

NATURAL LANGUAGE SYNTAX

CS662: Natural Language Processing

2015-02-05

How does word order relate to meaning?

If you told your friend you have to go home to do something, why can't your friend later ask you *What did you go home because you needed to do?*

If *Tim told Tom that he needed some time off*, who exactly wants a vacation?

(for another time) Can you use finite-state automata to parse natural languages? Can natural languages be parsed in polynomial time or better?

OUTLINE

- Constituency
- WH-movement (syntactic displacement)
- Case Theory (conditions on noun phrase position)
- Binding Theory (conditions on pronoun interpretation)

CONSTITUENCY

STRUCTURE DEPENDENCE

The dog *is* large.
Is the dog large?

The ragtime pianist who used to date Kitty *is* here.
Is the ragtime pianist who used to date Kitty here?

The rule of “yes/no” question formation reorders structural units — *constituents* — not just words.

STRUCTURAL AMBIGUITY

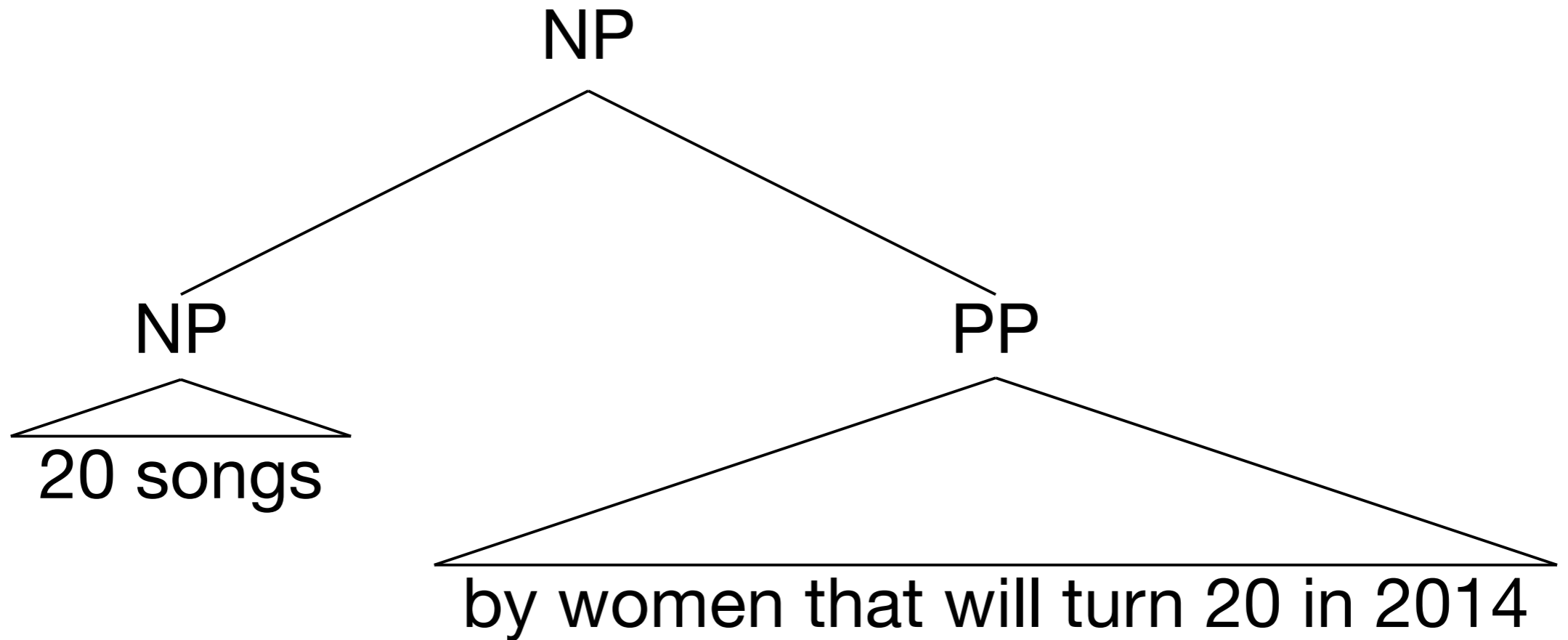
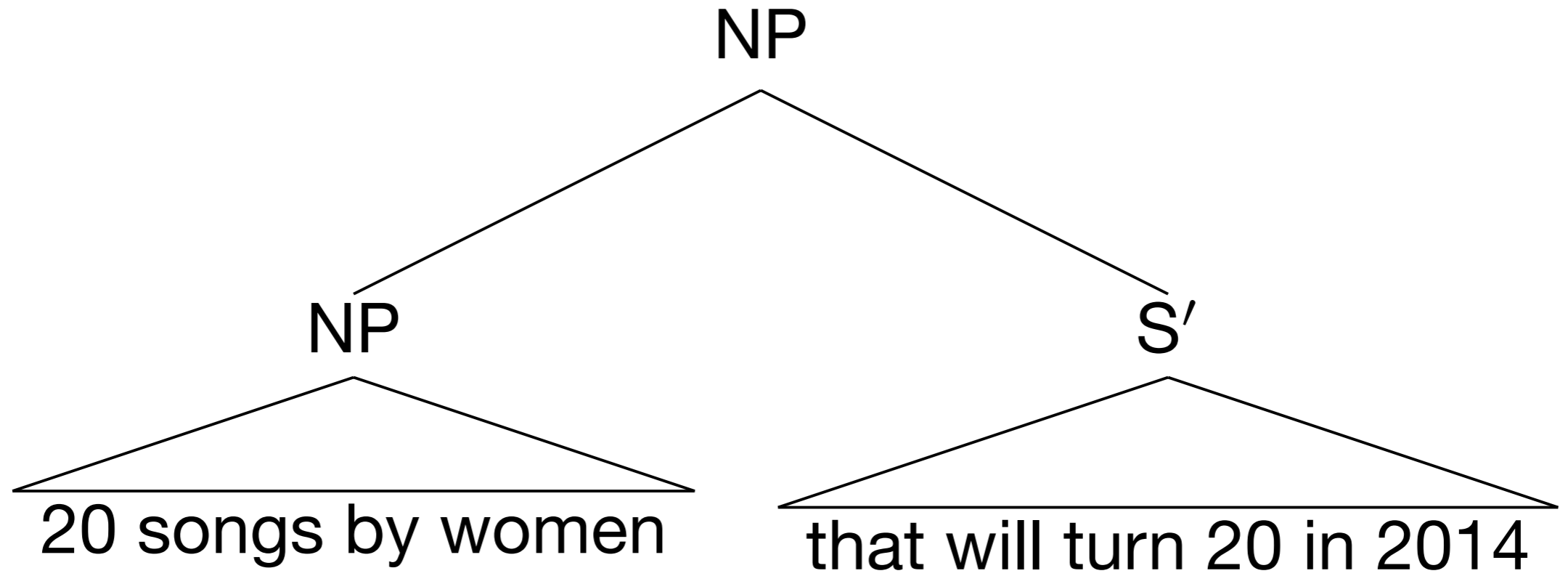
“20 Songs By Women That Will Turn 20 In 2014”:

[20 Songs By Women] [That Will Turn 20 In 2014]

e.g., Tori Amos - Cornflake Girl

[20 Songs] [By Women That Will Turn 20 In 2014]

e.g., Cher Lloyd - Want U Back



CONSTITUENCY TESTS

Substitution: e.g., pronouns can replace noun phrases

The ragtime pianist who used to date Kitty is here.
He is here.

Movement: e.g., adverb phrases can be preposed

The dog snatched the taco as quick as could be.
As quick as could be, the dog snatched the taco.

WH-MOVEMENT

WH-MOVEMENT

Direct WH-questions:

Gisi will pick the red horse.

[Which horse]_{*i*} will Gisi pick *t_i*?

Indirect WH-questions:

Johannes knows Gisi will pick the red horse.

Johannes wonders [which horse]_{*i*} Gisi will pick *t_i*.

WH-MOVEMENT ELEMENTS

[Which house]_{*i*} is Mary's *t_i*?

The subscript *i* indicates things that are *coindexed*

The square brackets denote the WH-phrase

t_i is a *trace* indicating the position the WH-phrase (with the same subscript) is to be interpreted at

APPARENT UNBOUNDEDNESS

What_{*i*} is he reading *t_{*i*}*?

What_{*i*} did he say he was reading *t_{*i*}*?

What_{*i*} does she believe he said he was reading *t_{*i*}*?

What_{*i*} are they claiming she believes he said he was reading *t_{*i*}*?

What_{*i*} do you think they are claiming she believes he said he was reading *t_{*i*}*?

ROSS'S ISLAND CONSTRAINTS

Coordinate structures:

*What_{*i*} did Morty cook *t_i* and wash the dishes?

*What_{*i*} did Morty cook chili and wash *t_i*?

Complex NPs:

*[How many cities]_{*i*} did you hear the rumor that Quinn has visited?

*What_{*i*} does Kiwi believe the report that Alan eats *t_i*?

CROSSLINGUISTIC VARIATION IN WH-MOVEMENT

WH-in-situ languages: e.g., Mandarin

Judou xiang-zhidao shei mai-le shenme

Judou want-know who buy-ASP what

‘For which person does J. wonder what that person bought?’

‘For which thing does J. wonder what person bought it?’

Multiple-WH languages: e.g., Romanian

Cine cui ce ziceai a i-a promis?

who to.whom what say.2SG.PST that to.him-has promise

‘Who do you say promised what to whom?’

CASE THEORY

TWO SENSES OF “CASE”

1. case (small-c): inflectional morphology encoding the grammatical function of noun phrases

Der Mann sieht den Hund.

(German)

Den Hund sieht *der* Mann.

Der Hund sieht den Mann.

Den Mann sieht *der* Hund.

...clāmōr ad cael-um uoluendus per aether-*a* uāgit (Latin)

2. Case (big-C): an abstract relationship between verbs and noun phrases governing the position of the latter

PRONOUN CASE

She will help them.

*Them will help *she*.

Ulf expected her to forget the caviar.

**Ulf* expected she to forget the caviar.

TRANSITIVITY AND CASE ALIGNMENT

Transitive verbs have at least one object:

Olga *loves* empanadas.

Intransitive verbs have no object:

Unergative verbs have an agent subject:

Björn *hunts*.

Unaccusative verbs have an patient subject:

The vase *broke*.

ERGATIVITY

Nominative-accusative language: (e.g., Japanese)

Otoko ga tsuita.
man NOM arrived
'the man arrived'

Otoko ga kodomo o mita.
man NOM child ACC saw
'the man saw the child'

Ergative-absolutive language: (e.g., Basque)

Gizon-a etorri da.
man-ABS has arrived

Gizon-ak mutil-a ikusi du.
man-ERG boy-ABS saw

RAISING & CONTROL VERBS

Raising to object:

Stephanie expects *her* to concede.

Object control:

Daniel asked *him* to leave.

Subject control:

Denny refuses to help out around the house.

PARENT ANNOTATION AND LEXICALIZATION

To capture these behaviors, probabilistic constituent parsing grammars use information about:

parent annotation: parent of a non-terminal

e.g., NP^S is an NP subject (probably)

lexicalization: *head* daughter of a non-terminal

e.g., VP(*expects*) is a raising verb-headed phrase

BINDING THEORY

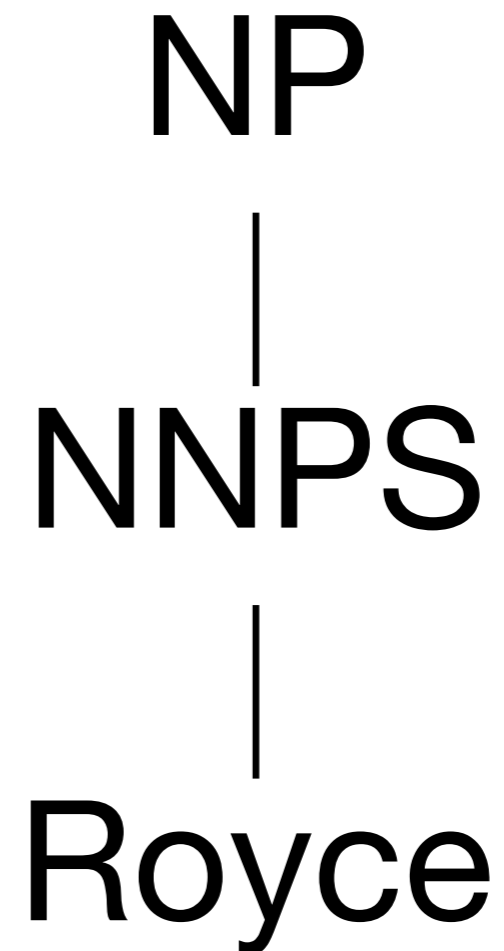
NODE TYPES

Royce is a terminal node

NP and *NNPS* are non-terminal nodes

NNPS is the mother of *Royce*

Royce is the daughter of *NNPS*



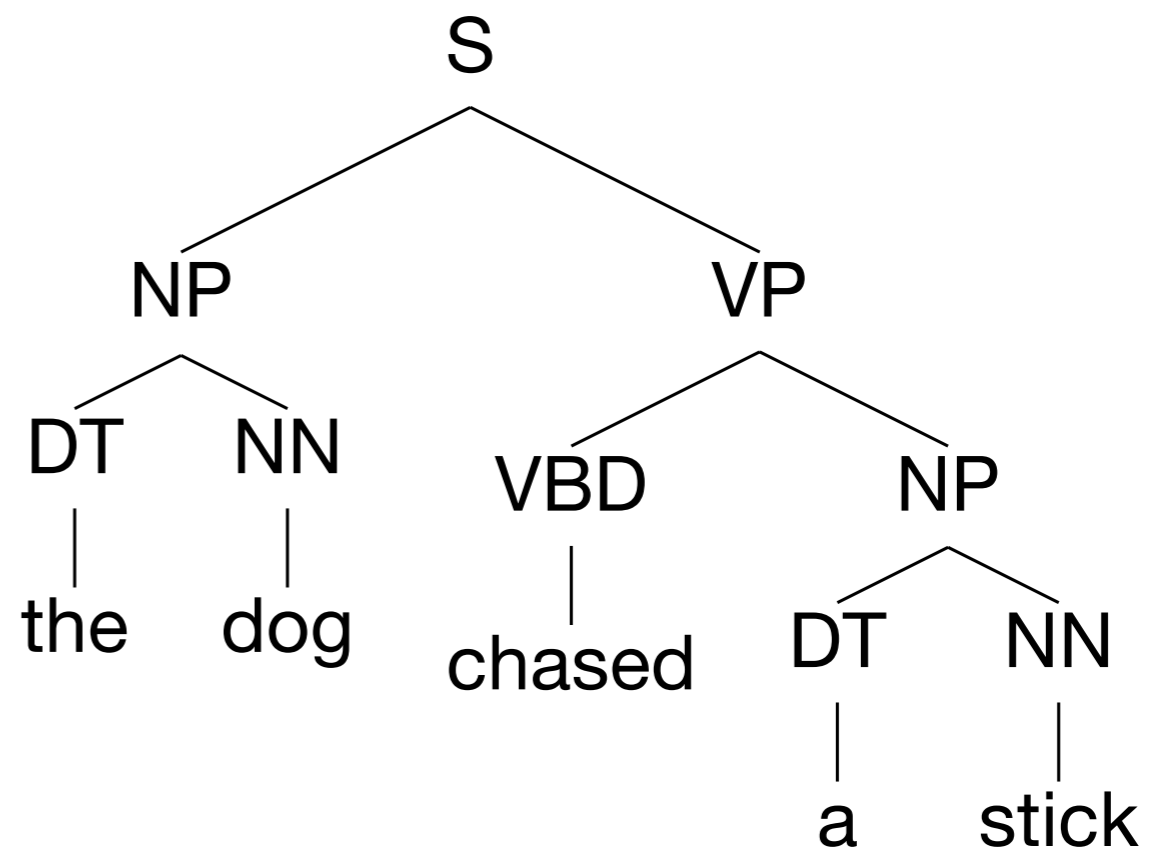
NODE RELATIONS (1/2)

A dominates B if and only if A is “above” B (A is a parent/grandparent, etc. of B)

VP dominates VBD , NP , DT , NN and the last 3 terminals

A immediately dominates B if and only if A dominates B and there is no node C ($\neq A, B$) such that A dominates C and C dominates B

S immediately dominates NP and VP



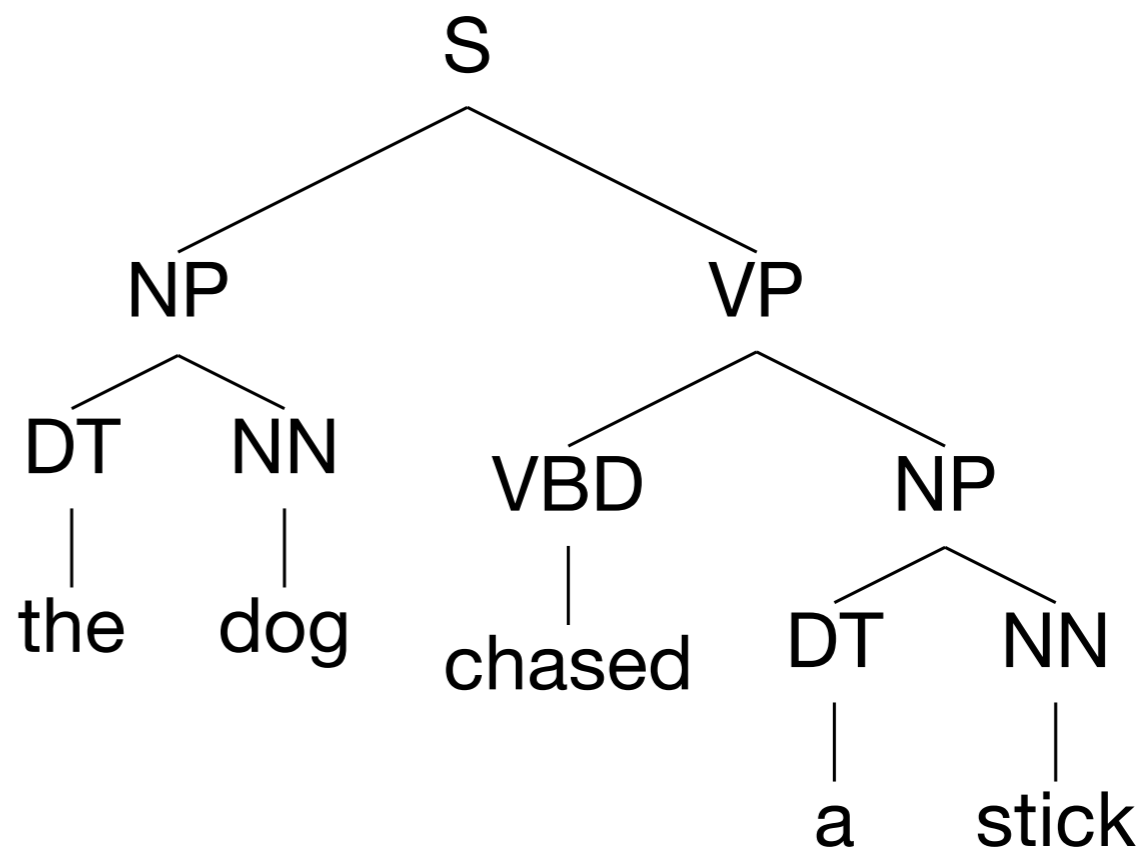
NODE RELATIONS (2/2)

A c-commands B if and only if neither A nor B dominates the other, and the lowest branching node that dominates A also dominates B

VP c-commands “dog”

“dog” doesn’t c-command VP

Sisters c-command each other



NOUN PHRASE TYPES

1. Full noun phrases: *the question, the student that asked the question*
2. Ordinary personal pronouns: *I, you, they*
3. Reflexive pronouns: *myself, yourself, themselves*
4. Reciprocal pronouns: *each other, one another*

ANAPHORA

Pronouns are noun phrases that necessarily refer to an *antecedent* in the discourse.

Determining the antecedents of pronouns is known as *anaphora resolution*.

This is an ill-posed problem: it is not always possible to determine what a pronoun's antecedent is:

Tim told *Tom* that *he* needed some time off.
(is *he* Tim, Tom, or some other male antecedent?)

PERSONAL PRONOUNS

*Paz_i helped her_i.

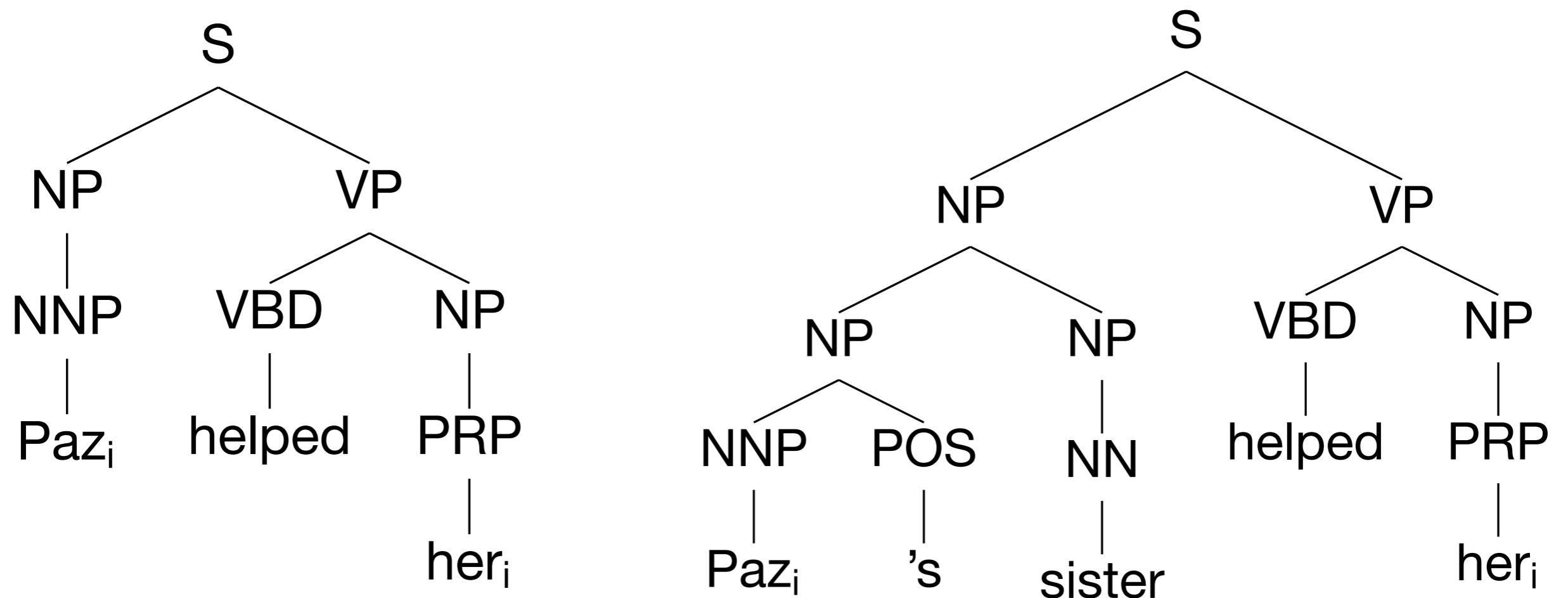
*[Paz's sister]_i helped her_i.

Paz_i's sister helped her_i.

*She_i knew Ginny_i.

PRINCIPLE B

(to be revised, but not here) a personal pronoun cannot be bound by a c-commanding antecedent



LINEAR ORDER DOES NOT MATTER

If Taissa_{*i*} tries, she_{*i*} will succeed.

If she_{*i*} tries, Taissa_{*i*} will succeed.

*At Sofía_{*i*}'s house she_{*i*} spent many days.

*She_{*i*} spent many days at Sofía_{*i*}'s house.

REFLEXIVES & RECIPROCAL

Paz_i helped herself_i.

[Paz's sister]_i helped herself_i.

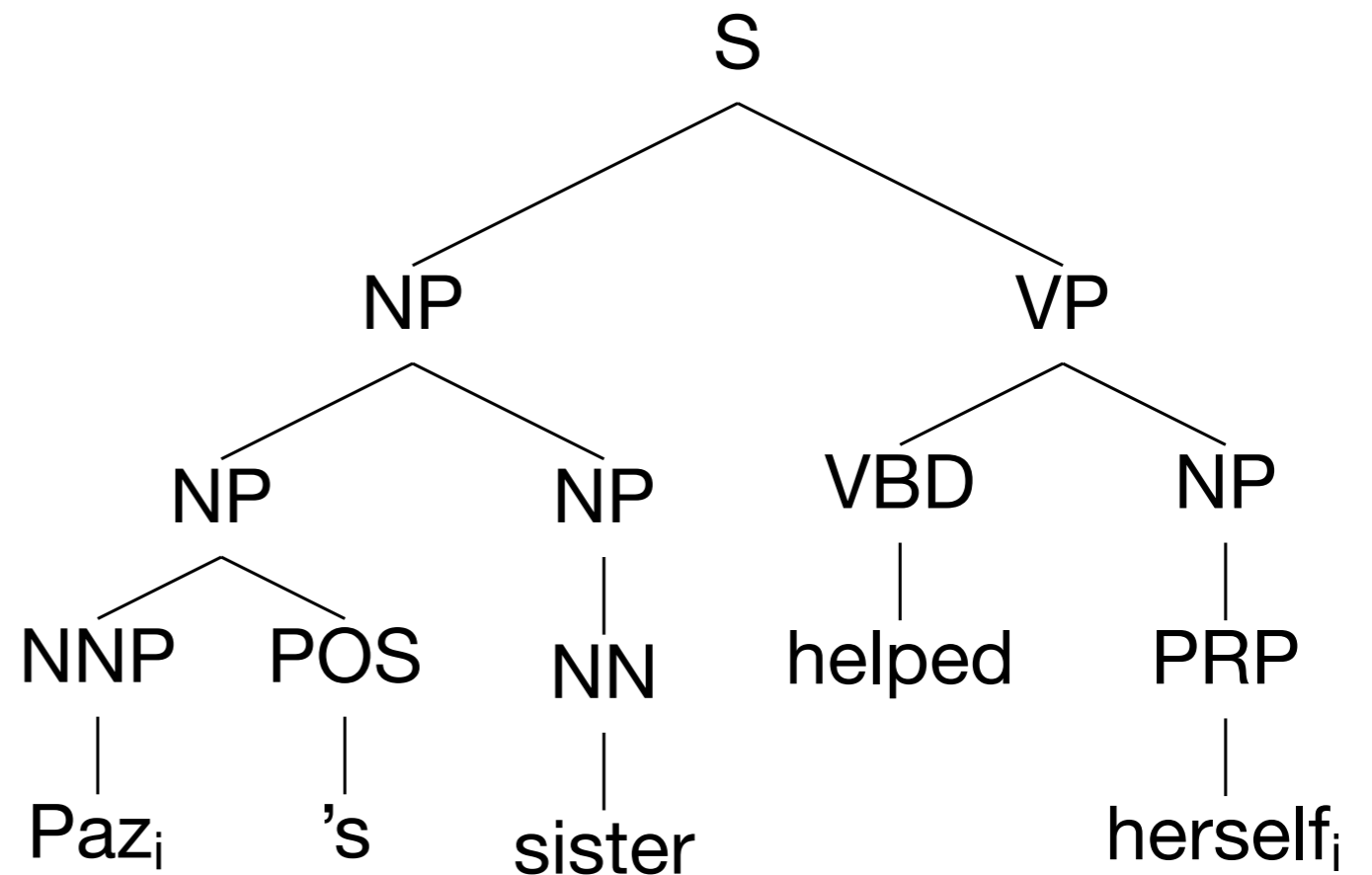
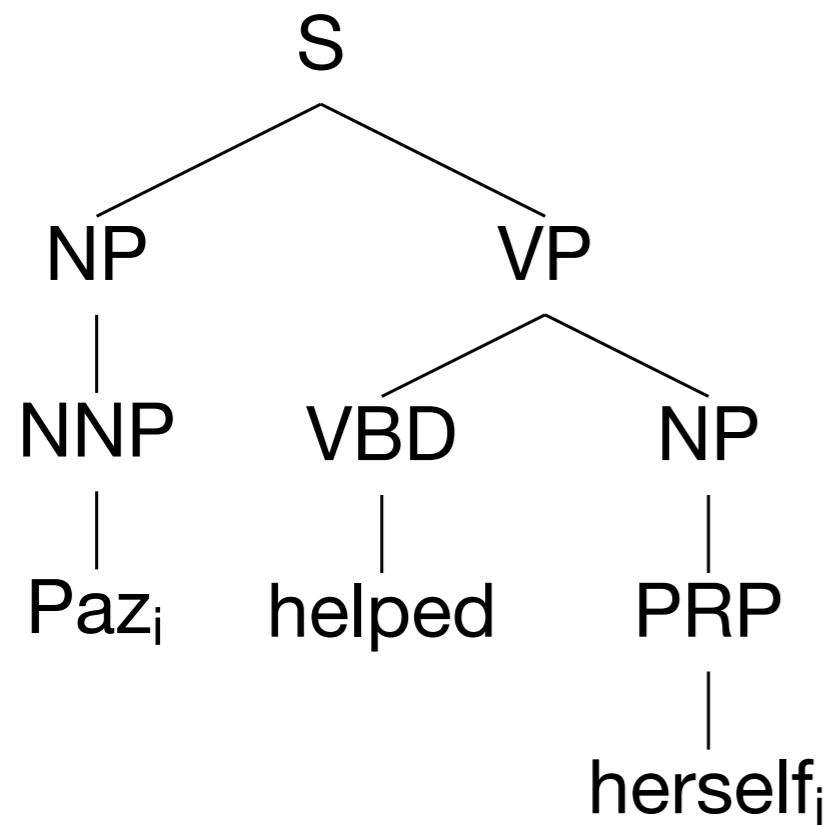
*Paz_i's sister helped herself_i.

*Herself_i knew Violet_i.

*Violet knows [each other]_i?

PRINCIPLE A

(to be revised, but not here) a reflexive or reciprocal must be bound by a c-commanding antecedent



COMPUTATIONAL MODELS OF ANAPHORA RESOLUTION

- Largely operate over constituency parse trees
- Soft (stochastic) preferences are very useful, e.g.:
 - antecedents tend to be syntactic subjects
 - anaphors and their antecedents tend to be “close” in linear order
- Hobb’s algorithm (for finding personal pronoun antecedents): breadth-first, left-to-right search for a candidate NP matching in gender and number