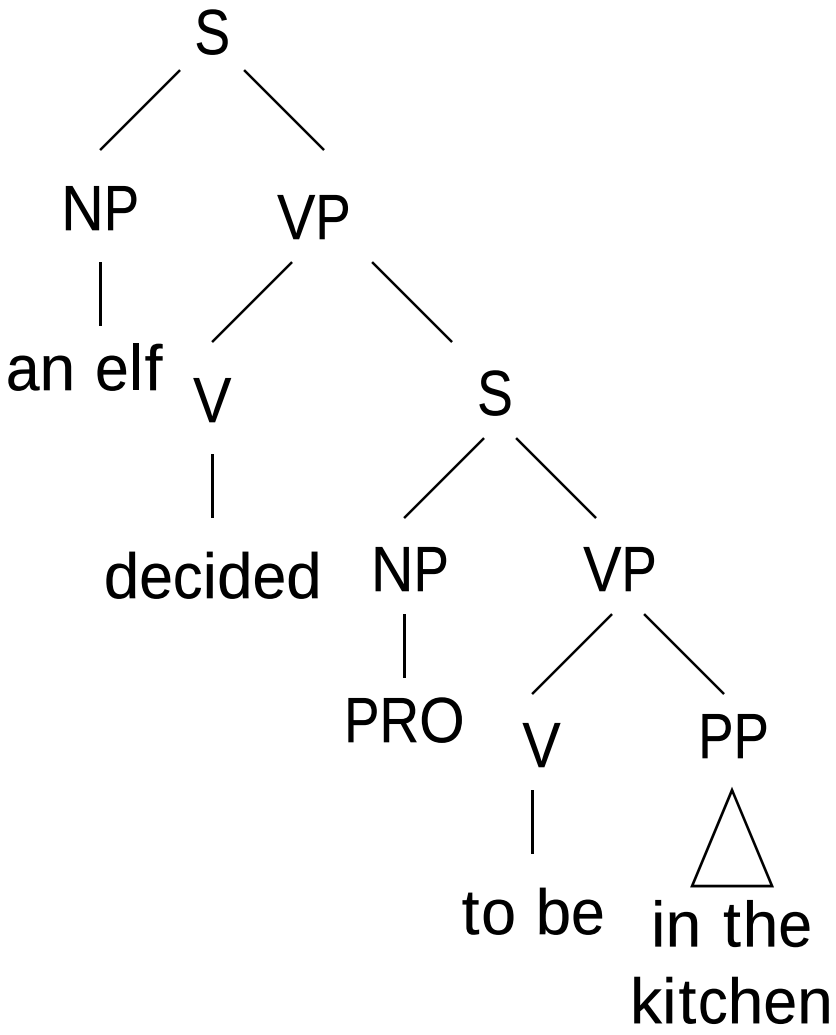


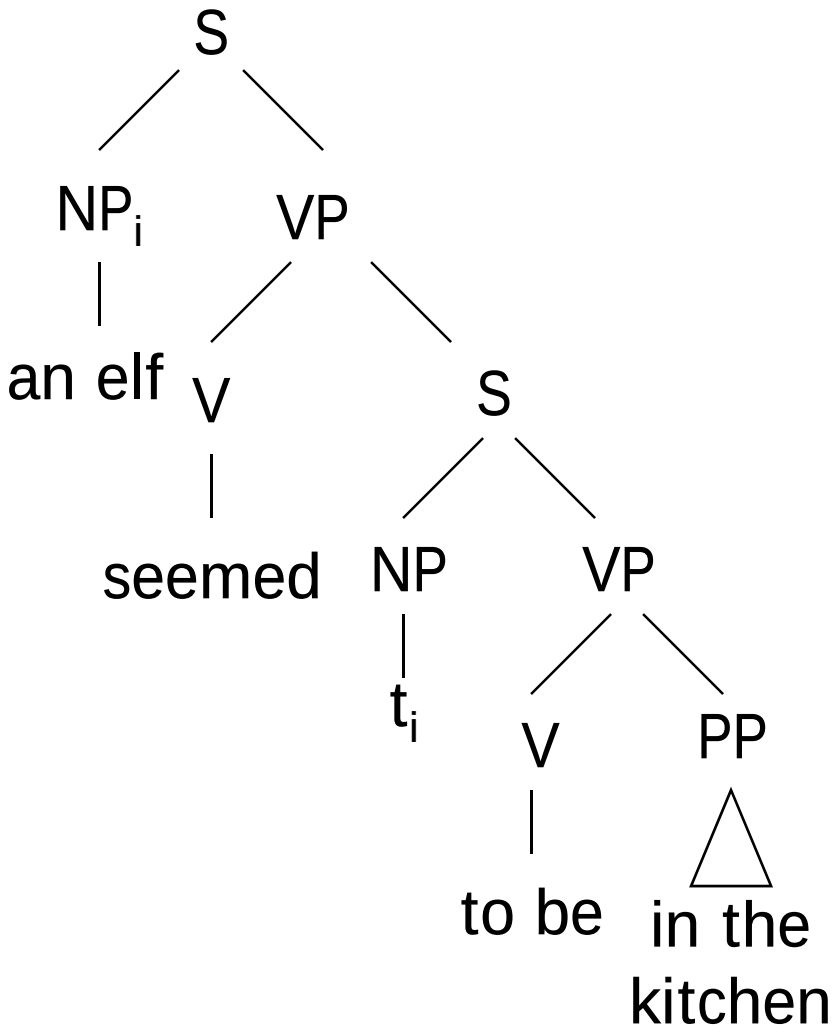


Types of syntactic constructions: Analysis



Types of syntactic constructions: Analysis





Types of syntactic constructions: Analysis

to seem: lower surface subject **raises** to upper clause; **raising verb**

seems (there to be an elf in the kitchen)

there seems (*t* to be an elf in the kitchen)

it seems (there is an elf in the kitchen)

Types of syntactic constructions: Analysis (ctd)

- *to decide:* subject is in upper clause and co-refers with an empty subject in lower clause; **control verb**

an elf decided (an elf to clean the kitchen)

an elf decided (PRO to clean the kitchen)

an elf decided (he cleans/should clean the kitchen)

*it decided (an elf cleans/should clean the kitchen)

Lessons Learned from the Raising/Control Issue

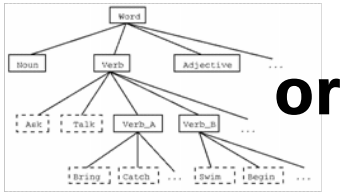
- Use distribution of data to group phenomena into classes
- Use different underlying structure as basis for explanations
- Allow things to “move” around from underlying structure -> **transformational grammar**
- Check whether explanation you give makes predictions

Examples from PTB

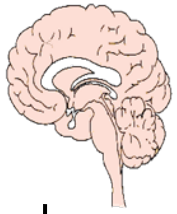
```
(S (NP-SBJ-1 The ropes)
      (VP seem
         (S (NP-SBJ *-1)
           (VP to
              (VP make
                 (NP much sound))))))

(S (NP-SBJ-1 The ancient church vicar)
    (VP refuses
       (S (NP-SBJ *-1)
         (VP to
            (VP talk
               (PP-CLR about
                           (NP it)))))
```

The Big Picture

Empirical Matter



or

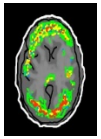


Formalisms

- Data structures
- Formalisms
- Algorithms
- Distributional Models

descriptive theory is about

Maud expects there to be a riot
*Teri promised there to be a riot
Maud expects the shit to hit the fan
*Teri promised the shit to hit the



uses

predicts

explanatory theory is about

Linguistic Theory

Content: Relate morphology to semantics
- Surface representation (eg, ps)
- Deep representation (eg, dep)
- Correspondence

Introduction to Syntax and Context-Free Grammars

http://www1.cs.columbia.edu/~rambow/teaching/lecture-2009-09-22.ppt

Owen Rambow

rambow@ccls.columbia.edu

Slides with contributions from Kathy McKeown, Dan Jurafsky and James Martin