

Syntax

M.Sc. Preparatory Course

October 2, 2009



What is Syntax?

- Study of structure of language
- Roughly, goal is to relate surface form (what we perceive when someone says something) to semantics (what that utterance means)

What is Syntax Not?

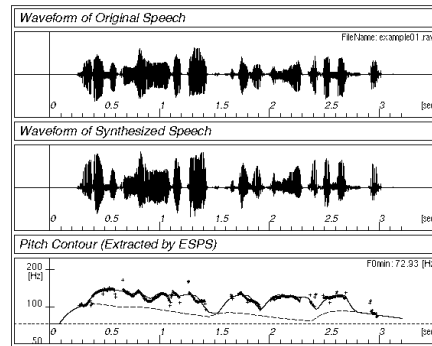
- Phonology: study of sound systems and how sounds combine
- Morphology: study of how words are formed from smaller parts (morphemes)
- Semantics: study of meaning of language

What is Syntax? (2)

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

Simplified Big Picture

Phonology



⇔ /waddyasai/

Morphology

/waddyasai/ ⇔ what did you say

Syntax

what did you say ⇔
$$\begin{array}{c} \text{say} \\ \text{subj} \quad \text{obj} \\ \text{you} \quad \text{what} \end{array}$$

Semantics

$$\begin{array}{c} \text{say} \\ \text{subj} \quad \text{obj} \\ \text{you} \quad \text{what} \end{array} \quad \Leftrightarrow \quad P[\lambda x. \text{say}(\text{you}, x)]$$

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
- These ideas adopted by almost all contemporary syntactic theories (“principles-and-parameters-type theories”)

Types of Linguistic Theories

- **Descriptive:** provide account of syntax of a language; often appropriate for NLP engineering work
- **Explanatory:** provide principles-and-parameters style account of syntax of (preferably) several languages
- ~~**Prescriptive:**~~ “prescriptive linguistics” is an oxymoron

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)

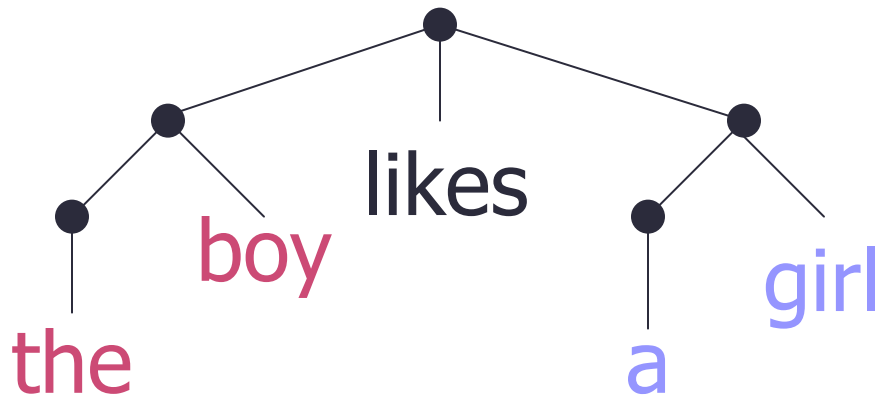
More Structure in Strings

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))

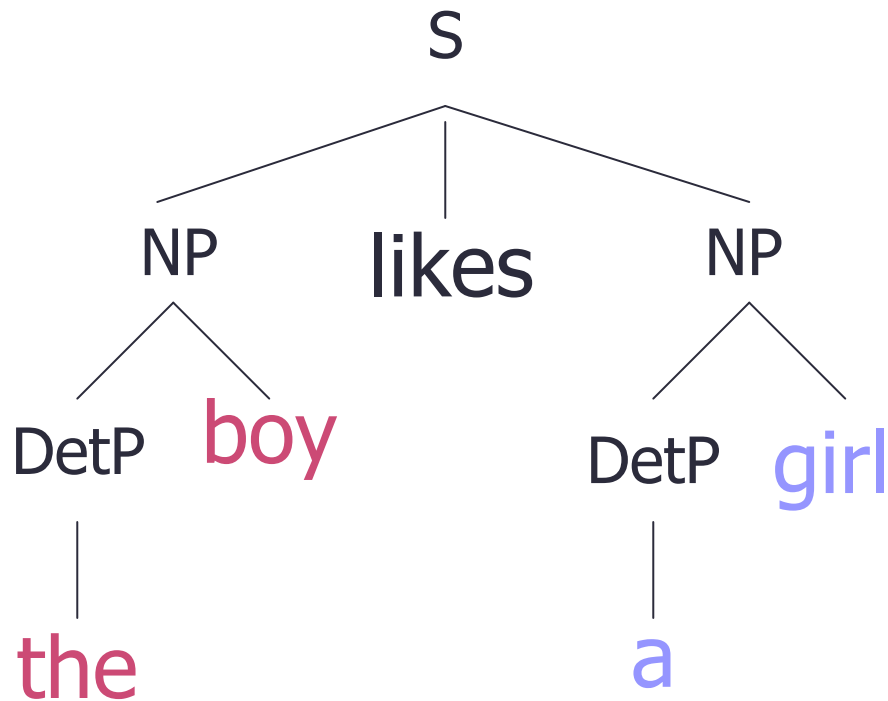


Node Labels?

- (((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

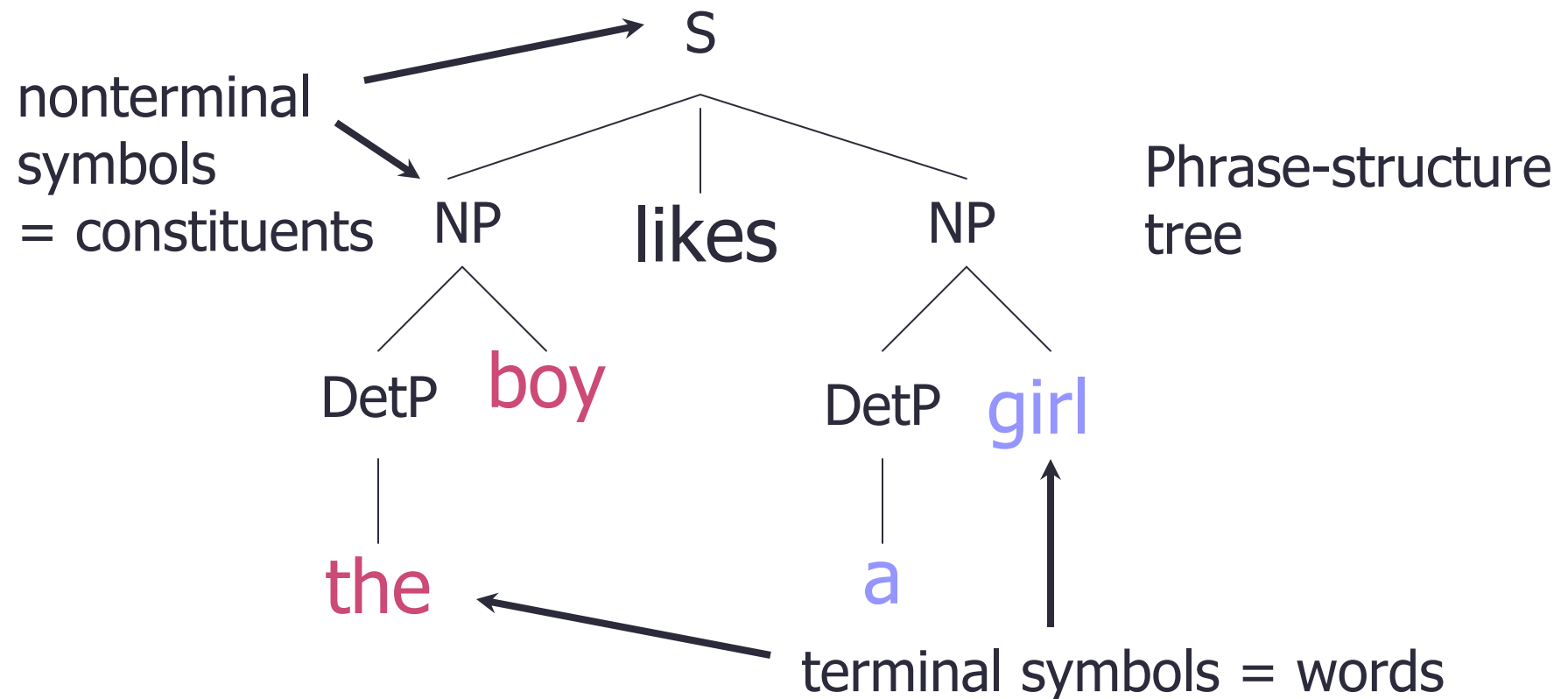
Node Labels

- (((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

o noun or adjective?

➤ a **blue** seat

a **child** seat

➤ a very **blue** seat

*a very **child** seat

➤ this seat is **blue**

*this seat is **child**

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

➤ **blue** is Adj, **child** is Noun

Determining Part-of-Speech (2)

o 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

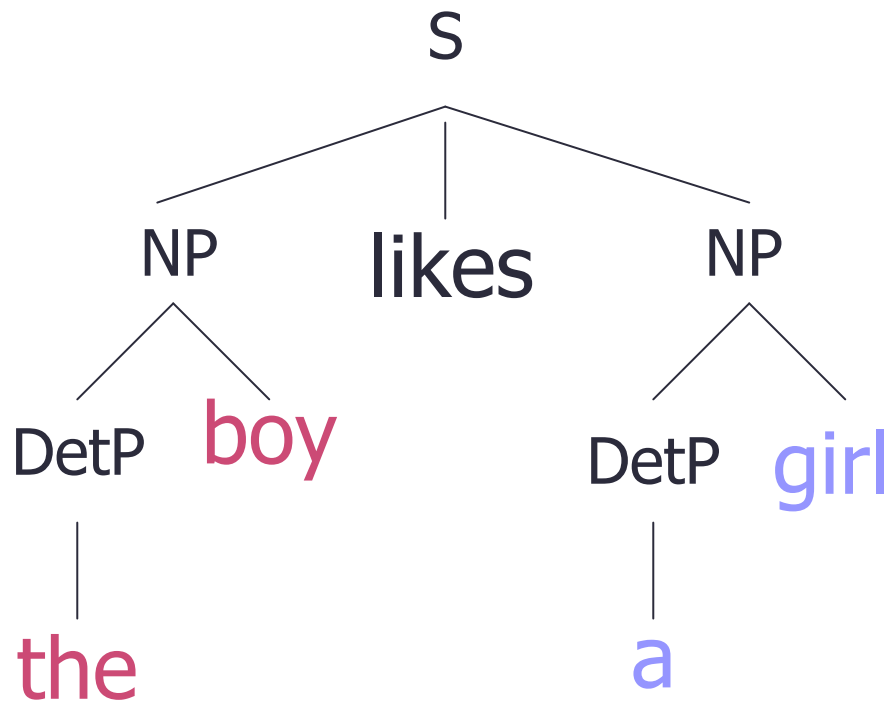
Word Classes (=POS)

- Heads of constituents fall into distributionally defined classes
- Additional support for class definition of word class comes from morphology

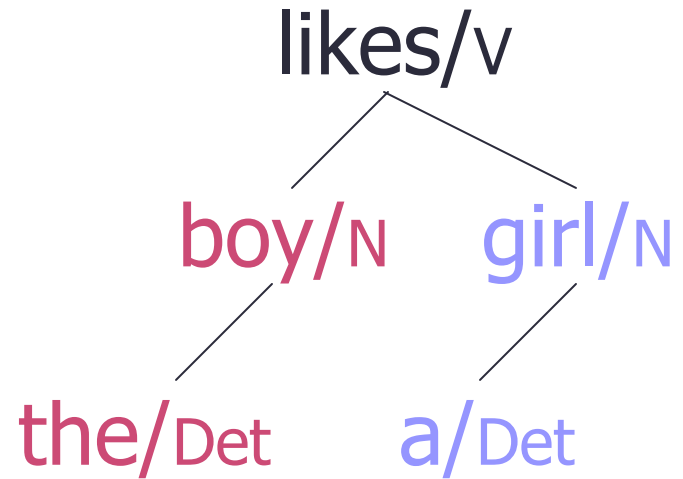
Some Points on POS Tag Sets

- Possible basic set: N, V, Adj, Adv, P, Det, Aux, Comp, Conj
- 2 supertypes: open- and closed-class
 - Open: N, V, Adj, Adv
 - Closed: P, Det, Aux, Comp, Conj
- Many subtypes:
 - eats/V \Rightarrow eat/VB, eat/VBP, eats/VBZ, ate/VBD, eaten/VBN, eating/VBG,
 - Reflect morphological form & syntactic function

Phrase Structure and Dependency Structure

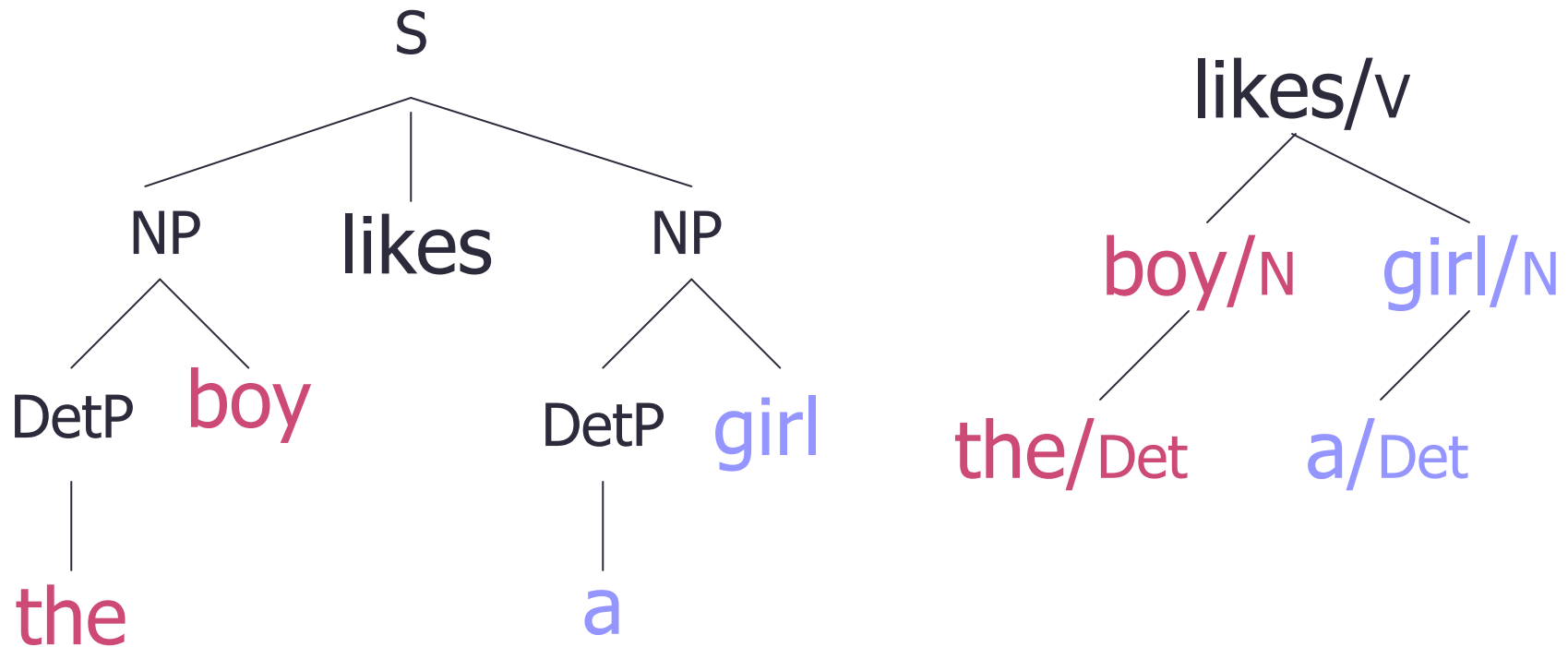


Only leaf nodes labeled with words!



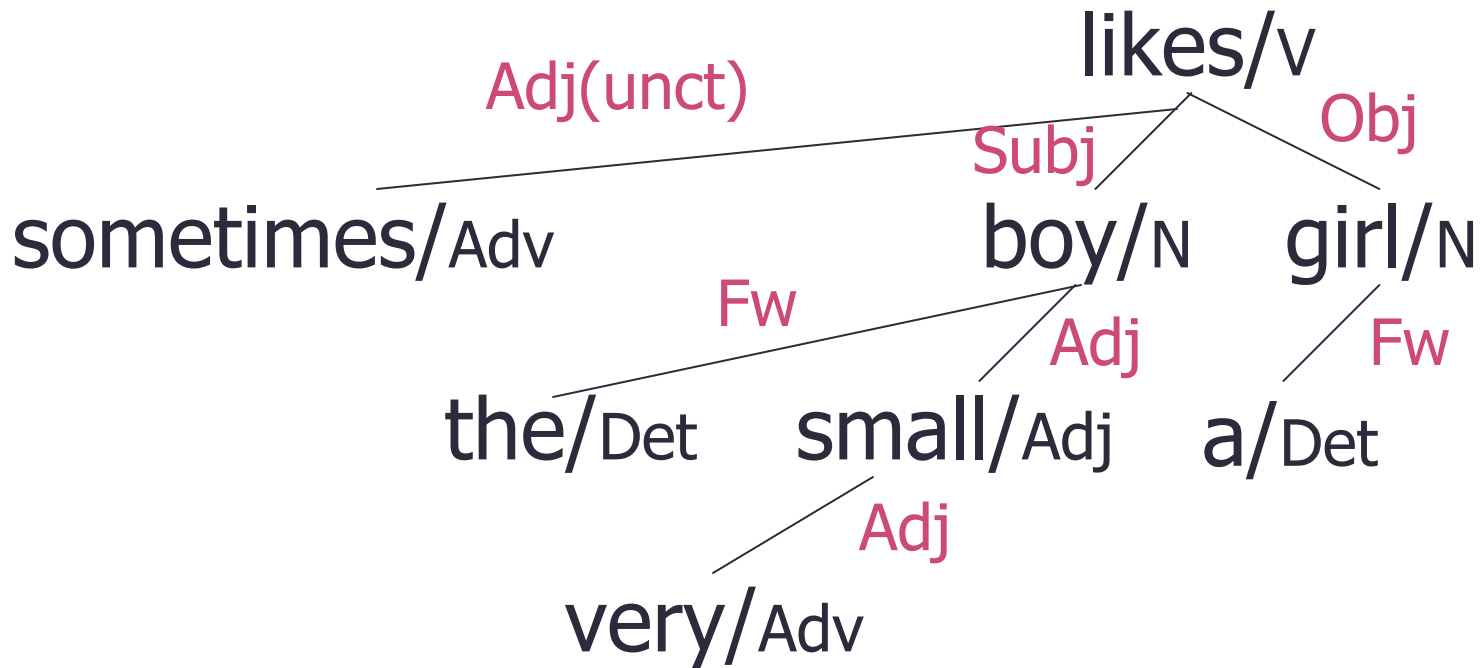
All nodes are labeled with words!

Phrase Structure and Dependency Structure (ctd)



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

Types of Dependency



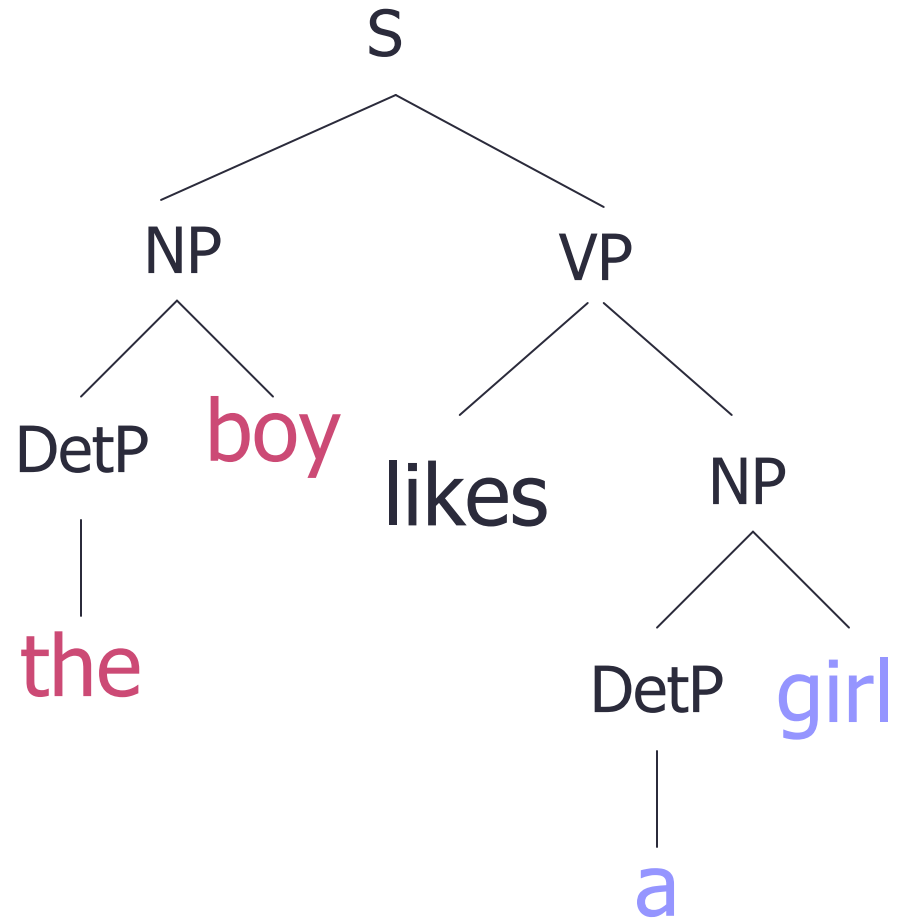
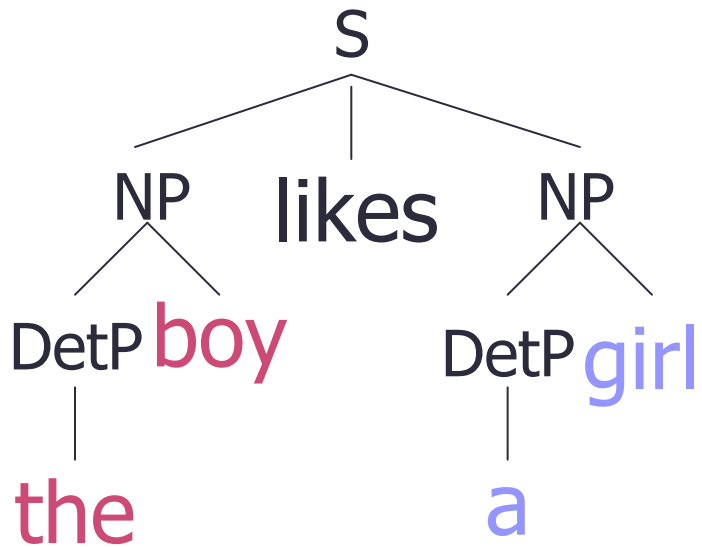
Grammatical Relations

- 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):
 - $S \rightarrow NP VP$
 - $VP \rightarrow V NP$
 - $NP \rightarrow DetP N \mid AdjP NP$
 - $AdjP \rightarrow Adj \mid Adv AdjP$
 - $N \rightarrow \text{boy} \mid \text{girl}$
 - $V \rightarrow \text{sees} \mid \text{likes}$
 - $Adj \rightarrow \text{big} \mid \text{small}$
 - $Adv \rightarrow \text{very}$
 - $DetP \rightarrow \text{a} \mid \text{the}$
- 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

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$NP \rightarrow \mathbf{DetP N} \mid \text{AdjP NP}$

$\text{AdjP} \rightarrow \text{Adj} \mid \text{Adv AdjP}$

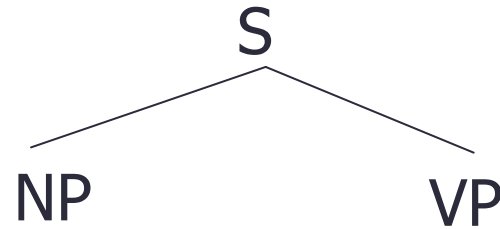
$N \rightarrow \text{boy} \mid \text{girl}$

$V \rightarrow \text{sees} \mid \text{likes}$

$\text{Adj} \rightarrow \text{big} \mid \text{small}$

$\text{Adv} \rightarrow \text{very}$

$\text{DetP} \rightarrow \text{a} \mid \text{the}$



Derivations in a CFG

DetP N VP

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$NP \rightarrow DetP N \mid AdjP NP$

$AdjP \rightarrow Adj \mid Adv AdjP$

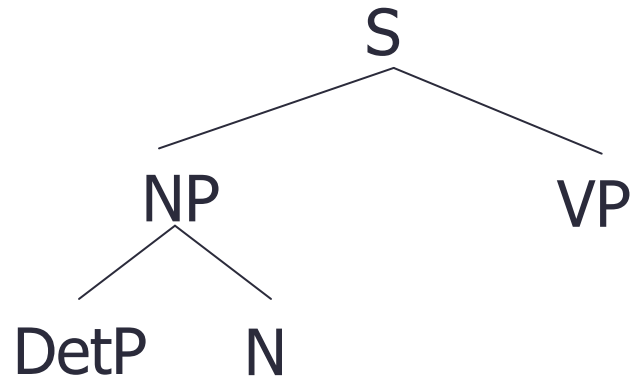
$N \rightarrow \mathbf{boy} \mid \text{girl}$

$V \rightarrow \text{sees} \mid \text{likes}$

$Adj \rightarrow \text{big} \mid \text{small}$

$Adv \rightarrow \text{very}$

$DetP \rightarrow a \mid \mathbf{the}$



Derivations in a CFG

the boy VP

$S \rightarrow NP VP$

$VP \rightarrow \mathbf{V NP}$

$NP \rightarrow DetP N \mid AdjP NP$

$AdjP \rightarrow Adj \mid Adv AdjP$

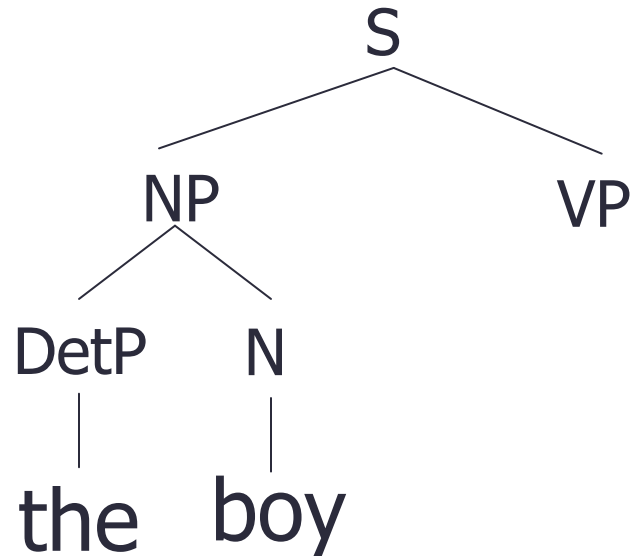
$N \rightarrow boy \mid girl$

$\mathbf{V} \rightarrow sees \mid \mathbf{likes}$

$Adj \rightarrow big \mid small$

$Adv \rightarrow very$

$DetP \rightarrow a \mid the$



Derivations in a CFG

the boy likes NP

$S \rightarrow NP VP$

$VP \rightarrow V NP$

$NP \rightarrow \mathbf{DetP N} \mid \text{AdjP NP}$

$\text{AdjP} \rightarrow \text{Adj} \mid \text{Adv AdjP}$

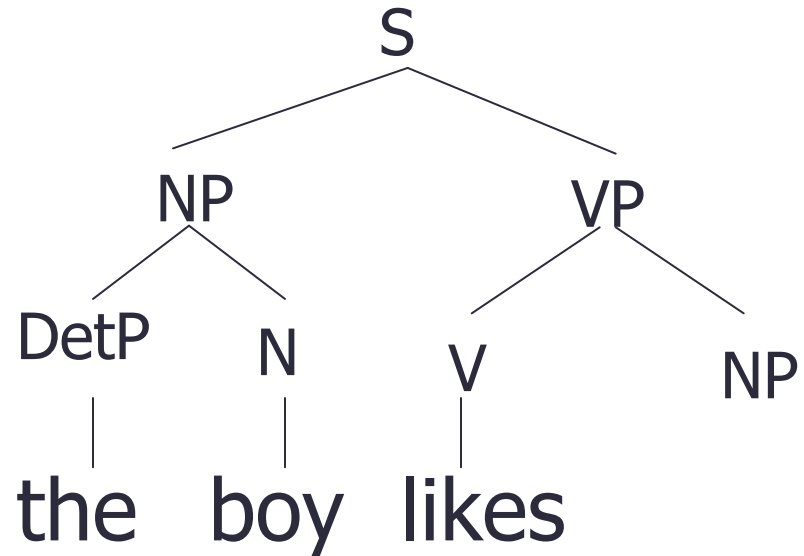
$N \rightarrow \text{boy} \mid \mathbf{girl}$

$V \rightarrow \text{sees} \mid \text{likes}$

$\text{Adj} \rightarrow \text{big} \mid \text{small}$

$\text{Adv} \rightarrow \text{very}$

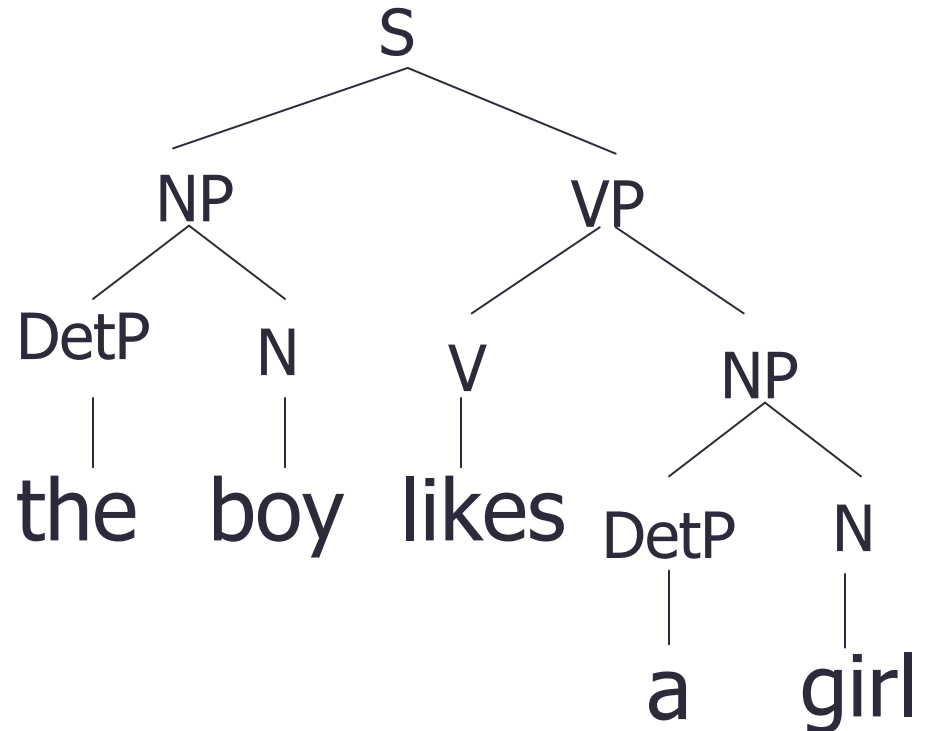
$\mathbf{DetP} \rightarrow \mathbf{a} \mid \mathbf{the}$



Derivations in a CFG

the boy likes a girl

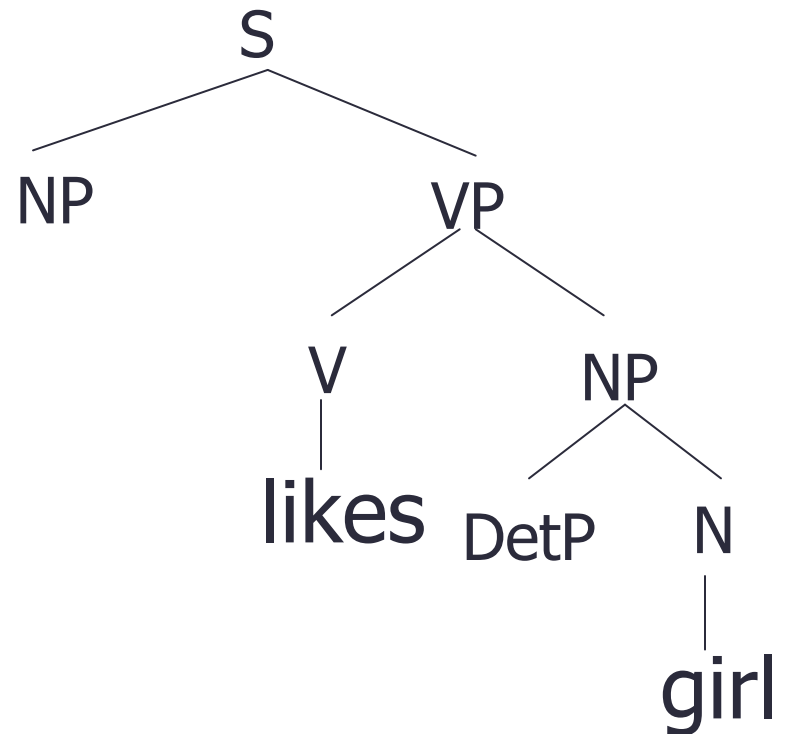
$S \rightarrow NP VP$
 $VP \rightarrow V NP$
 $NP \rightarrow DetP N \mid AdjP NP$
 $AdjP \rightarrow Adj \mid Adv AdjP$
 $N \rightarrow boy \mid girl$
 $V \rightarrow sees \mid likes$
 $Adj \rightarrow big \mid small$
 $Adv \rightarrow very$
 $DetP \rightarrow a \mid the$



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)!

Grammar Equivalence

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