

RECORDING!!!

CFGs

LING566 — Oct. 5, 2026

Overview

- Some admin
- Two insufficient theories of grammar
- Formal definition of CFG
- Constituency, ambiguity, constituency tests
- Central claims of CFG
- CFG Mechanics
- Weaknesses of CFG
- RQs
- Group Activities

Admin

- Be sure to form study groups!
- We recommend groups of 4-5 with diverse academic backgrounds
- **Do not record large group sessions.**
- The purpose of the assignments is to help you learn the material
- Working in small groups is beneficial as you can help each other work through the problems
- Gathering together in large groups is likely to turn into a “passing answers around” session

Two theories that won't work

Insufficient Theory #1

- “A grammar is simply a list of sentences”
- What's wrong with this?
 - Languages are infinite, our brains are finite
 - We can always come up with sentences we've never heard before
 - Our experience is finite too, how are we going to learn the list of sentences?

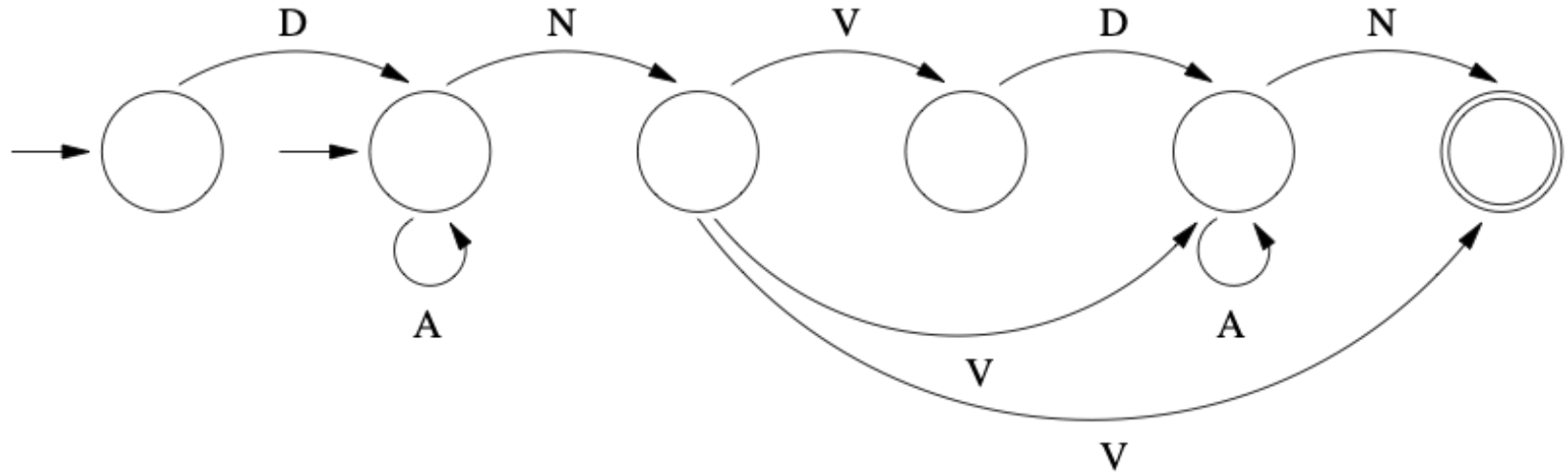
Insufficient Theory #2: FSMs

- the noisy dogs left
- D A N V
- the noisy dogs chased the innocent cats
- D A N V D A N
- $(D) A^* N V ((D) A^* N)$

Kleene* and Kleene+:

- $a^* = \{\emptyset, a, aa, aaa, \dots\}$
- $a^+ = \{a, aa, aaa, \dots\}$

A Finite State Machine



FSMs for Grammar

- Why are FSMs insufficient as a representation of natural language syntax?
 - Doesn't capture the notion of constituents or relationships between them
 - can't model structural ambiguity
 - Can't capture dependencies between elements
 - Can technically model agreement (i.e. sg/plural) but would be very cumbersome to do so

What does a theory of grammar do?

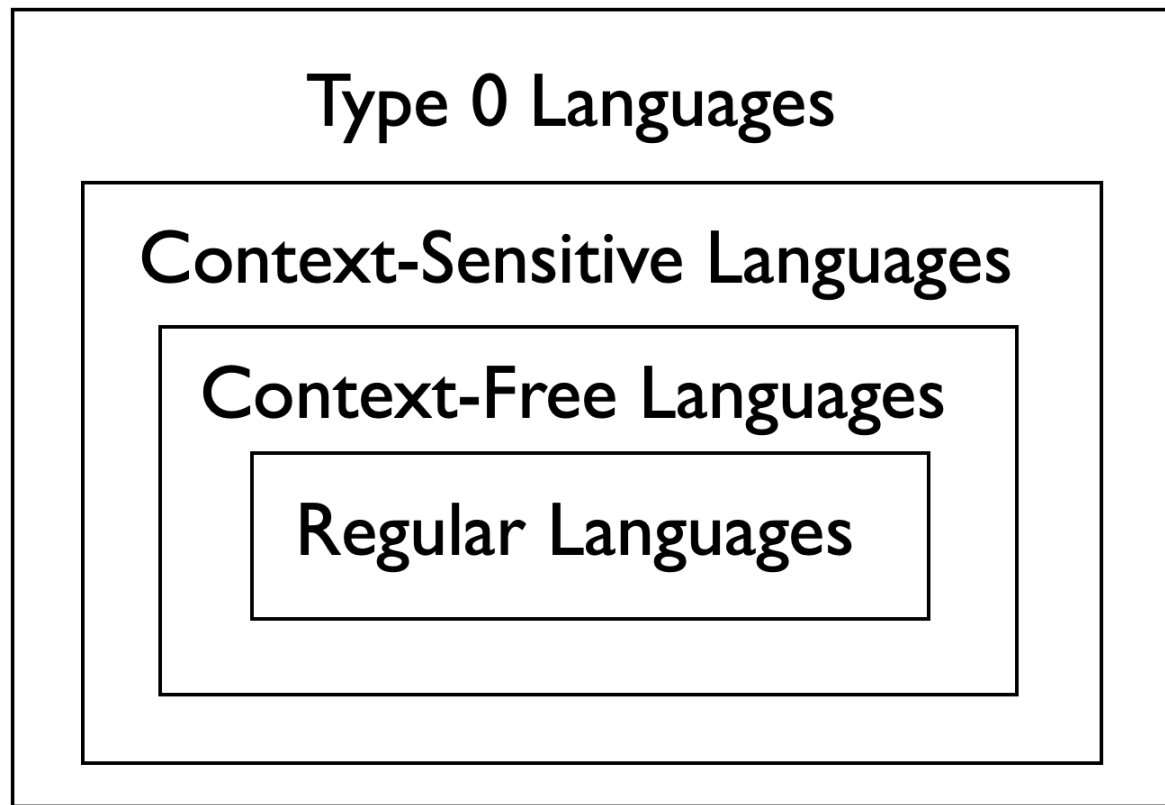
- Monolingual theory
 - Model grammaticality/acceptability
 - Model relationships between sentences (internal structure)
 - e.g. how sentences are similar/different to one another, esp wrt to meaning
- Multilingual theory
 - Model relationships between languages
 - Capture generalizations about possible languages

Summary

- Grammars as lists of sentences
 - Doesn't account for creativity of language
- Grammars as finite-state machines
 - No representation of structural ambiguity
 - Misses generalizations about structure
 - (Not formally powerful enough)

Context-free Grammar

Chomsky Hierarchy



RQ: Regular Expressions vs. CFGs

“I was a bit confused about the properties and limitations of regular expressions as compared to CFGs. Are regular expressions obligatorily 1:1 mappings of individual words to categories (noun, verb, etc.), making it that they can't be grouped into NPs or VPs, which CFGs address?”

Within the “Chomsky Hierarchy of languages,” languages are understood to be sets of strings. Regular expressions and CFGs are two different types of machinery that are formally capable of describing different classes of languages. CFGs are more powerful and are able to describe a broader set of languages.

ex. $a^n b^n$ is not describable by a regular expression, but a CFG could be defined that would describe this language

Context-Free Grammar

- A quadruple $\langle C, \Sigma, P, S \rangle$
 - C : set of non-terminal symbols (in this context, named for part of speech categories)
 - Σ : set of terminals (vocabulary)
 - P : set of rewrite rules, $\alpha \rightarrow \beta_1, \beta_2, \dots, \beta_n$
 - S : a member of C , the start symbol
 - For each rule, $\alpha \rightarrow \beta_1, \beta_2, \dots, \beta_n \in P$
 - $\alpha \in C$
 - $\beta_i \in C \cup \Sigma$
 - $1 \leq i \leq n$

A Toy Grammar

RULES

$S \rightarrow NP VP$

$NP \rightarrow (D) A^* N PP^*$

$VP \rightarrow V (NP) (PP)$

$PP \rightarrow P NP$

LEXICON

D: the, some

A: big, brown, old

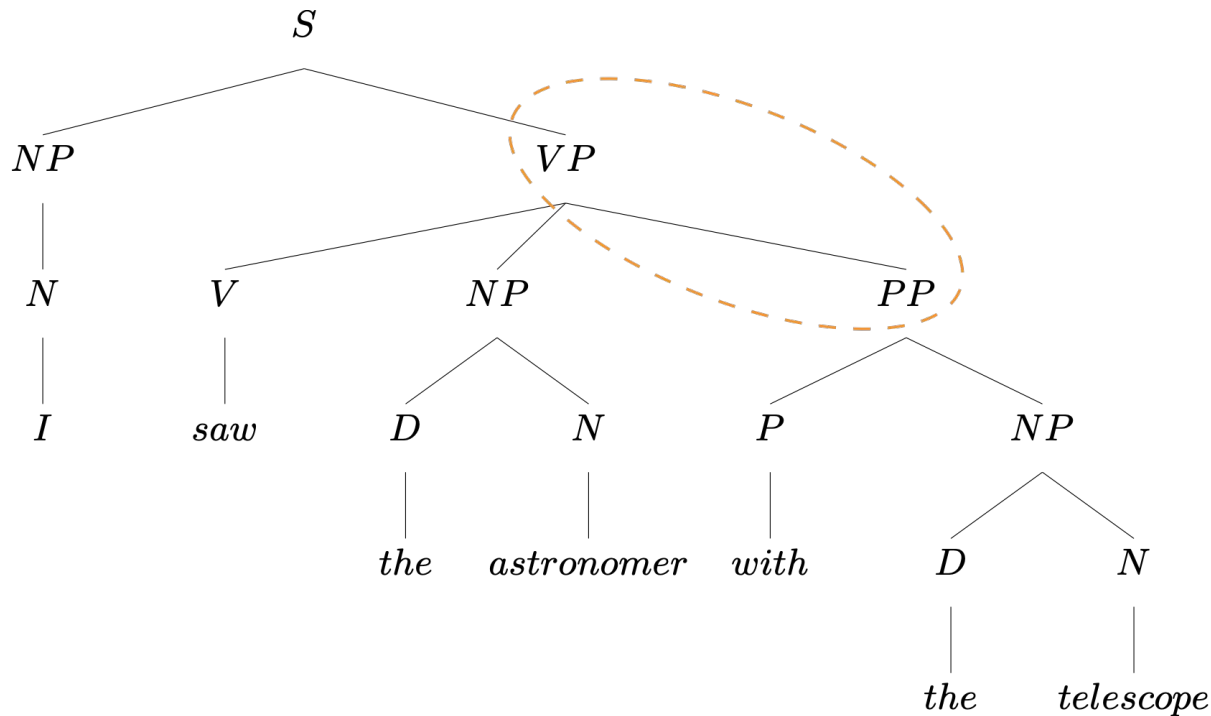
N: birds, fleas, dog, hunter, I

V: attack, ate, watched

P: for, beside, with

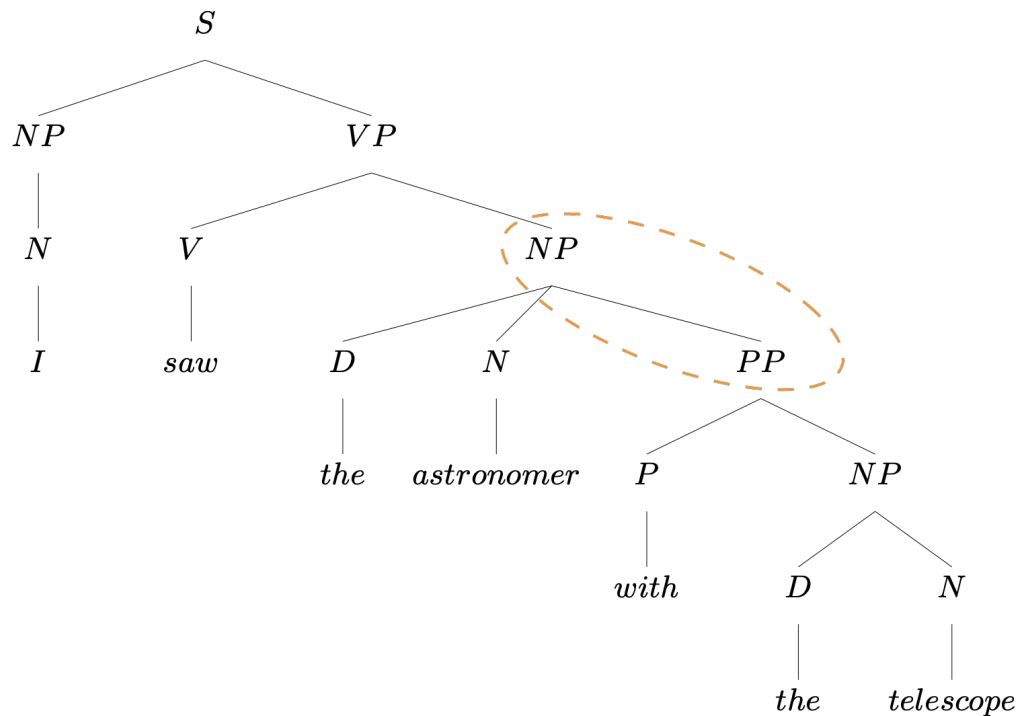
Structural Ambiguity

I saw the astronomer with the telescope.



Structural Ambiguity

I saw the astronomer with the telescope.



Constituency

Constituents

- How do constituents help us? (What's the point?)
 - Can represent structural ambiguity; efficiency (e.g. D A N can be represented as NP)
- What aspect of the grammar determines which words will be modeled as a constituent?
 - The rules
- How do we tell which words to group together into a constituent?
 - Constituency tests (upcoming)
- What does the model claim or predict by grouping words into constituents?
 - If the grammar claims a group of words is a constituent, then it is predicting, for example, that it can be a constituent in a conjunction construction; or that any NP can be used where any other NP can be used, etc

Constituency Tests

- Recurrent patterns
 - ***The quick brown fox with the busy tail** jumped over **the lazy brown dog with one little ear***
- Coordination
 - ***The quick brown fox with the busy tail** and **the lazy brown dog with one little ear** are friends*
- Sentence-initial position
 - ***The election of 2000**, everyone will remember for a long time*
- Cleft sentences
 - *It was **a book about syntax** they were reading*

RQ: Constituency Tests

“p. 33: I would like some clarification on what exactly the "occurrence before the subject of a sentence" diagnostic means”:

“Probably the most common such diagnostic is occurrence before the subject of a sentence. In the appropriate contexts, various types of phrases are acceptable at the beginning of a sentence.”

- Most elections are quickly forgotten, but ***the election of 2000***, everyone will remember for a long time (NP)
- You asked me to fix the drain, and ***fix the drain***, I shall (VP)
- ***In the morning***, they drink tea (PP)
- ****Election of***, everyone will remember for a long time

RQ: Constituency Tests

- **But this test isn't ironclad!**
- ****Election of***, everyone will remember for a long time
 - fails the test, and also is not a constituent
- ****Election***, everyone will remember for a long time
 - fails the test, but IS a constituent
 - but is NOT an NP, so in that sense the test works if we assume it is only able to identify NPs, VPs, and PPs

General Types of Constituency Tests

- Distributional — which words show up with other words?
- Intonational — prosody and how people pronounce things might give clues
- Semantic — how are things relating to each other?
- Psycholinguistic — probe with experiments using things like eye tracking
... but they don't always agree.

In this class we're not going to focus on the details of determining constituents

Central Claims of CFG

Central claims implicit in CFG formalism

1. Parts of sentences (larger than single words) are linguistically significant units
 - a. i.e. phrases play a role in determining meaning, pronunciation, and/or the acceptability of sentences.
2. Phrases are contiguous portions of a sentence
 - a. no discontinuous constituents
3. Two phrases are either disjoint or one fully contains the other
 - a. no partially overlapping constituents
4. What a phrase can consist of depends only on what kind of a phrase it is (that is, the label on its top node), not on what appears around it

Central claims implicit in CFG formalism

- Claims 1-3 characterize what is called ‘phrase structure grammar’
- Claim 4 (that the internal structure of a phrase depends only on what type of phrase it is, not on where it appears) is what makes it ‘context-free’
- There is another kind of phrase structure grammar called ‘context-sensitive grammar’ (CSG) that gives up 4. That is, it allows the applicability of a grammar rule to depend on what is in the neighboring environment
 - So rules can have the form $A \rightarrow X$, in the context of Y_Z

Possible Counterexamples

- To Claim 2 (no discontinuous constituents):
 - **A technician** arrived **who could solve the problem**.
- To Claim 3 (no overlapping constituents):
 - I read **what** was written about me.
- To Claim 4 (context independence):
 - **He arrives** this morning.
 - ***He arrive** this morning.
 - ***They arrives** this morning.
 - **They arrive** this morning.

CFG Mechanics / Examples

A Trivial CFG

$S \rightarrow NP VP$

$NP \rightarrow D N$

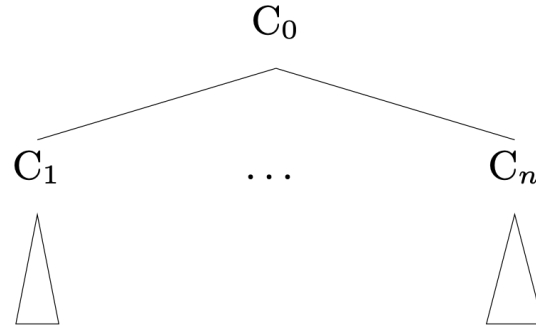
$VP \rightarrow V NP$

D: *the*

V: *chased*

N: *dog, cat*

Trees and Rules



is a well-formed nonlexical tree if (and only if)

$C_1, \dots, C_n$ are well-formed trees, and

Below the text, there are two identical triangles, each with a small dot centered below its base, representing the trees C_1 and C_n .

$C_0 \rightarrow C_1 \dots C_n$ is a grammar rule.

Bottom-up Tree Construction

D: *the*

V: *chased*

N: *dog, cat*

D

|

the

V

|

chased

N

|

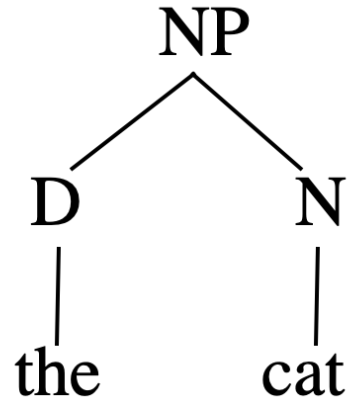
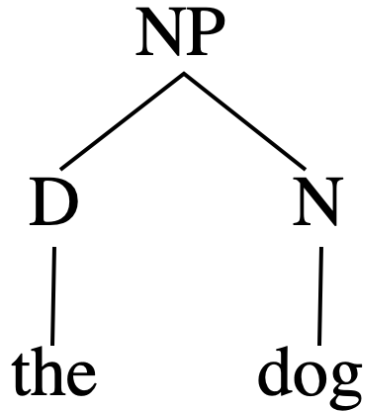
dog

N

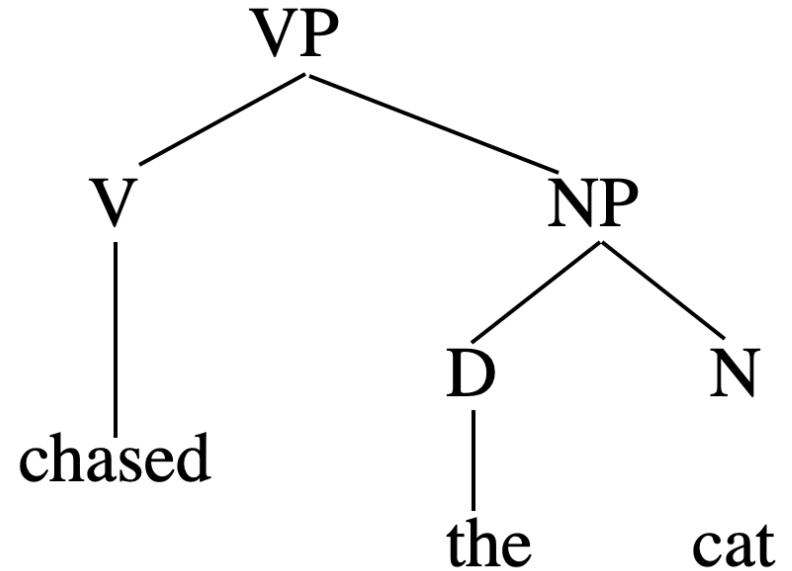
|

cat

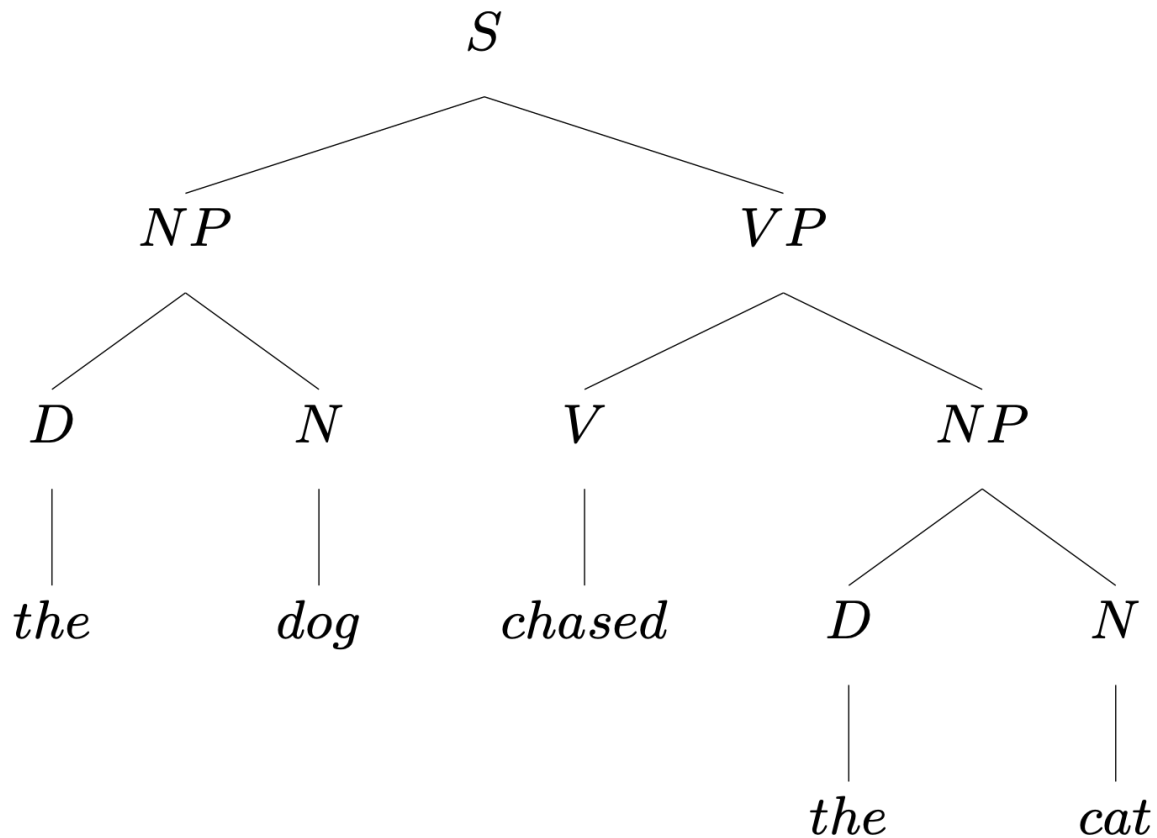
$NP \rightarrow D \ N$



$VP \rightarrow V \ NP$

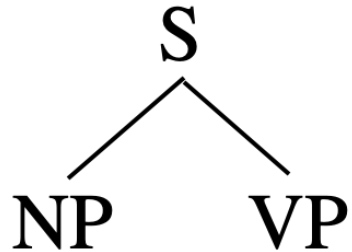


$S \rightarrow NP VP$

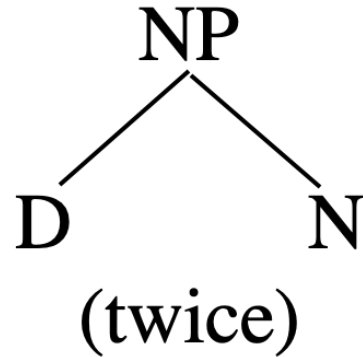


Top-down Tree Construction

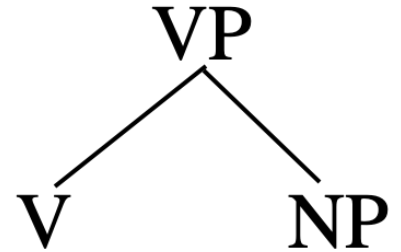
$S \longrightarrow NP \ VP$

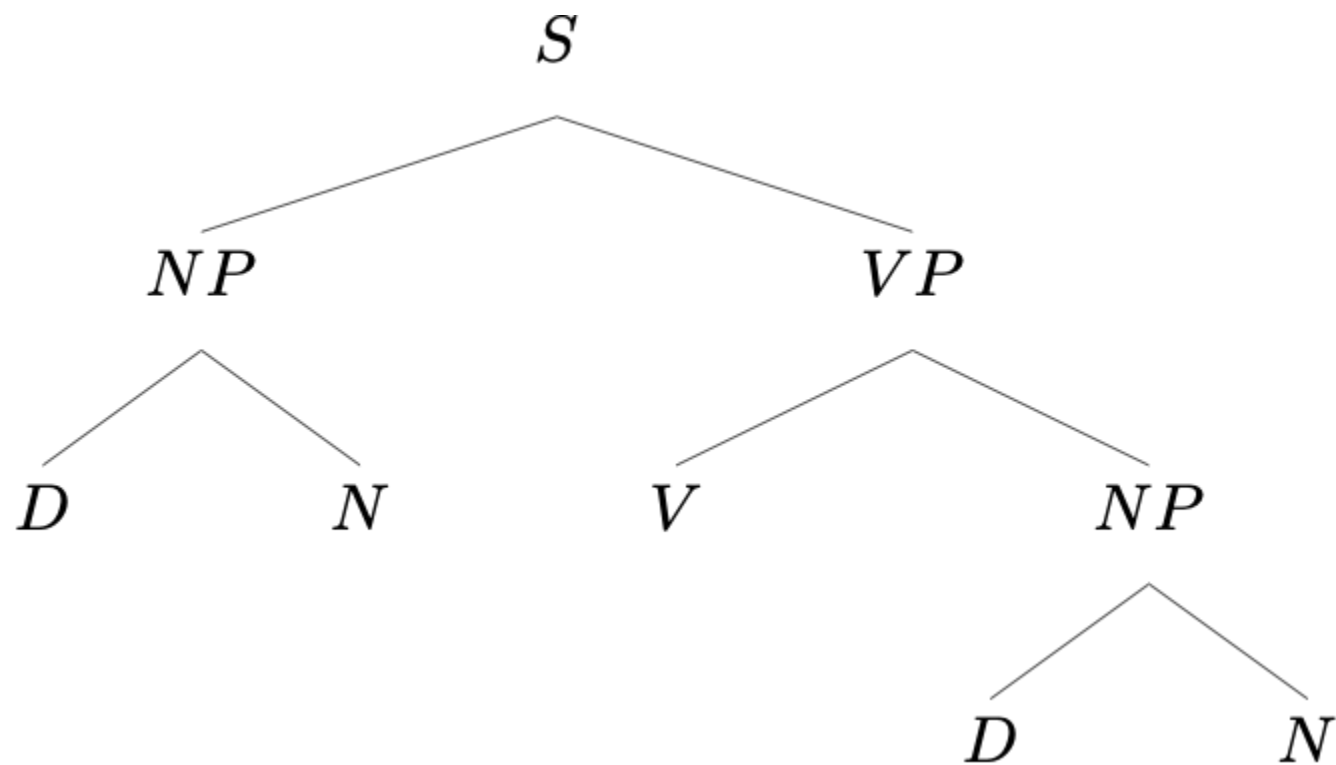


$NP \longrightarrow D \ N$



$VP \longrightarrow V \ NP$





D



the

V



chased

N

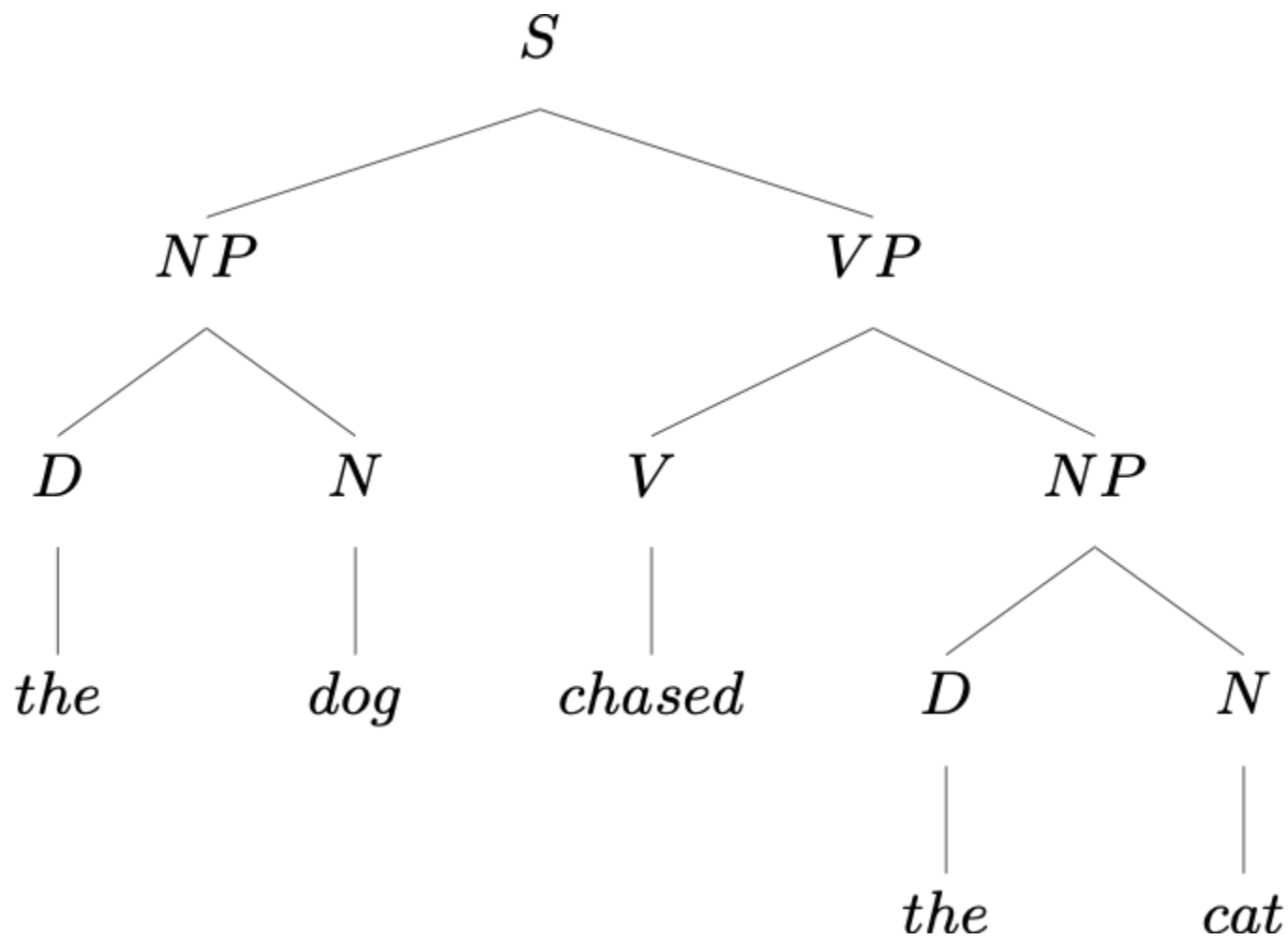


dog

N



cat

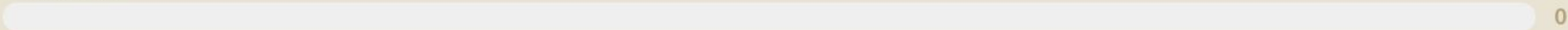




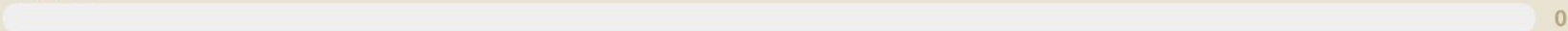
For parsing, which style feels most intuitive to you?



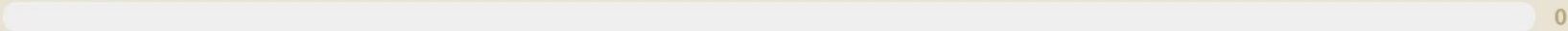
Bottom up



Top down



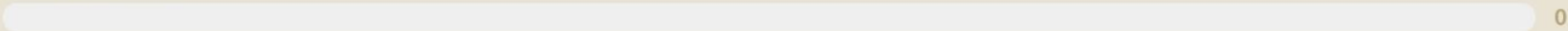
Left to right



Something else



None of the above



RQs: Tree Terminology

“What does "on a branch" mean in the definition for domination ("A node above another on a branch is said to dominate it")? Is it just excluding leaf nodes? i.e., would just saying "A node above another is said to dominate it" be the same, why or why not?”

The idea is that a node must be directly above a node (i.e. immediately dominates) or be an ancestor of it via a series of branches mother-daughter. A node that is “below” another in the tree (i.e. lower down) but is a “niece” would not be said to be dominated by the “aunt”

I think the “on a branch” is trying to get at this but admittedly not very clearly

RQs: Tree Terminology

“I was introduced to trees through the terminology of parents, children, and siblings (with node/branch/terminal/leaf staying the same), so I was wondering if this terminology or the mother/daughter/sister terminology is more common.”

I believe in the syntax world the mother/daughter/sister terminology is more common, and is even encoded into some feature structure names (at least in the ERG), though parent/child/sibling works just as well

Grammar Rules for your HW (ex. 23 from the book)

$S \rightarrow NP VP$

$NOM \rightarrow N$

$NOM \rightarrow NOM PP$

$NP \rightarrow (D) NOM$

$VP \rightarrow V (NP) (NP)$

$VP \rightarrow VP PP$

$PP \rightarrow P NP$

$X \rightarrow X^+ CONJ X$

RQ: Eliding nodes?

“I’m confused about when we should elide nodes with exactly one parent and one child in the trees we draw. In (22), we see NOM leading directly to N (painting) and PP (by Miro). Our rules at this point allow $NOM \rightarrow NOM\ PP$ or $NOM \rightarrow N$ but not directly $NOM \rightarrow N\ PP$ as we see there. Do we just by convention skip the intermediate NOM above painting that leads only to N (painting), substituting that $NOM \rightarrow N$ rule in on the left hand side?”

It’s an error in the book :)

RQ: X?

“Concerning the grammar used for our homework question, how do we apply X to a tree? For example, if it’s two nouns NOM $\rightarrow$ N CONJ N? Or would we want NP $\rightarrow$ NOM CONJ NOM? Does this depend on the rest of the phrase? So maybe $S \rightarrow$ NP CONJ NP VP or $VP \rightarrow$ V NP CONJ NP?”

The X serves as a variable where any of the other non-terminal symbols can take its place (but it must be the same one when you apply it, i.e. you can coordinate NPs with each other, but not NPs with VPs)

RQ: NOM

“I am unclear about why we need to add Nominal (NOM) as an intermediate layer in noun phrases.”

“I didn’t fully understand why we use NOM for the NP $\rightarrow$ (D) NOM rule instead of letting it be a recursive NP rule, NP $\rightarrow$ (D) NP. I noticed the later explanation about NOM not passing the same constituency tests as NP. I also see that within the recursion you could not have a D within the second NP. Are either of these the reason for defining NOM?”

Which rules (w/PP->P NP, NP-> D N, and coord rule) license "no painting by A or drawing by B was displayed"?



0

NP -> NP PP

0

NP -> D NOM; NOM -> NOM PP; NOM -> N

0

NP -> (D) N (PP)

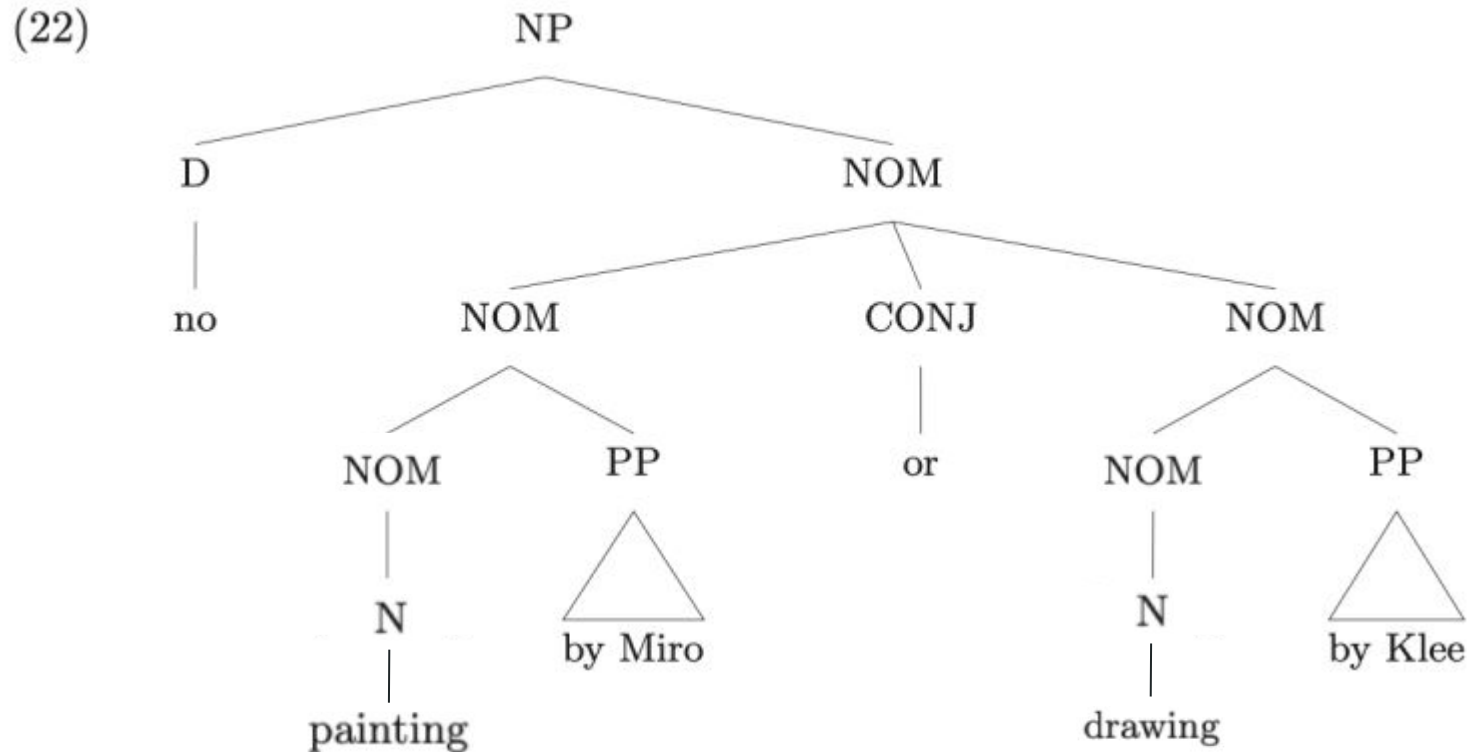
0

None of the above

0



RQs: NOM



RQ: NOM

“The Coordinate conjunction test sorta fails for the NOM examples. The test generates false positives (Eg 21A).”

- a. The love of my life and mother of my children would never do such a thing
- b. The museum displayed no painting by Miro or drawing by Klee

In (a) the individual NOM constituents are *love of my life* and *mother of my children* ... and the whole NOM is *love of my life and mother of my children* and *The* serves as the determiner

Weaknesses of CFGs

Weaknesses of CFG (w/ atomic node labels)

- It doesn't tell us what constitutes a linguistically natural rule

$VP \rightarrow P\ NP$

$NP \rightarrow VP\ S$

- Rules get very cumbersome once we try to deal with things like agreement and transitivity
- It has been argued that certain languages (notably Swiss German and Bambara) contain constructions that are probably beyond the descriptive capacity of CFG.

Agreement & Transitivity

S	→	NP-SG VP-SG	VP-SG	→	IV-SG
S	→	NP-PL VP-PL	VP-PL	→	IV-PL
NP-SG	→	(D) NOM-SG	VP-SG	→	TV-SG NP
NP-PL	→	(D) NOM-PL	VP-PL	→	TV-PL NP
NOM-SG	→	NOM-SG PP	VP-SG	→	DTV-SG NP NP
NOM-PL	→	NOM-PL PP	VP-PL	→	DTV-PL NP NP
NOM-SG	→	N-SG	VP-SG	→	CCV-SG S
NOM-PL	→	N-PL	VP-PL	→	CCV-PL S
NP	→	NP-SG	VP-SG	→	VP-SG PP
NP	→	NP-PL	VP-PL	→	VP-PL PP
	...			...	

Shieber 1985

- Swiss German example:

... mer d'chind em Hans es huus lönd hälfe aastriche

... we the children-ACC Hans-DAT the house-ACC let help paint

... we let the children help Hans paint the house

- Cross-serial dependency:
 - *let* governs case on *children*
 - *help* governs case on *Hans*
 - *paint* governs case on *house*

Shieber 1985

- Define a new language $f(\text{SG})$

$f(\text{d'chind})$	=	a	$f(\text{Jan sait das mer})$	=	w
$f(\text{em Hans})$	=	b	$f(\text{es huus})$	=	x
$f(\text{londe})$	=	c	$f(\text{aastriiche})$	=	y
$f(\text{halfe})$	=	d	$f([\text{other}])$	=	z

- Let r be the regular language $wa^*b^*xc^*d^*y$
- $f(\text{SG}) \cap wa^mb^nc^md^ny$
- $wa^mb^nc^md^ny$ is not context free
- But context free languages are closed under intersection w/ regular languages
- $\therefore f(\text{SG})$ (and by extension Swiss German) must not be context free

Strongly/weakly CF

- A language is weakly context-free if the set of strings in the language can be generated by a CFG
- A language is strongly context-free if the CFG furthermore assigns the “correct” structures to the strings
- Shieber’s argument is that SG is not weakly context-free which entails that it is not strongly context-free

On the other hand...

- It's a simple formalism that can generate infinite languages and assign linguistically plausible structures to them
- Linguistic constructions that are beyond the descriptive power of CFG are rare
- It's computationally tractable and techniques for processing CFGs are well understood

So...

- CFG has been the starting point for most types of generative grammar
- The theory we develop in this course is an extension of CFG
 - but at some point will go beyond the descriptive power of CFG!

Additional RQs

RQs: Transformations

“...I am confused on how transformations can help solve the complexity of English subject-verb agreement for CFGs when it is also just a rule that is adding an s- wouldn't this also require some kind of context?”

“Should we understand transformations to basically always be intermediary representations/mappings between two grammatically correct trees? Also, in the future will we be expected to apply transformations ourselves given an input and output tree?”

“What makes HPSG more appealing than transformational generative grammar for natural language processing? How do they differ in their integration of semantic information?”

RQs: Transformations

- TL;DR ... not our problem <3
- ... but on a more serious note
- If you don't have experience with transformational theories of grammar, don't fret, you don't need to in order to learn the material in this class
- If you DO have experience with transformational theories of grammar, hopefully this class will provide an interesting alternative
 - and hopefully you'll be converted >:)

RQs: Designing a Grammar

“Since a grammar is a model of a language, not the language itself, it seems like some property will always be missing compared to a language. Do linguists accept that every attempt in this work might fall short somewhere (Or, is that not the case)? If so, how do we decide which shortcomings matter and in what degree?”

“I was a little confused about how specific a grammar should be. For example, dividing verbs into different categories helps CFG distinguish which verbs can take objects, but it also says that this can create a lot of redundancy. How do we decide when adding more specific categories is helpful and when it just makes the grammar unnecessarily complicated?”

RQs: Sociolinguistics

“The basic premise of the study of syntax is that languages are governed by rules. If this is really the case, how can we account for the difference in acceptability of some sentences (e.g. Some sentences are definitely bad while some are partly acceptable for native speakers)?”

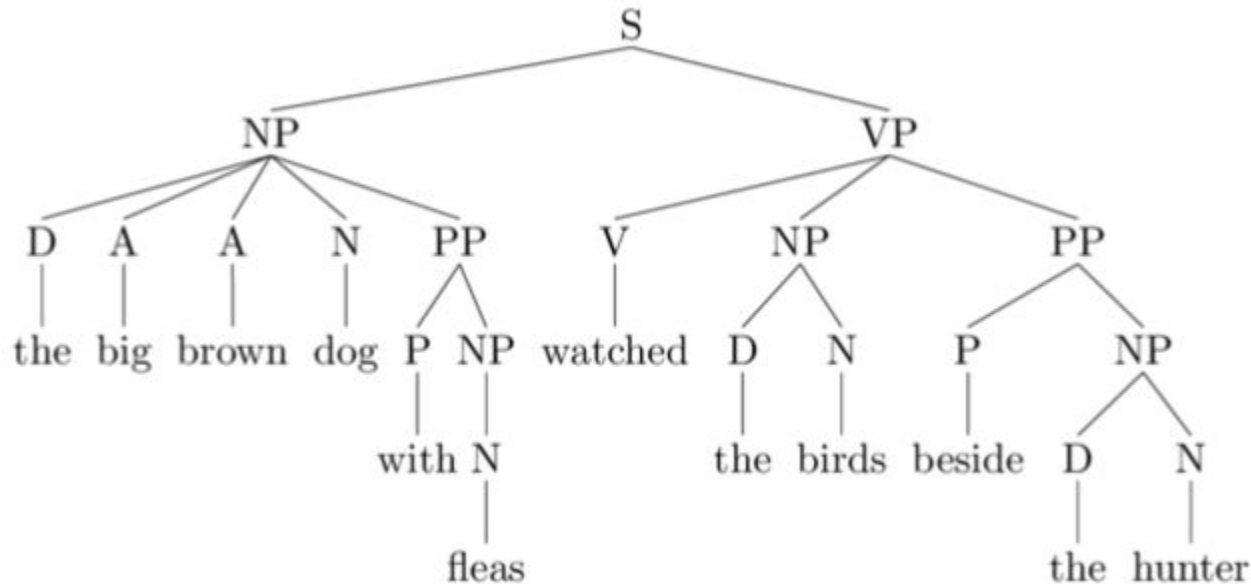
“When speakers show variations in their grammatical patterns, how can we determine whether the variation reflects differences in their linguistic backgrounds rather than nonstandard language use?”

Group Activity

Exercise 1: Practice with CFG

Assume the CFG grammar given in (11). Draw the tree structure for the other interpretation (i.e. not the one shown in (12)) of *The big brown dog with fleas watched the birds beside the hunter*.

(12)



This would keep the number of VP-SG and VP-PL rules down to three each (rather than seven each), but it introduces extra noun phrase categories. Either way, the rules are full of undesirable redundancy.

Matters would get even worse when we examine a wider range of verb types. So far, we have only considered how many NPs must follow each verb. But there are verbs that only appear in other environments; for example, some verbs require following PPs or Ss, as in (42).

- (42) a. Terry wallowed in self-pity.
b.*Terry wallowed.
c.*Terry wallowed the self-pity.
d. Kerry remarked (that) it was late.
e.*Kerry remarked.
f.*Kerry remarked the time.

Exercise 2: Wallowing in Categories

- A. Provide examples showing that the verbs *wallow* and *remark* exhibit the same agreement patterns as the other types of verbs we have been discussing.
- B. What additional categories and rules would be required to handle these verbs?
-

HW Reminders

HW Reminders

- RQ Ch. 3 due **Wednesday, 10/7 @ 6am**
- HW1 due **Friday, 10/9 @ 11:59pm**