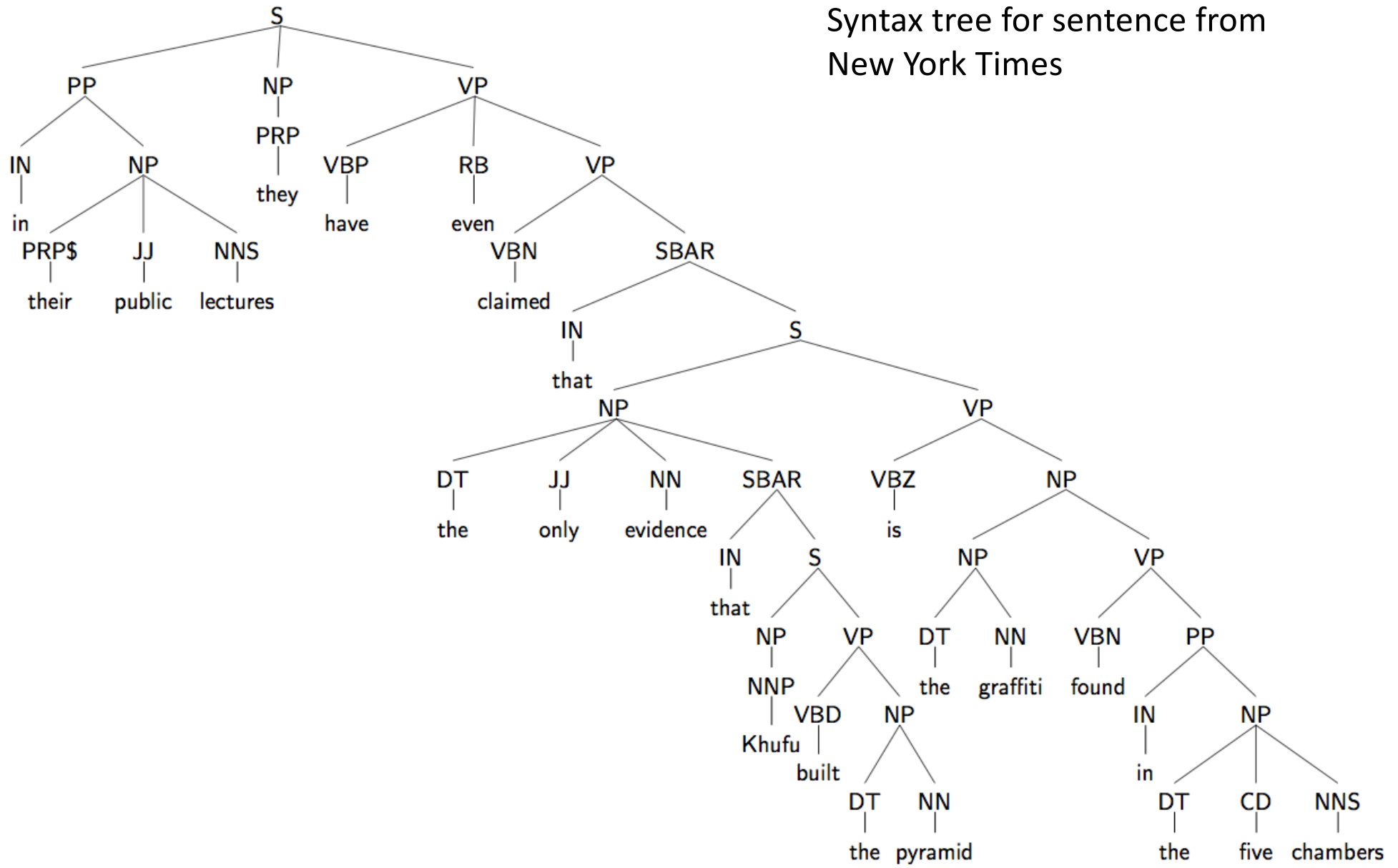


Tree Syntax of Natural Language

Computational Linguistics

Spring 2023

Syntax tree for sentence from
New York Times



Pre-terminal labels

| label | long name | example |
|--------------|--|----------------|
| NN | singular noun | pyramid |
| NNS | plural noun | lectures |
| NNP | proper noun | Khufu |
| VBD | past tense verb | claimed |
| VBZ | 3rd person singular present tense verb | is |
| VBP | non-3rd person sin- gular present tense verb | have |
| VBN | past participle | found |

Pre-terminal labels cont.

| | | | |
|-------|--------------------|--------|-------------|
| PRP | pronoun | they | D |
| PRP\$ | possessive pronoun | their | D[case=gen] |
| JJ | adjective | public | A |
| IN | preposition | in | P |
| | complementizer | that | C |
| DT | determiner | the | D |

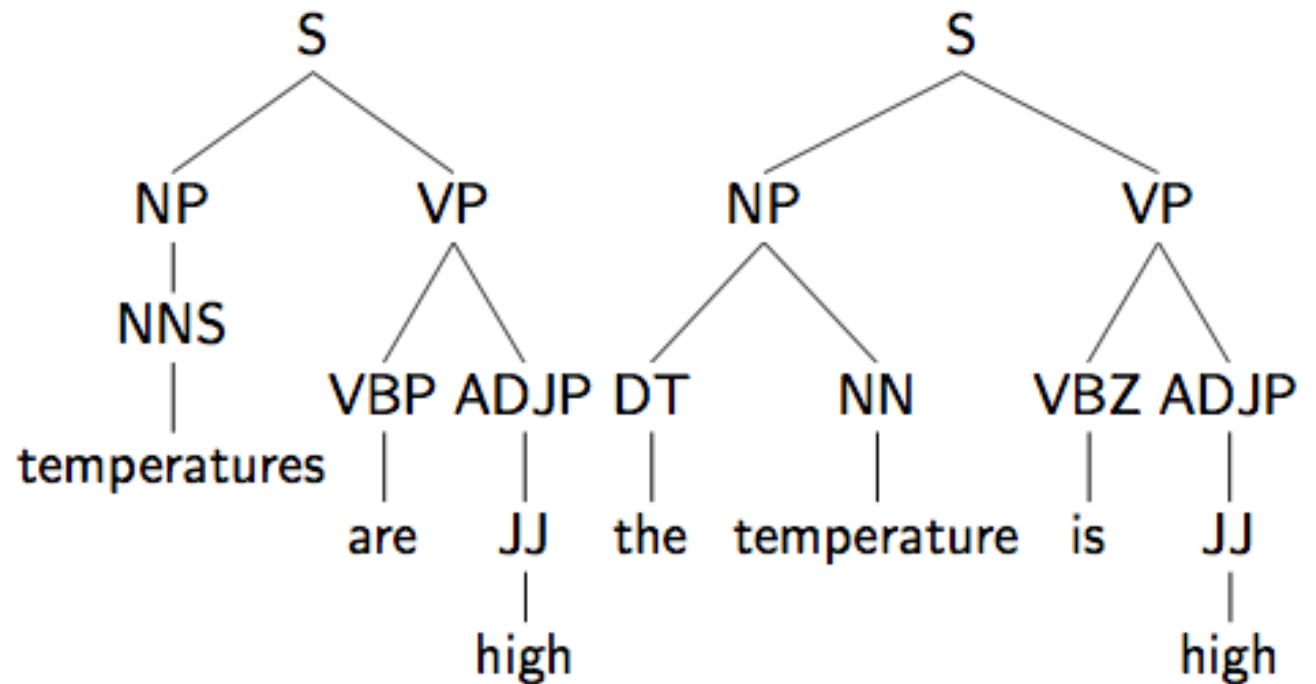
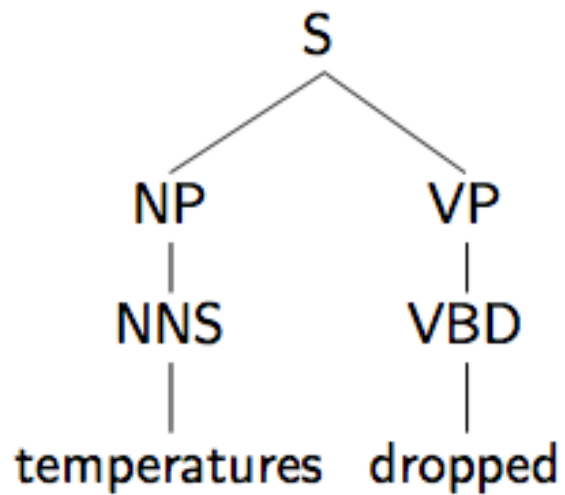
Phrasal labels
Quasi-substitutability
Reference for NP

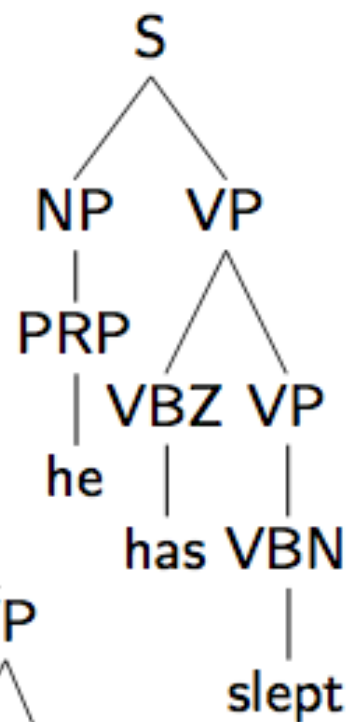
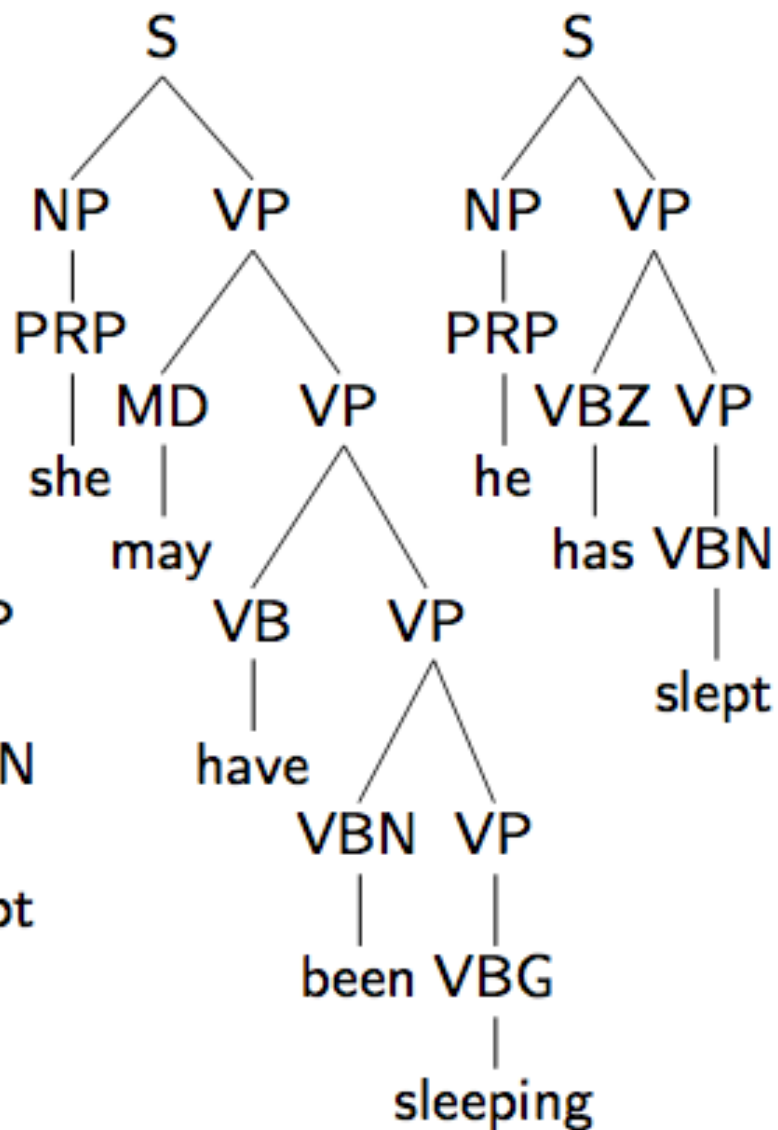
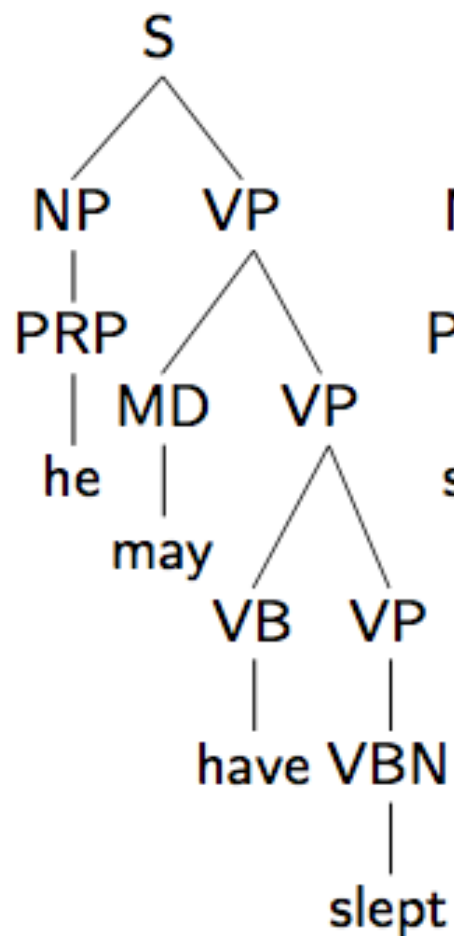
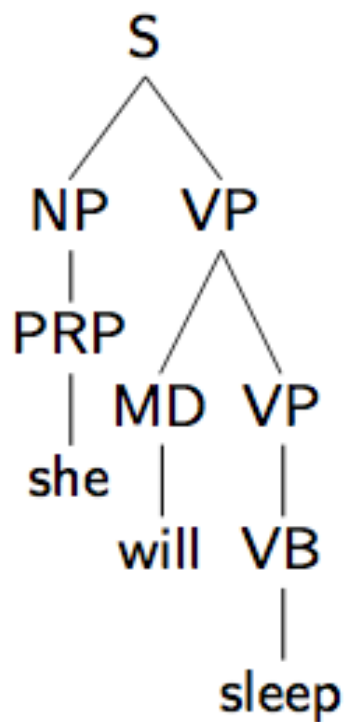
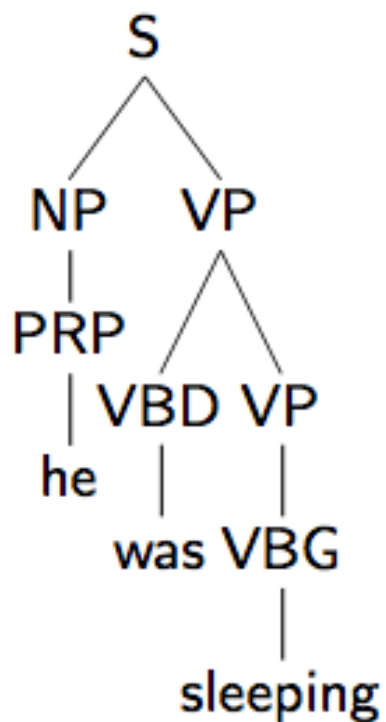
| label | long name | example (represented by terminal string) |
|--------------|----------------------|---|
| NP | noun phrase | their public lectures |
| VP | verb phrase | built the pyramid |
| PP | prepositional phrase | in the five chambers |
| S | sentence | Khufu built the pyramid |
| SBAR | sbar | that Khufu built the pyramid |

Tensed sentences

Truth value

Complete Independent
utterance





VP recursion
Preterminal
TP

| | | |
|-----|---------------|---|
| VB | | |
| VBD | past tense | He ate/VBD the cookies. She answered/VBD the question. |
| VBZ | present tense | He likes/VBZ cookies. |
| VB | | |
| VBP | present tense | They like/VBP cookies. |
| | 3rd person | They answer/VBP such questions. |
| | plural | They are/VBP tired. |

Verb tags as feature structures

V[tns=pst]

V[tns=prs]

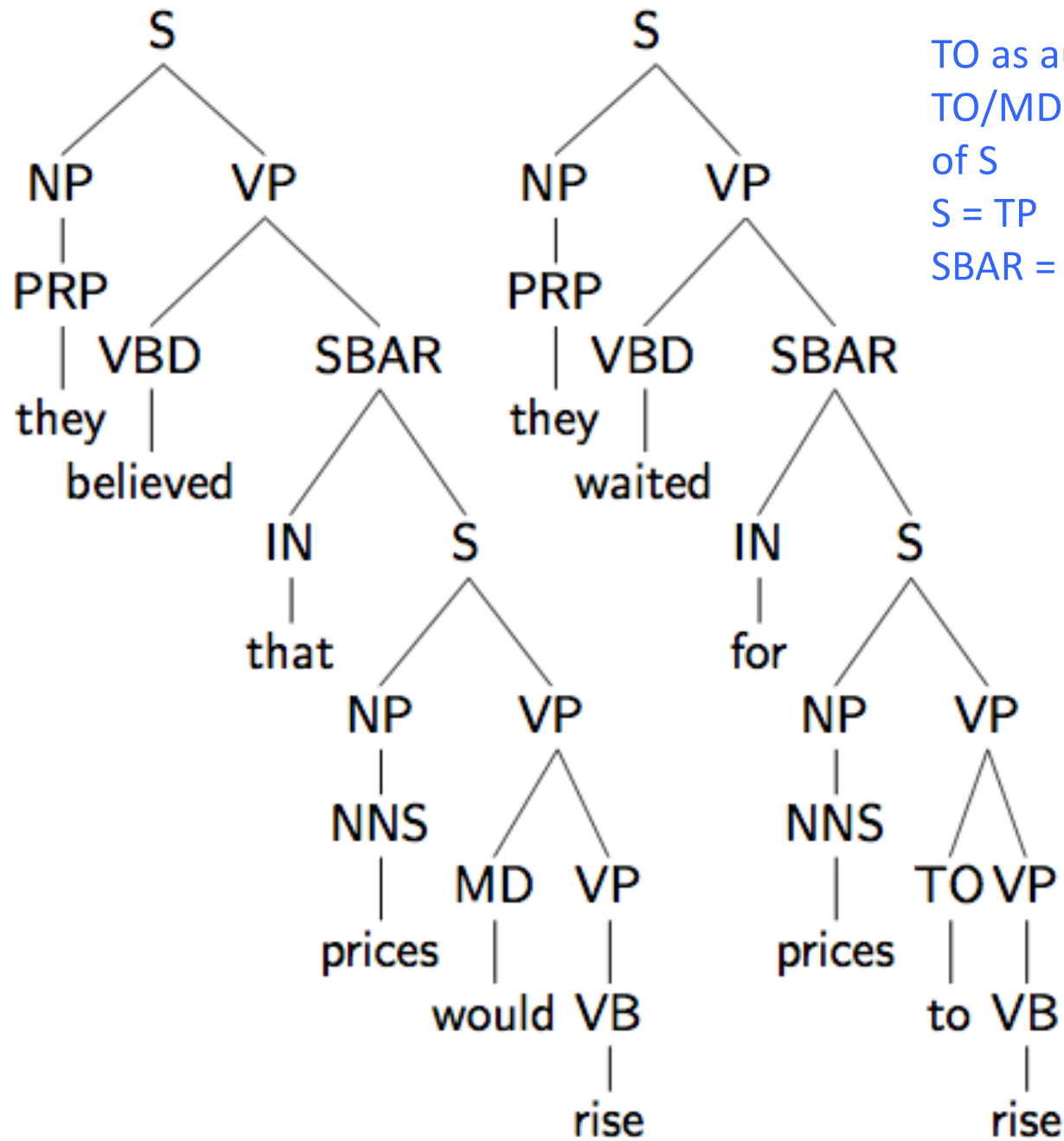
V[vform=s]

V[vform=d]

V[vform=base]

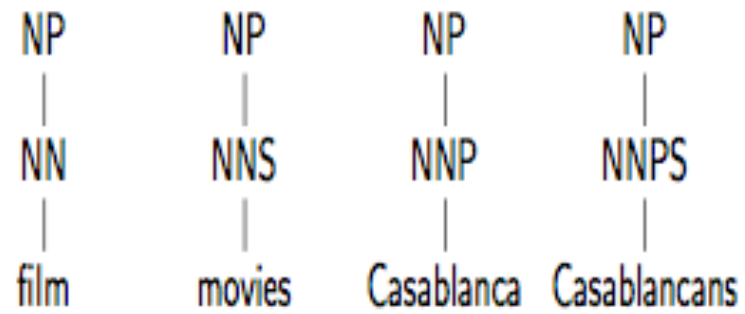
| Tag | Long name | Example |
|-----|----------------------------|---|
| VB | base | He may like/VB cookies. I heard her answer/VB the question. They may be/VB tired. |
| VBG | present participle, G-form | Eating/VG cookies is unhealthy. He likes eating/VG cookies. |
| VBN | past participle, N-form | He has eaten/VBN the cookies. She has answered/VBN the questions. My question was not answered/VBN. |
| MD | modal | She will/MD prevail. |
| TO | auxiliary <i>to</i> | She expects to/TO prevail. |

Form / place in tree

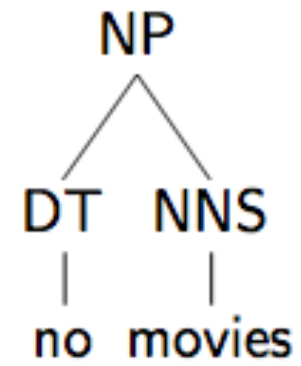
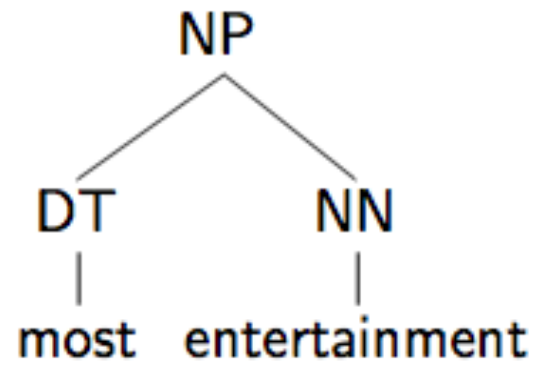
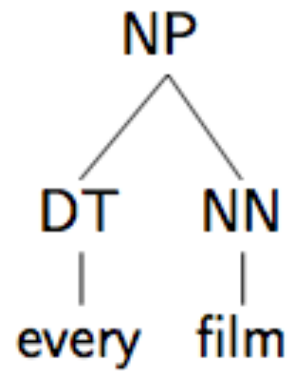


TO as aux verb
TO/MD as head
of S
S = TP
SBAR = CP

Minimal NP



Determiner
head
DP hypothesis
Xbar system



Many impressed me.
Each impressed me.
Some impressed me.
*The impressed me.
*A impressed me.
*Every impressed me.

Isolated DT

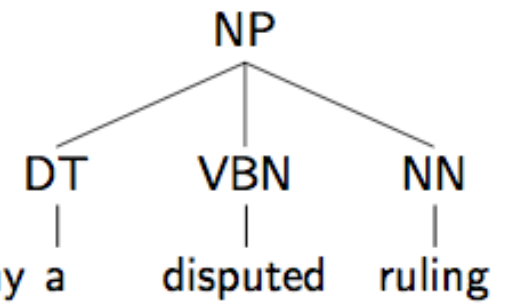
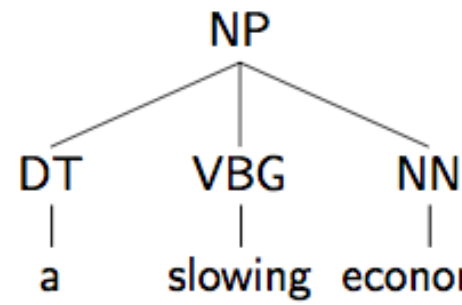
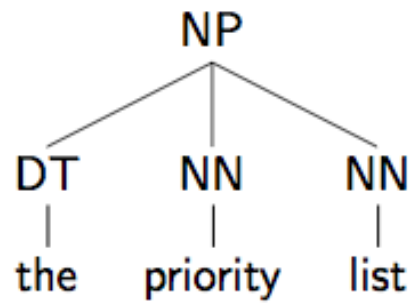
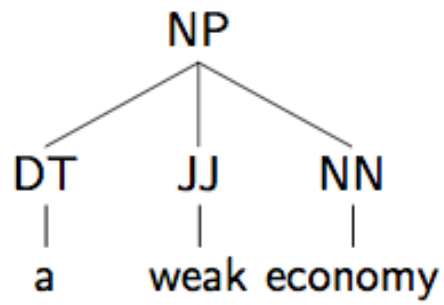
* Notation

ellipsis

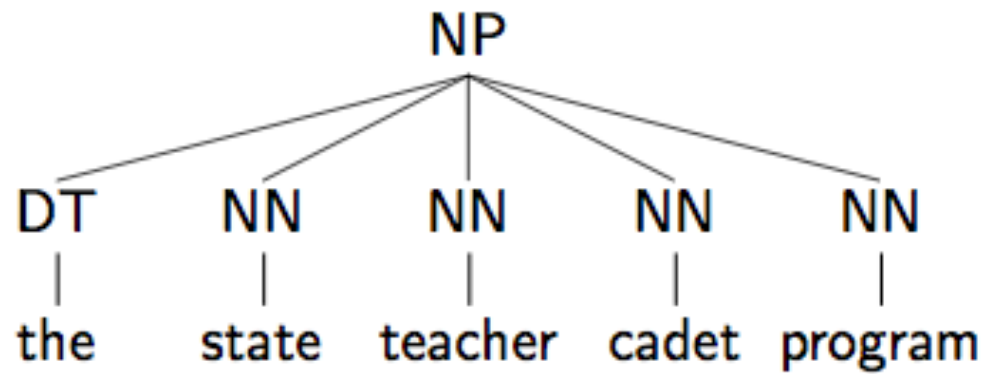
empty category

NP
|
DT
|
some

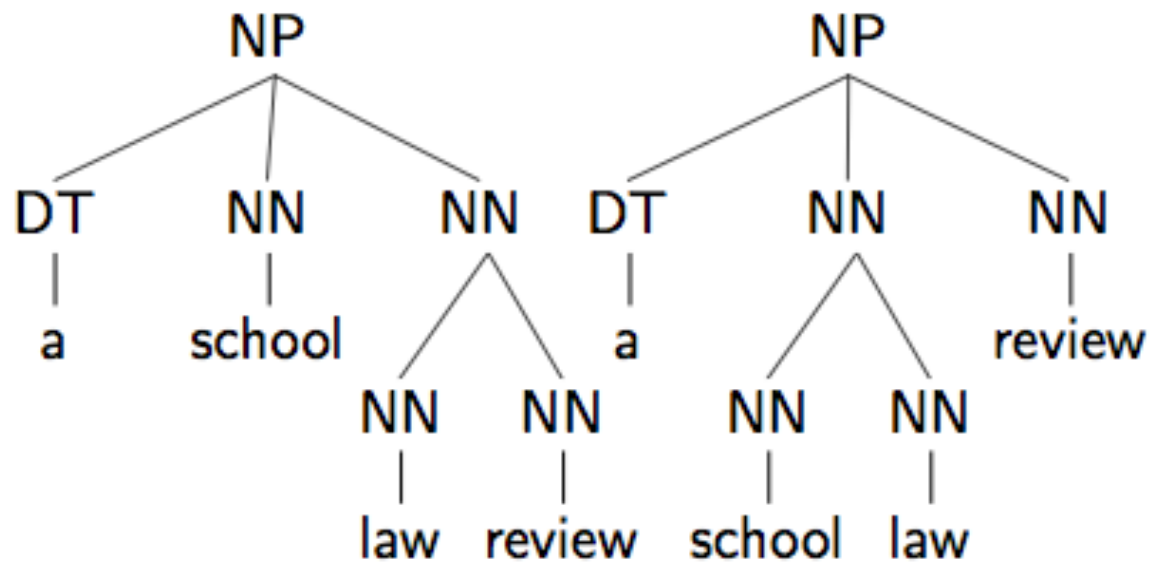
Modifiers



More modifiers



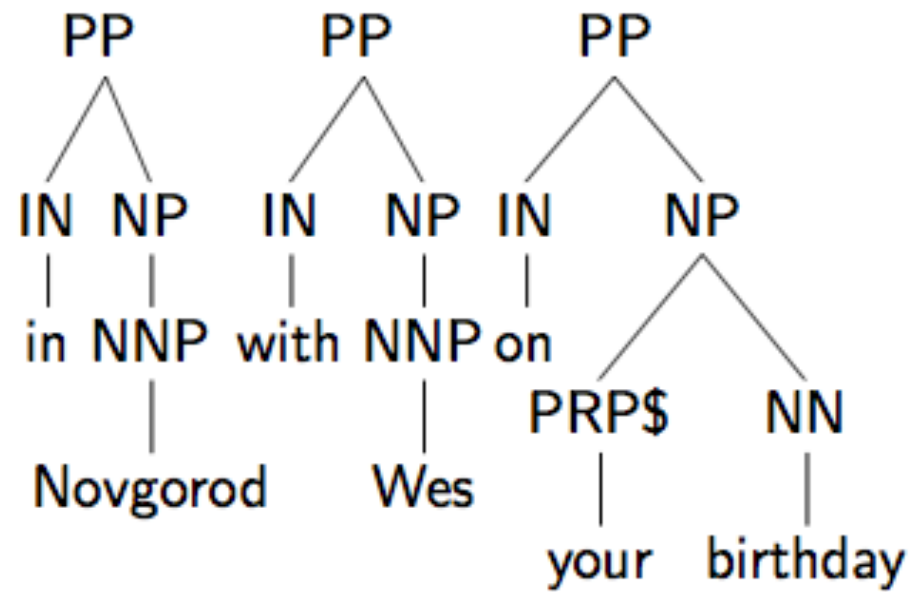
Internal structure in NP
pronunciation
meaning



Prepositional phrases

IN P

XP with X=P



Semantic classes of PPs

denotation

semantic type

| class of PPs | examples |
|---------------------|---------------------------------------|
| temporal | on Monday, in November, after lunch |
| locative | in Ithaca, on campus, under the sheet |
| path | through downtown, into Barcelona |

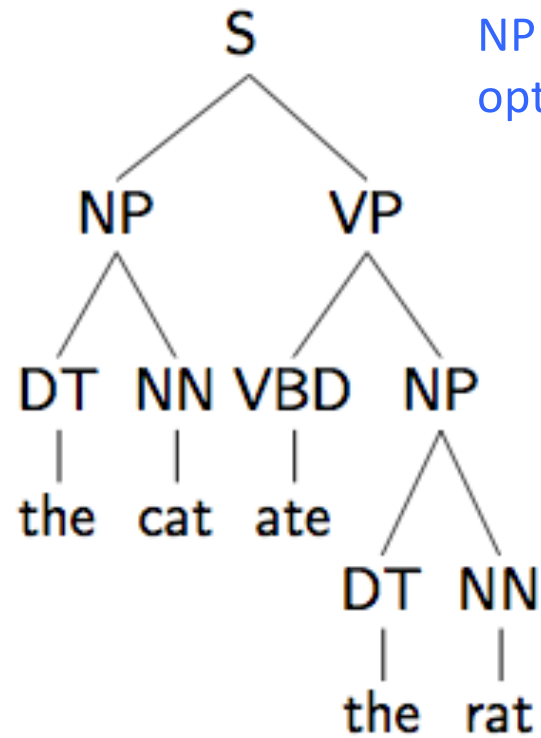
Complementation

transitive

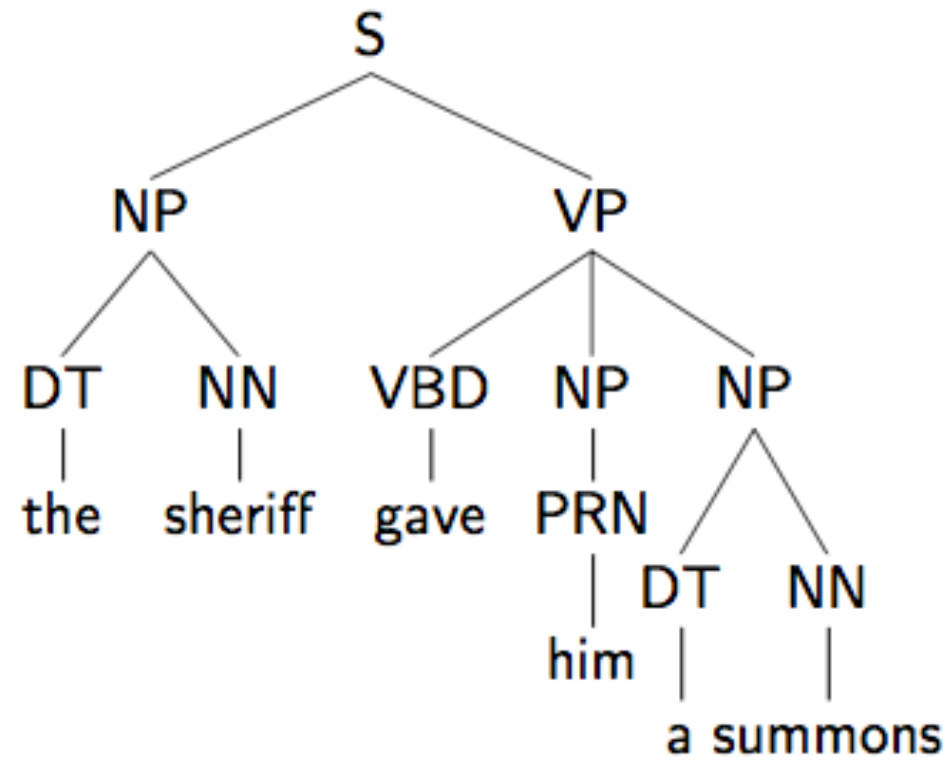
wants to combine with

NP to form VP

optionality



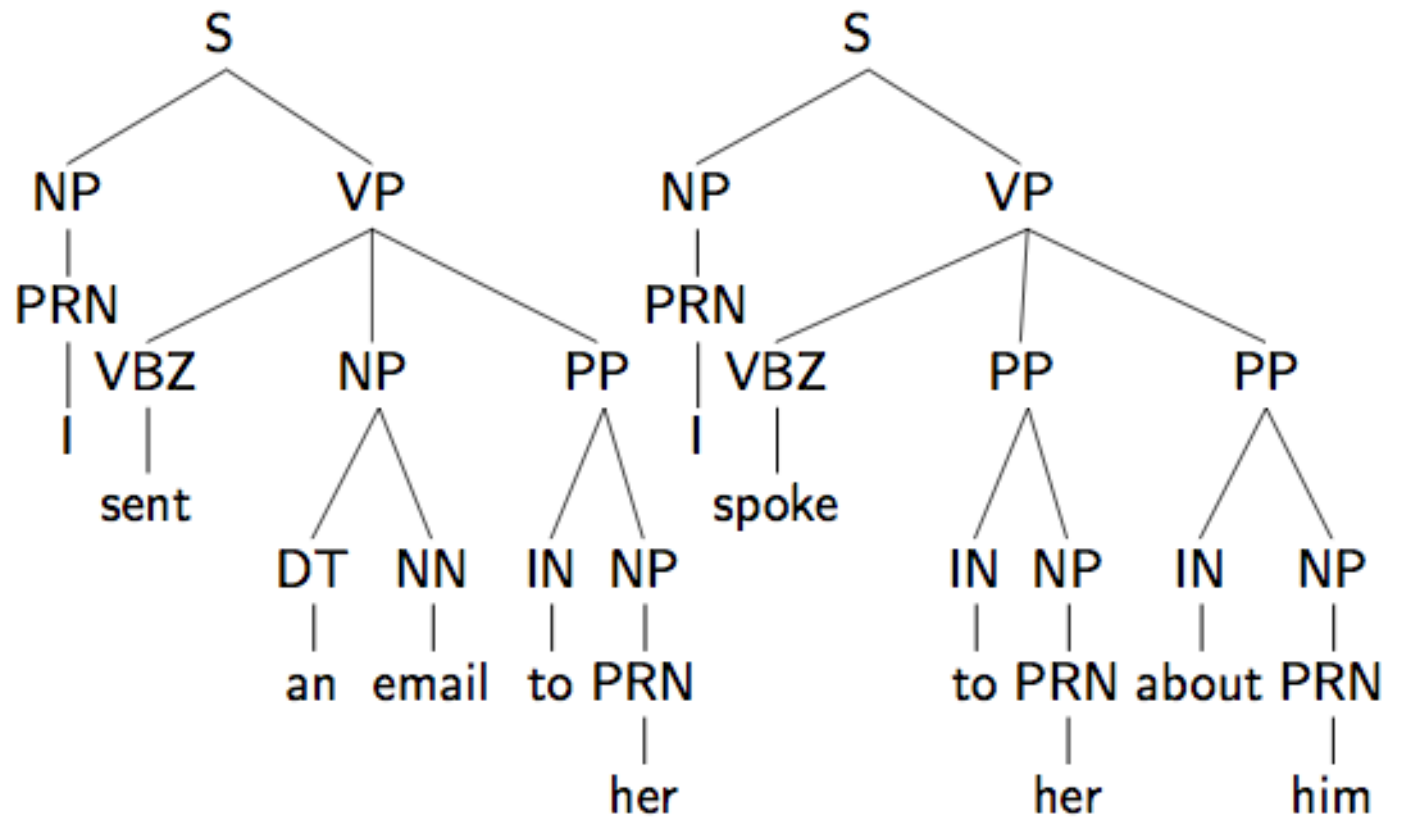
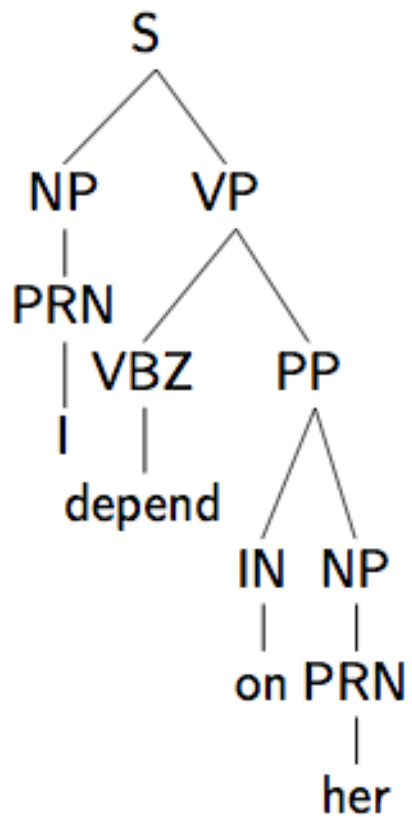
Ditransitive

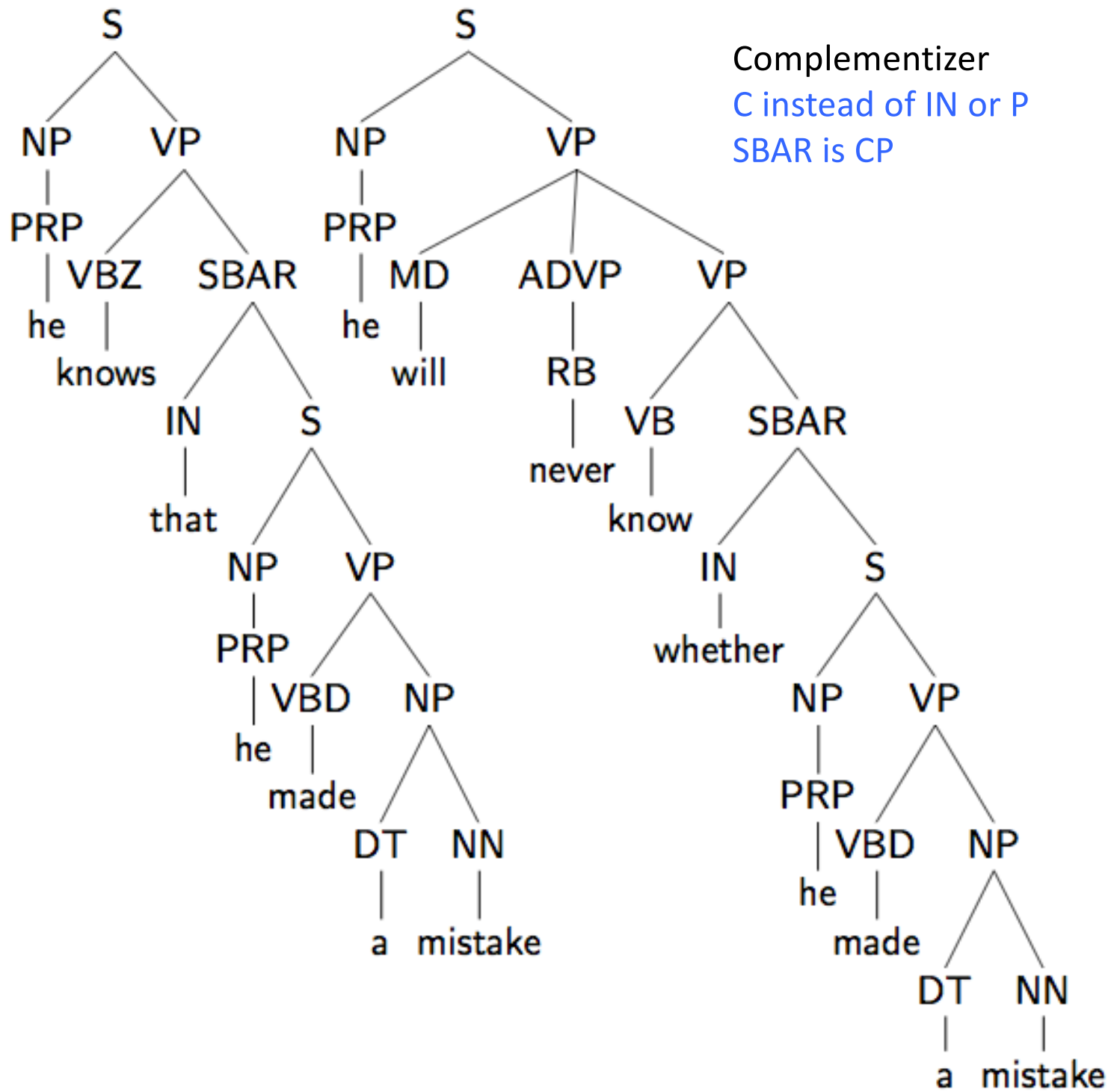


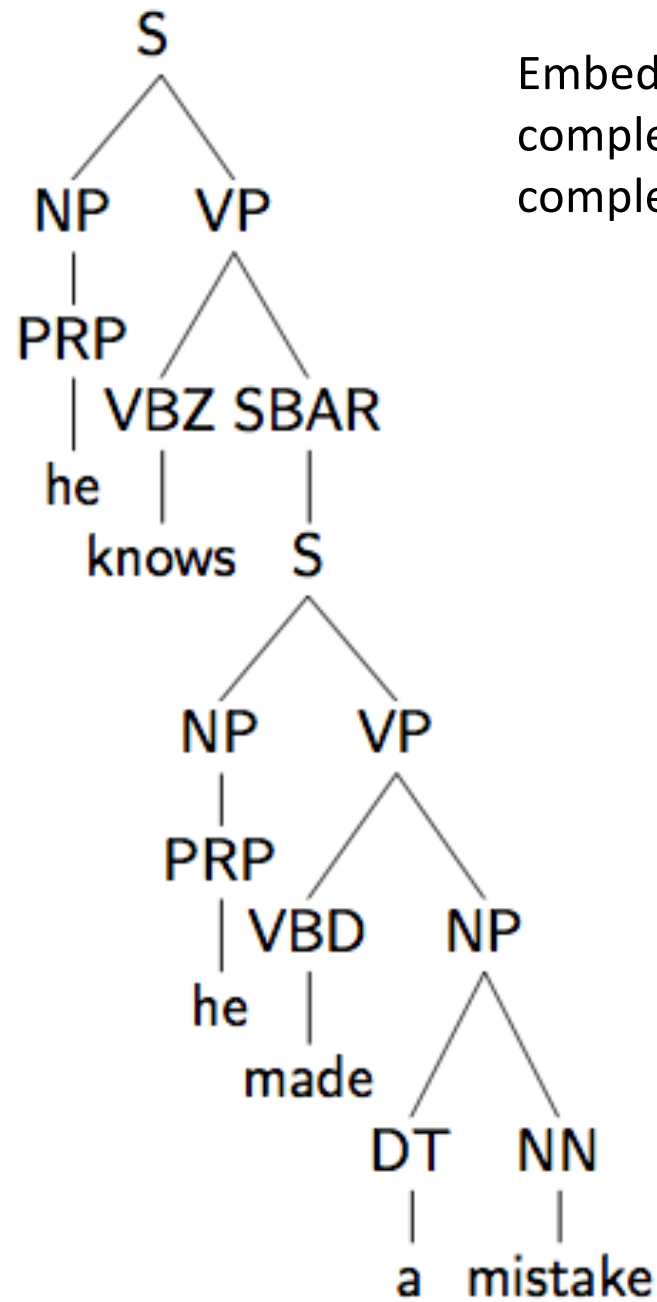
Prepositional complement

goal

theta role







Embedded sentence with no complementizer, or empty complementizer.

Complement selection
Valence

- * I depend her.
- * I ate to her about him.
- * He believed to her.
- * He spoke whether he made a mistake.

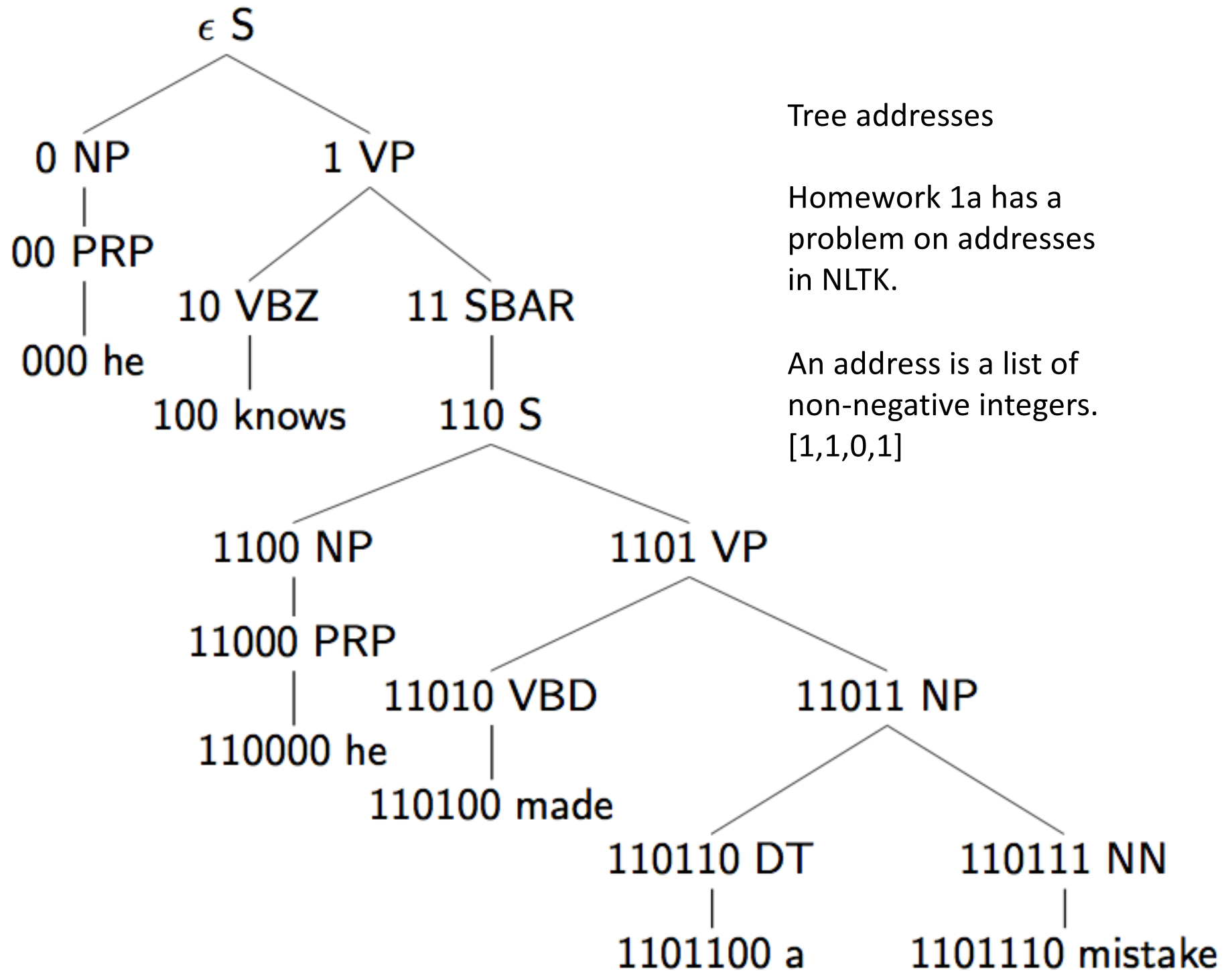
Particular P or
semantic class of PP

I depend on/*in her.

He yearned for/*to an icecream cone.

He left the paper in the trash. (Location)

*He left the paper into the trash. (Path)



Tree addresses

Homework 1a has a problem on addresses in NLTK.

An address is a list of non-negative integers. [1,1,0,1]

An address is a finite sequence of non-negative integers. The address of the root is the empty sequence ϵ . If a is a non-terminal address in the tree, then the address of the first child of a is $a0$. The address of the second child (if there is one) is $a1$, and the address of the i th child (counting from 0) is ai .

Definition A tree domain X is a subset of N^* satisfying the following conditions.

- (i) If $\alpha k \in X$ and $0 \leq i < k$ then $\alpha i \in X$.
- (ii) If $\alpha i \in X$ then $\alpha \in X$.

Definition A labeled tree t is a function such that $Dom(t)$ is a tree domain.

| x | $t(x)$ | x | $t(x)$ | x | $t(x)$ | x | $t(x)$ |
|------------|--------|------|--------|--------|--------|---------|---------|
| ϵ | S | 10 | VBZ | 11000 | PRP | 11011 | NP |
| 0 | NP | 100 | knows | 110000 | he | 110110 | DT |
| 00 | PRP | 11 | SBAR | 1101 | VP | 1101100 | a |
| 000 | he | 110 | S | 11010 | VBD | 110111 | NN |
| 1 | VP | 1100 | NP | 110100 | made | 110111 | mistake |

Formal conception of natural language syntax: a syntactic natural language (such as English) is a set of labeled trees. How big a set?

This is the dog that worried the cat that chased the rat that ate the carrot that lay in the house that Jack built.

String formal language

$ab^+ = \{ab, abb, abbb, \dots\}$

$a^n b^n$ (n greater than 0)

$= \{ab, aabb, aaabbb, \dots\}$

Tree formal language

Phonology

Semantics

