Language Technologies Institute

Carnegie Mellon University

Advanced NLP

11-711 · October 2021

Syntax and parsing 1

(Some slides adapted from Emma Strubell and J&M)

Syntax

The mailman bit my dog

- Some early AI natural language work tried to avoid using syntax
  - *(Including me in grad school, at first)*
- You *cannot* understand this sentence based solely on statistics or semantics
- You need syntax (language-specific patterns) to understand statements about weird, unlikely things
  - Also *probably* as a learning bias, for all language

Syntax

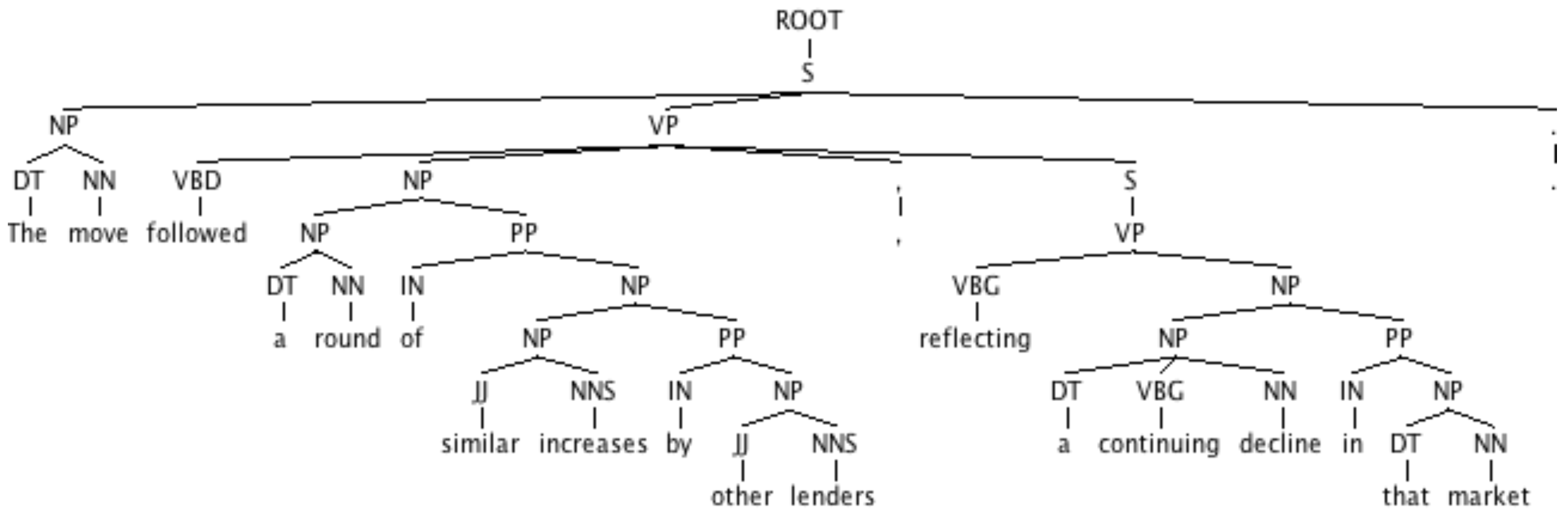
- The study of the patterns of formation of sentences and phrases from words
  - my dog **Pron N**
  - the dog **Det N**
  - the cat **Det N**
  - the large cat **Det Adj N**
  - the black cat **Det Adj N**
  - ate a sausage **V Det N**
- Compositional, recursive patterns

Syntactic parsing

- **Input:**

  The move followed a round of similar increases by other lenders, reflecting a continuing decline in that market.

- **Output:**



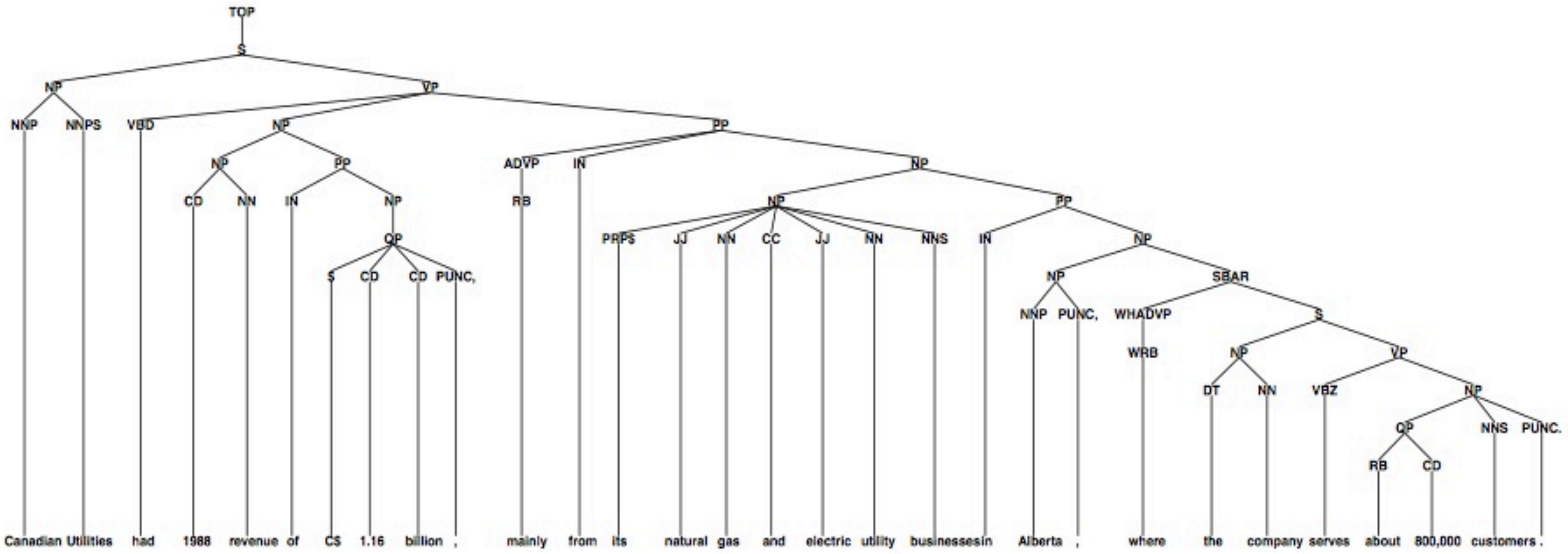
Ambiguity

I saw the woman with the telescope wrapped in paper.

- Who has the telescope?
- Who or what is wrapped in paper?
- Event of perception or assault?

Parsing as supervised ML

- Data for parsing experiments:
  - WSJ portion of the Penn Treebank = 50k sentences annotated with trees
  - Usual train/test split: 40k training, 1700 development, 2400 test



Canadian Utilities had 1988 revenue of $ 1.16 billion , mainly from its natural gas and electric utility businesses in Alberta , where the company serves about 800,000 customers .

Morphology + syntax + semantics

- **Syntax**: The study of the patterns of formation of sentences and phrases from words.
- Borders with semantics and morphology are sometimes blurred.

Afyonkarahisarlılaştırabildiklerimizdenmişsinizcesinee

as if you are one of the people that we thought to be originating from Afyonkarahisar

Parts of Speech

- 8 (ish) traditional parts of speech:
  - Noun, verb, adjective, preposition, adverb, article, interjection, pronoun, conjunction, etc
    - Called: parts-of-speech, lexical categories, word classes, morphological classes, lexical tags...
    - Lots of debate within linguistics about the number, nature, and universality of these
      - We'll completely ignore this debate.

POS examples

- N noun *chair, bandwidth, pacing*
- V verb *study, debate, munch*
- ADJ adjective *purple, tall, ridiculous*
- ADV adverb *unfortunately, slowly*
- P preposition *of, by, to*
- PRO pronoun *I, me, mine*
- DET determiner *the, a, that, those*

POS Tagging

- The process of assigning a part-of-speech or lexical class marker to each word in a collection.

WORD	tag
the	**DET**
koala	**N**
put	**V**
the	**DET**
keys	**N**
on	**P**
the	**DET**
table	**N**

Why is POS Tagging Useful?

- First step of a vast number of practical tasks
- Speech synthesis
  - How to pronounce "lead"?
  - INsult inSULT
  - OBject objECT
  - OVERflow overFLOW
  - DIScount disCOUNT
  - CONtent conTENT
- Parsing
  - Need to know if a word is an N or V before you can parse
- Information extraction
  - Finding names, relations, etc.
- Machine Translation

Open and Closed Classes

- Closed class: a small fixed membership
  - Prepositions: of, in, by, …
  - Auxiliaries: may, can, will had, been, …
  - Pronouns: I, you, she, mine, his, them, …
  - Usually function words (short common words which play a role in grammar)
- Open class: new ones can be created all the time
  - English has 4: Nouns, Verbs, Adjectives, Adverbs
  - Many languages have these 4, but (maybe) not all!

Open Class Words

- Nouns
  - Proper nouns (Wilmerding, Graham, Eli Manning)
    - English capitalizes these.
  - Common nouns (the rest).
  - Count nouns and mass nouns
    - Count: have plurals, get counted: goat/goats, one goat, two goats
    - Mass: don’t get counted (snow, salt, communism) (*two snows)
- Adverbs: tend to modify things
  - Unfortunately, John walked home extremely slowly yesterday
  - Directional/locative adverbs (here, home, downhill)
  - Degree adverbs (extremely, very, somewhat)
  - Manner adverbs (slowly, slinkily, delicately)
- Verbs
  - In English, have morphological changes (eat/eats/eaten)

Closed Class Words

Examples:

- prepositions: *on, under, over, …*
- particles: *up, down, on, off, …*
- determiners: *a, an, the, …*
- pronouns: *she, who, I, ..*
- conjunctions: *and, but, or, …*
- auxiliary verbs: *can, may should, …*
- numerals: *one, two, three, third, …*

Prepositions from CELEX

of	540,085	through	14,964	worth	1,563	pace	12
in	331,235	after	13,670	toward	1,390	nigh	9
for	142,421	between	13,275	plus	750	re	4
to	125,691	under	9,525	till	686	mid	3
with	124,965	per	6,515	amongst	525	o'er	2
on	109,129	among	5,090	via	351	but	0
at	100,169	within	5,030	amid	222	ere	0
by	77,794	towards	4,700	underneath	164	less	0
from	74,843	above	3,056	versus	113	midst	0
about	38,428	near	2,026	amidst	67	o'	0
than	20,210	off	1,695	sans	20	thru	0
over	18,071	past	1,575	circa	14	vice	0

POS Tagging
Choosing a Tagset

- There are so many parts of speech, potential distinctions we can draw
- To do POS tagging, we need to choose a standard set of tags to work with
- Could pick very coarse tagsets
  - N, V, Adj, Adv.
- More commonly used set is finer grained, the "Penn TreeBank tagset", 45 tags
  - PRP$, WRB, WP$, VBG
- Even more fine-grained tagsets exist

Penn TreeBank POS Tagset

Tag	Description	Example	Tag	Description	Example
CC	coordin. conjunction	*and, but, or*	SYM	symbol	+,%, &
CD	cardinal number	*one, two, three*	TO	"to"	*to*
DT	determiner	*a, the*	UH	interjection	*ah, oops*
EX	existential 'there'	*there*	VB	verb, base form	*eat*
FW	foreign word	*mea culpa*	VBD	verb, past tense	*ate*
IN	preposition/sub-conj	*of, in, by*	VBG	verb, gerund	*eating*
JJ	adjective	*yellow*	VBN	verb, past participle	*eaten*
JJR	adj., comparative	*bigger*	VBP	verb, non-3sg pres	*eat*
JJS	adj., superlative	*wildest*	VBZ	verb, 3sg pres	*eats*
LS	list item marker	*1, 2, One*	WDT	wh-determiner	*which, that*
MD	modal	*can, should*	WP	wh-pronoun	*what, who*
NN	noun, sing. or mass	*llama*	WP$	possessive wh-	*whose*
NNS	noun, plural	*llamas*	WRB	wh-adverb	*how, where*
NNP	proper noun, singular	*IBM*	$	dollar sign	$
NNPS	proper noun, plural	*Carolinas*	#	pound sign	#
PDT	predeterminer	*all, both*	"	left quote	' or "
POS	possessive ending	*'s*	"	right quote	' or "
PRP	personal pronoun	*I, you, he*	(	left parenthesis	[, (, {, <
PRP$	possessive pronoun	*your, one's*	)	right parenthesis	],), }, >
RB	adverb	*quickly, never*	,	comma	,
RBR	adverb, comparative	*faster*	.	sentence-final punc	. ! ?
RBS	adverb, superlative	*fastest*	:	mid-sentence punc	: ; ... – -
RP	particle	*up, off*			

Using the Penn Tagset

- The/DT grand/JJ jury/NN commmented/VBD on/IN a/DT number/NN of/IN other/JJ topics/NNS ./.
- Prepositions and subordinating conjunctions marked IN ("although/IN I/PRP..")
- Except the preposition/complementizer "to" is just marked "TO".

POS Tagging

- Words often have more than one POS: *back*
  - The *back* door = JJ
  - On my *back* = NN
  - Win the voters *back* = RB
  - Promised to *back* the bill = VB
- The POS tagging problem is to determine the POS tag for a particular instance of a word.

These examples from Dekang Lin

How Hard is POS Tagging? Measuring Ambiguity

		87-tag Original Brown		45-tag Treebank Brown	
Unambiguous (1 tag)		**44,019**		**38,857**	
Ambiguous (2–7 tags)		**5,490**		**8844**	
Details:	2 tags	4,967		6,731	
	3 tags	411		1621	
	4 tags	91		357	
	5 tags	17		90	
	6 tags	2	(*well*, *beat*)	32	
	7 tags	2	(*still*, *down*)	6	(*well*, *set*, *round*, *open*, *fit*, *down*)
	8 tags			4	(*'s*, *half*, *back*, *a*)
	9 tags			3	(*that*, *more*, *in*)

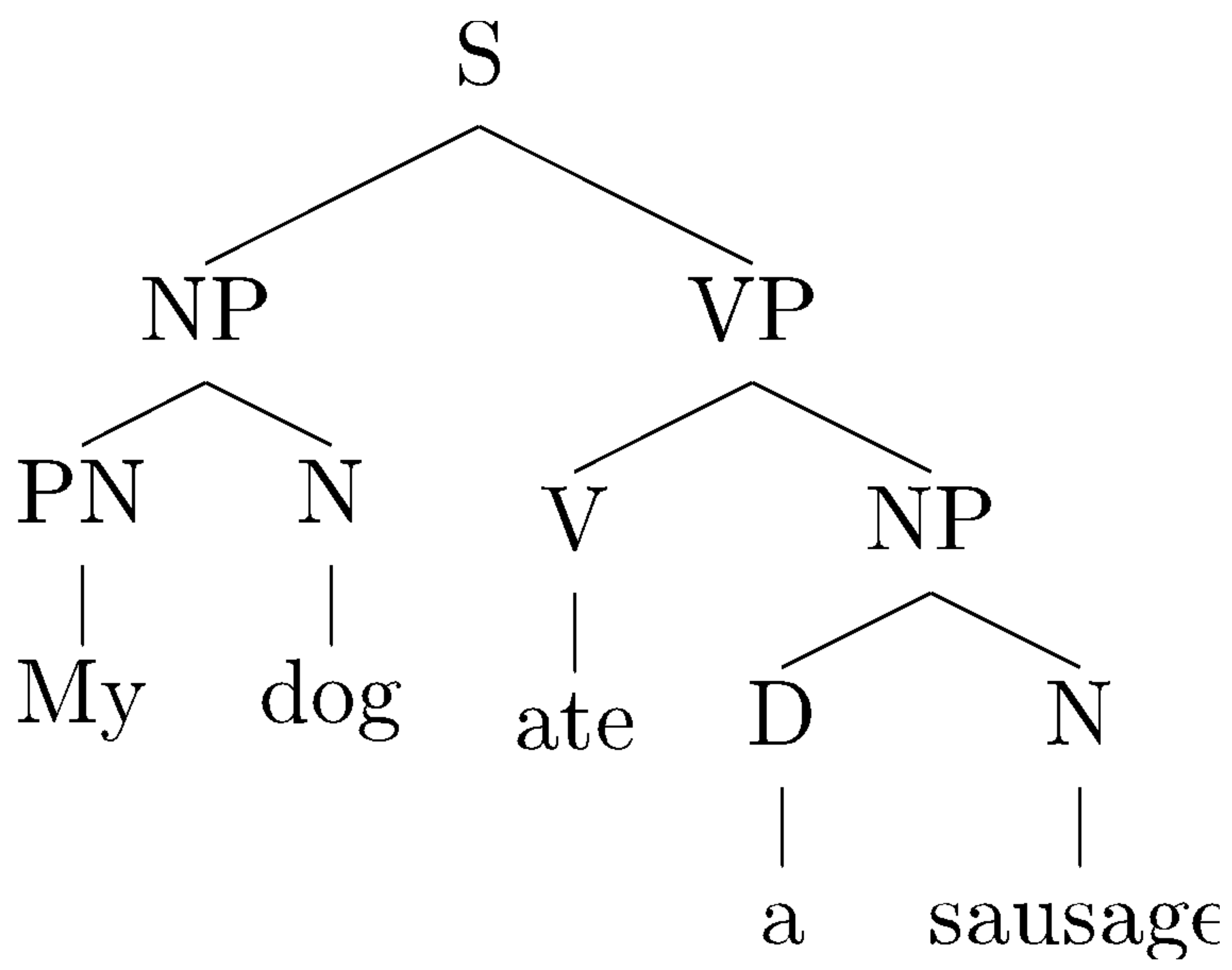
Three Methods for POS Tagging

1. Rule-based tagging
   - (ENGTWOL)
2. Stochastic/Probabilistic sequence models
   - HMM (Hidden Markov Model) tagging
   - MEMMs (Maximum Entropy Markov Models)
3. Neural

   Just use BERT

Parsing

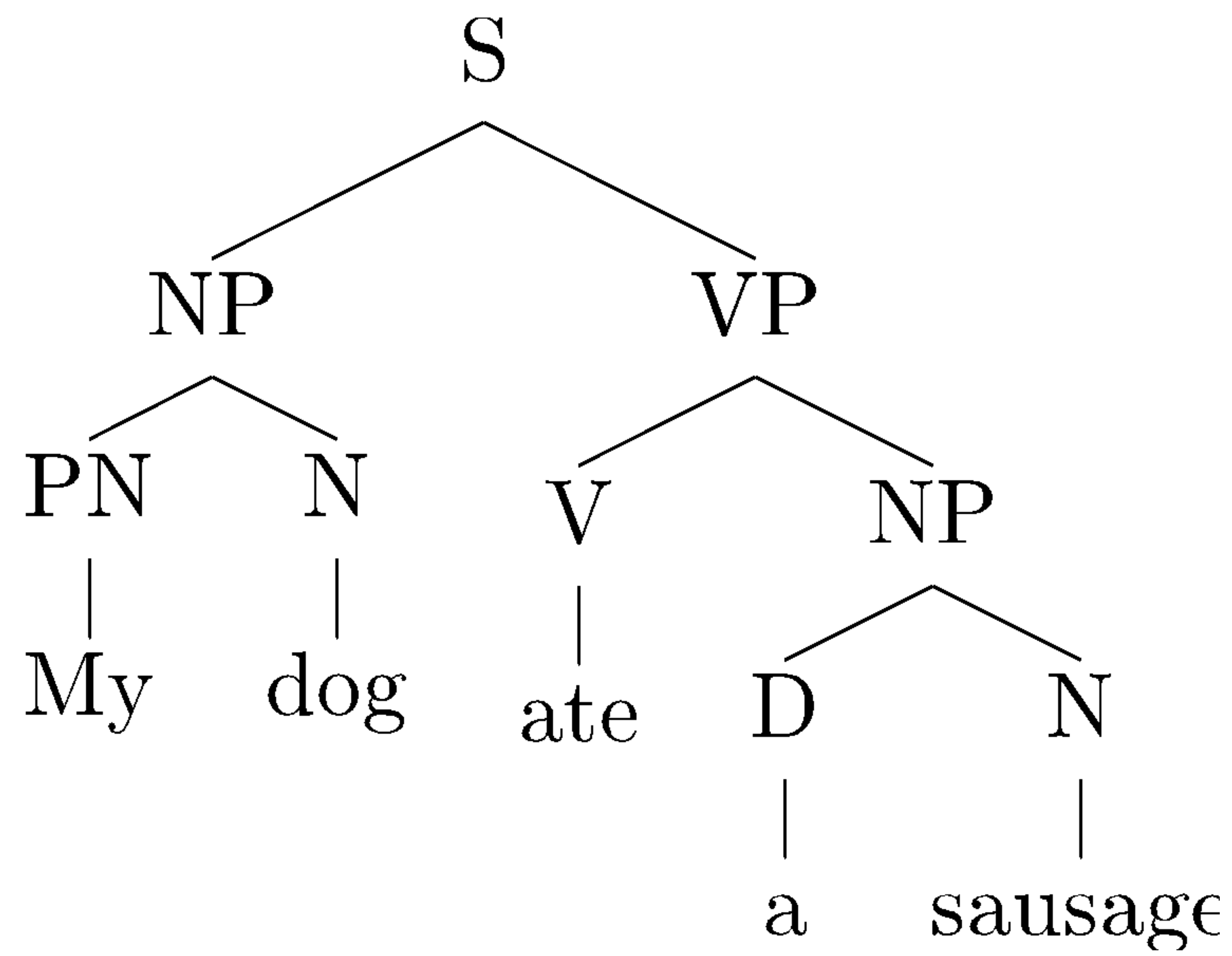
- The process of predicting **syntactic representations**
- Different types of syntactic representations are possible, for example:



constituency (aka phrase-structure) tree

Constituency trees

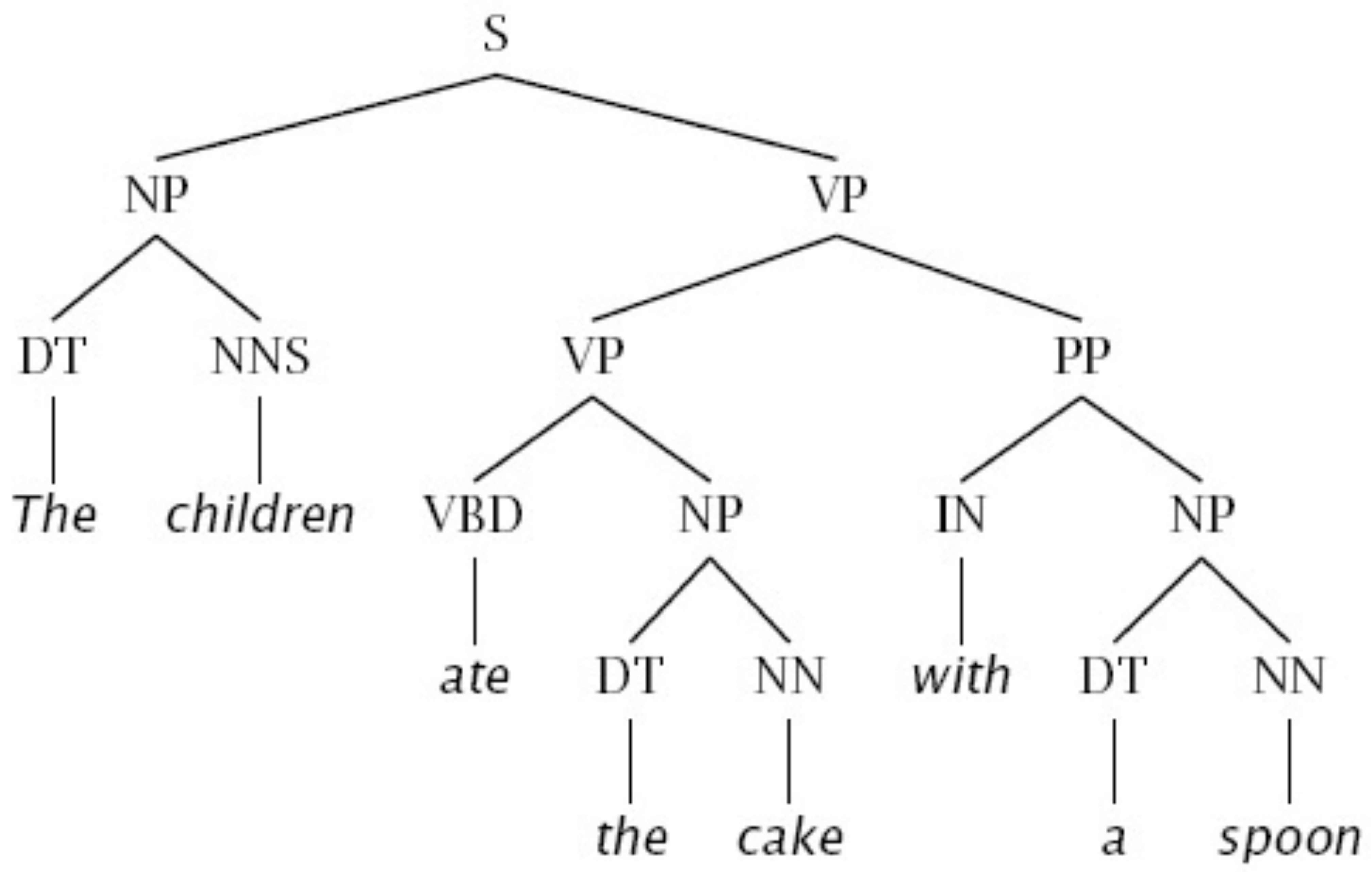
- Internal nodes correspond to **phrases**.
  - **S**: a sentence
  - **NP** (noun phrase): My dog, a sandwich, lakes, …
  - **VP** (verb phrase): ate a sausage, barked, …
  - **PP** (prepositional phrases): with a friend, in a car, …



- Nodes immediately above words are **part-of-speech** tags (or **preterminals**).
  - **PN**: pronoun
  - **D**: determiner
  - **V**: verb
  - **N**: noun
  - **P**: preposition

Constituency tests

- How do we know what nodes go in the tree?
- Classic constituency tests:
  - Replacement
  - Substitution by *proform*
  - Movement: Clefting, preposing, passive
  - Modification
  - Coordination / conjunction
  - Ellipsis / deletion



Conflicting tests

- Constituency is not always clear.
  - Coordination:

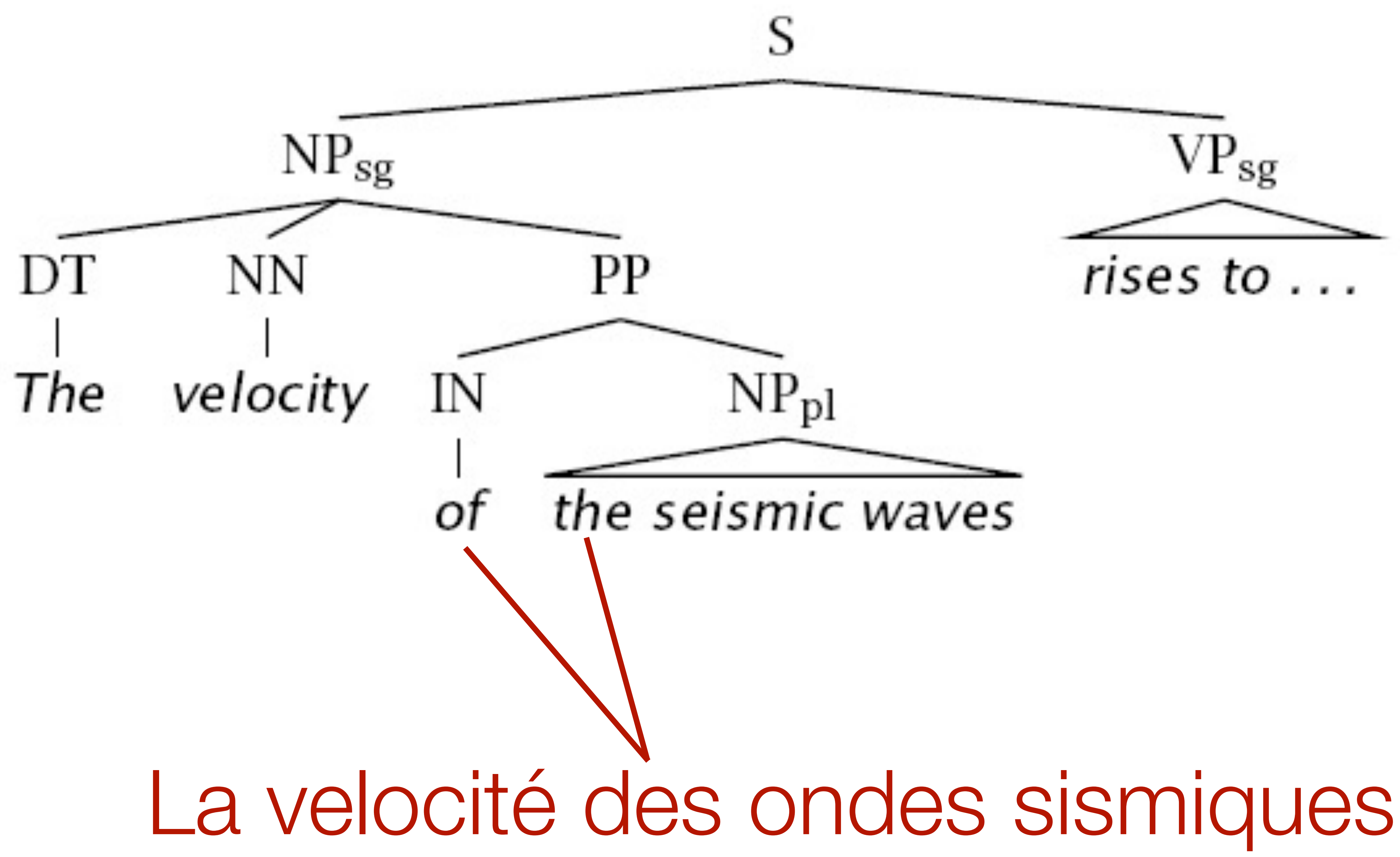
    He went to and came from the store.

  - Phonological reduction:

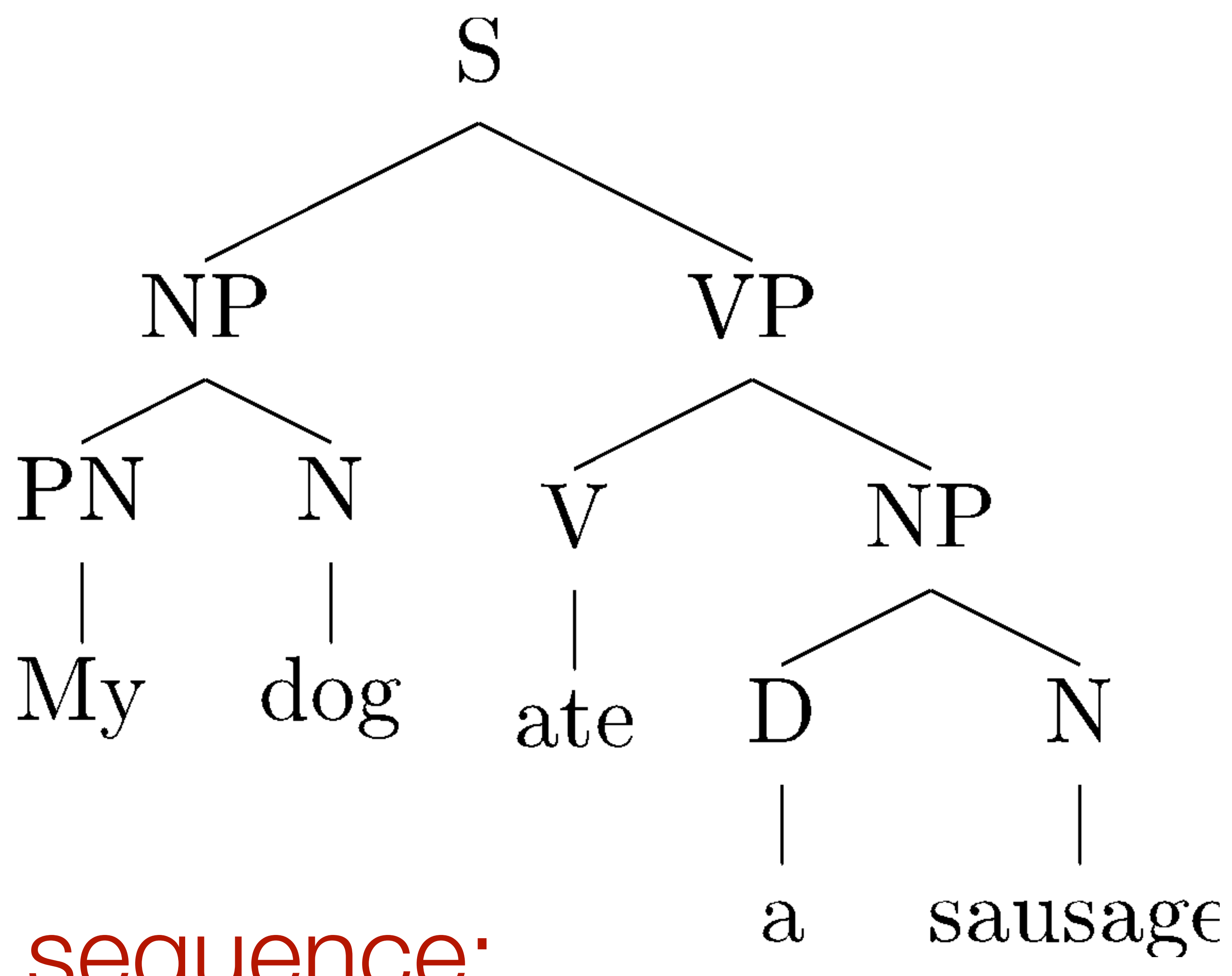
    I will go → I'll go

    I want to go → I wanna go

    a le centre → au centre



Bracketing notation

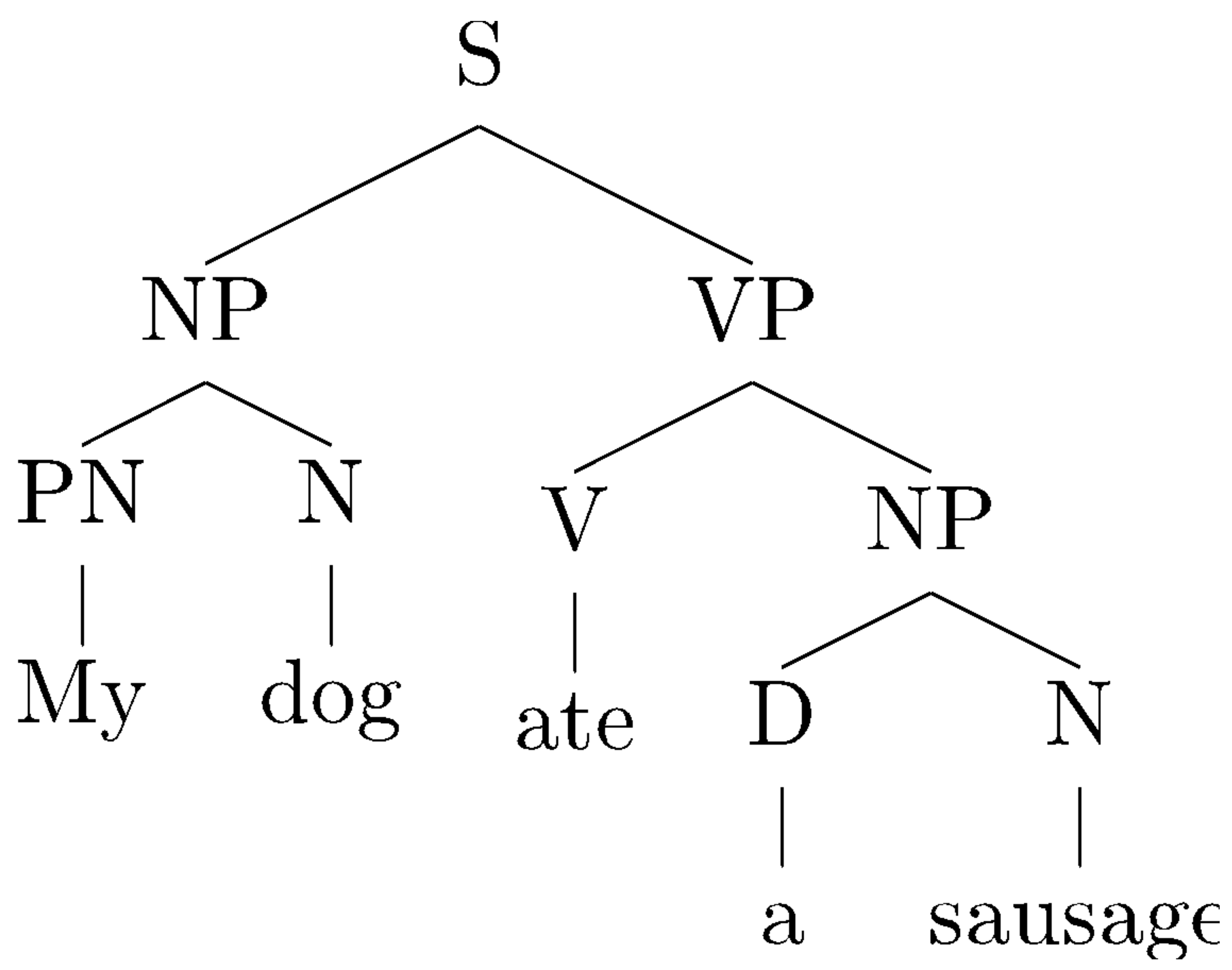


- Often convenient to represent a tree as a bracketed sequence:

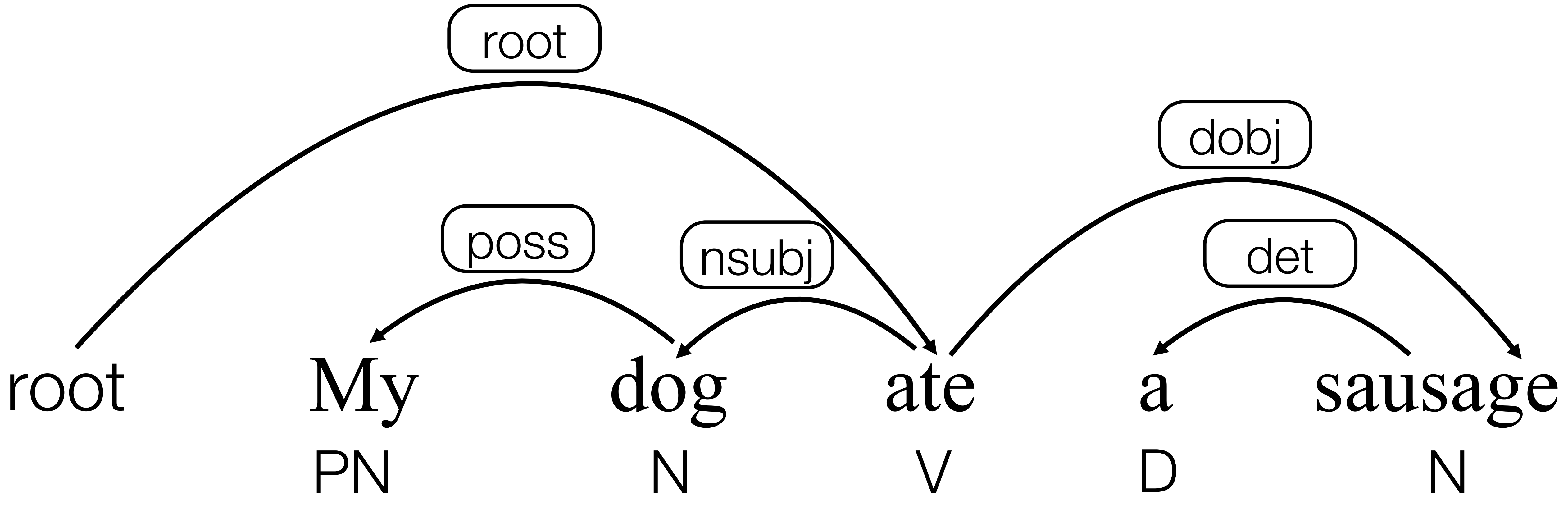
```
(S
   (NP (PN My) (N dog) )
   (VP (V ate)
      (NP (D a) (N sausage) )
   )
)
```

Parsing

- The process of predicting **syntactic representations**
- Different types of syntactic representations are possible, for example:

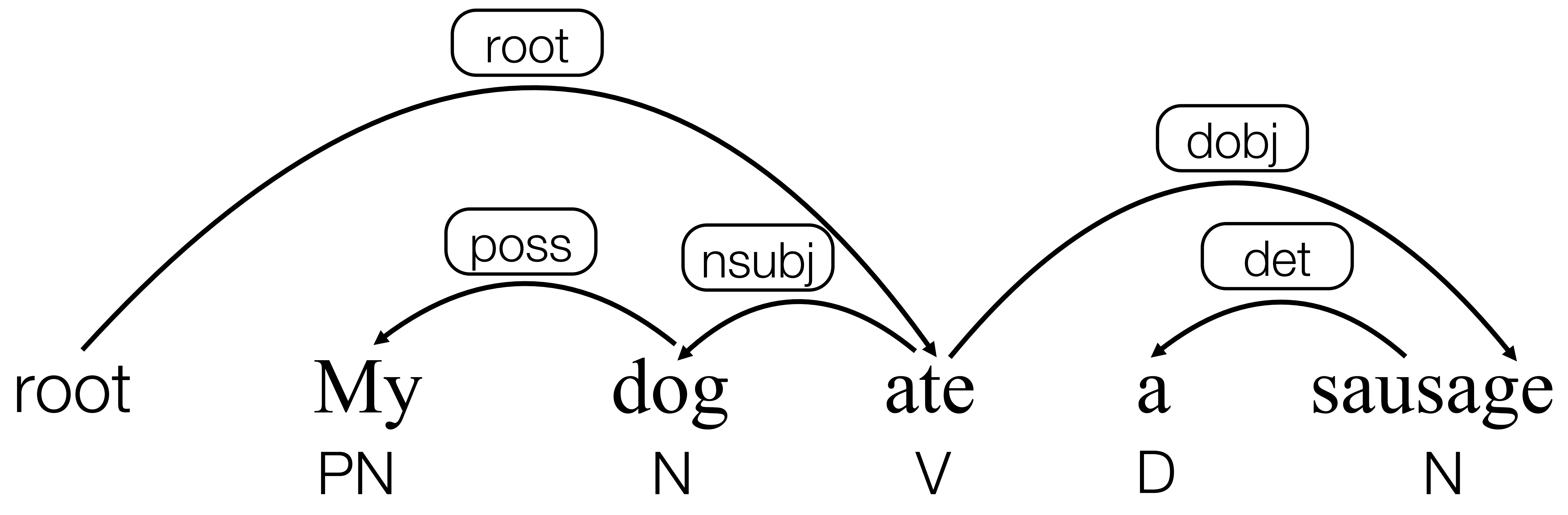


constituency (aka phrase-structure) tree



dependency tree

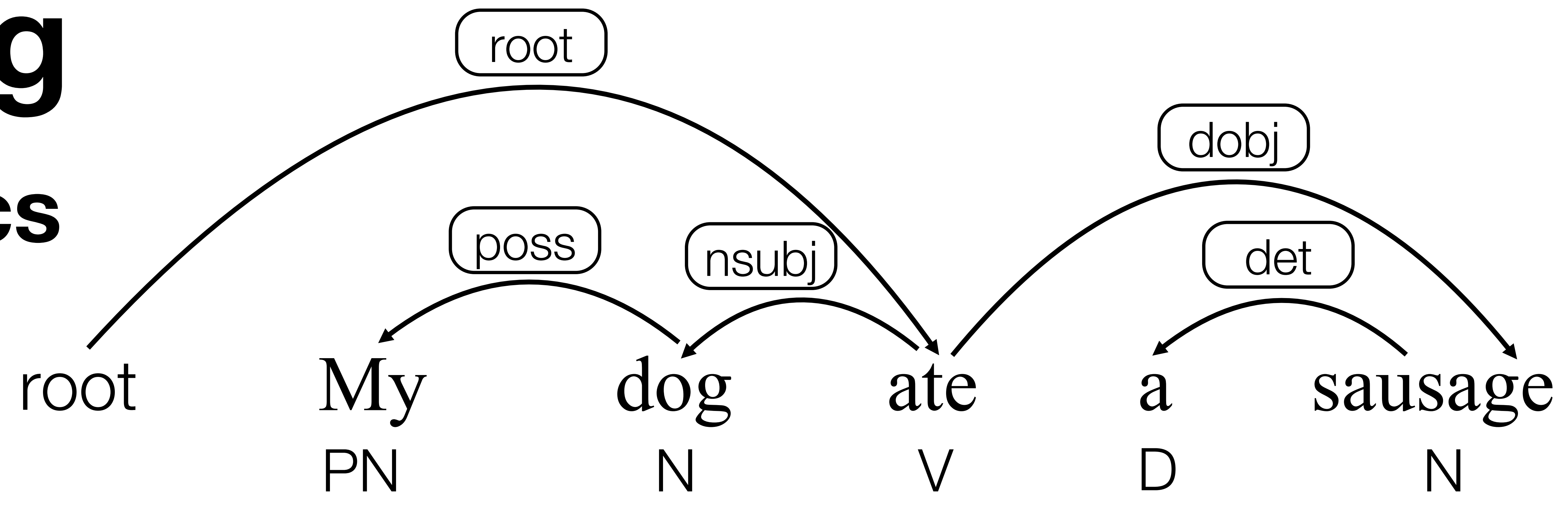
Dependency trees



- Nodes are words (along with part-of-speech tags)
- Directed arcs encode syntactic dependencies between words
- Labels are types of relations between words
  - **poss**: possessive
  - **dobj**: direct object
  - **nsubj**: (noun) subject
  - **det**: determiner

Dependency parsing

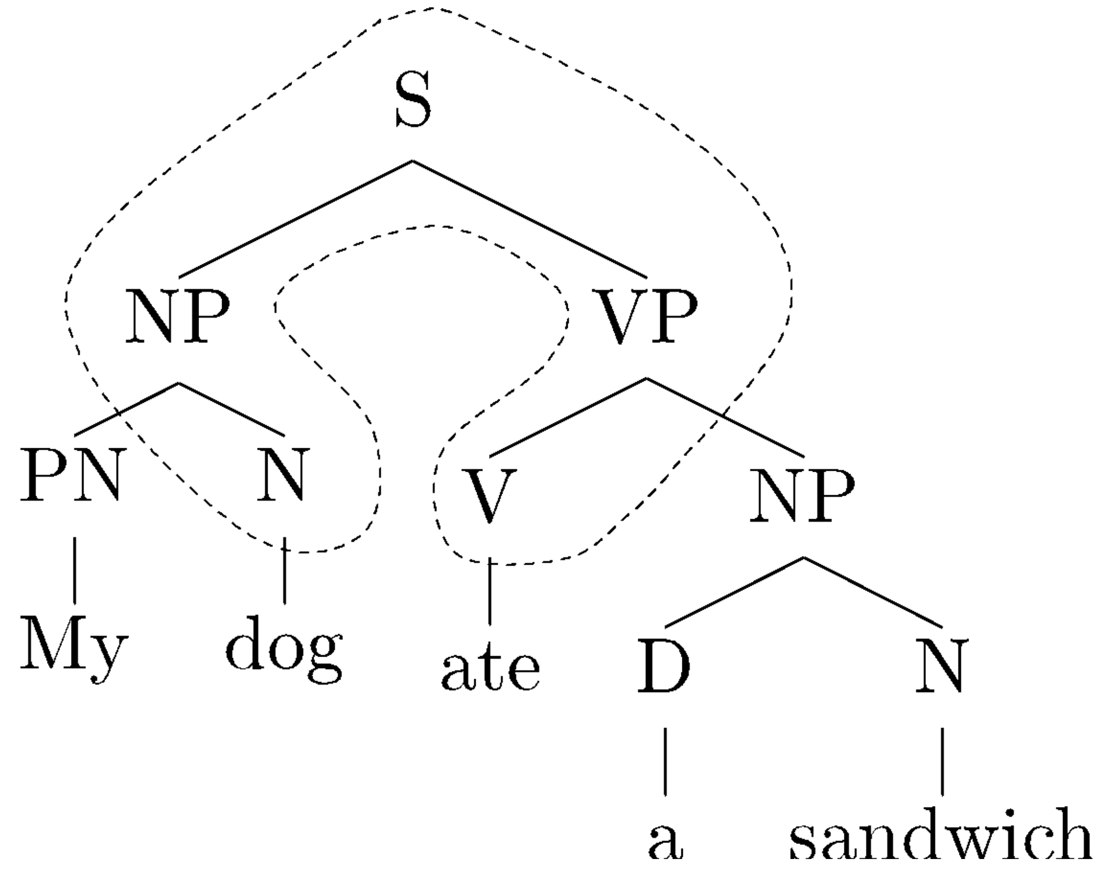
Recovering shallow semantics



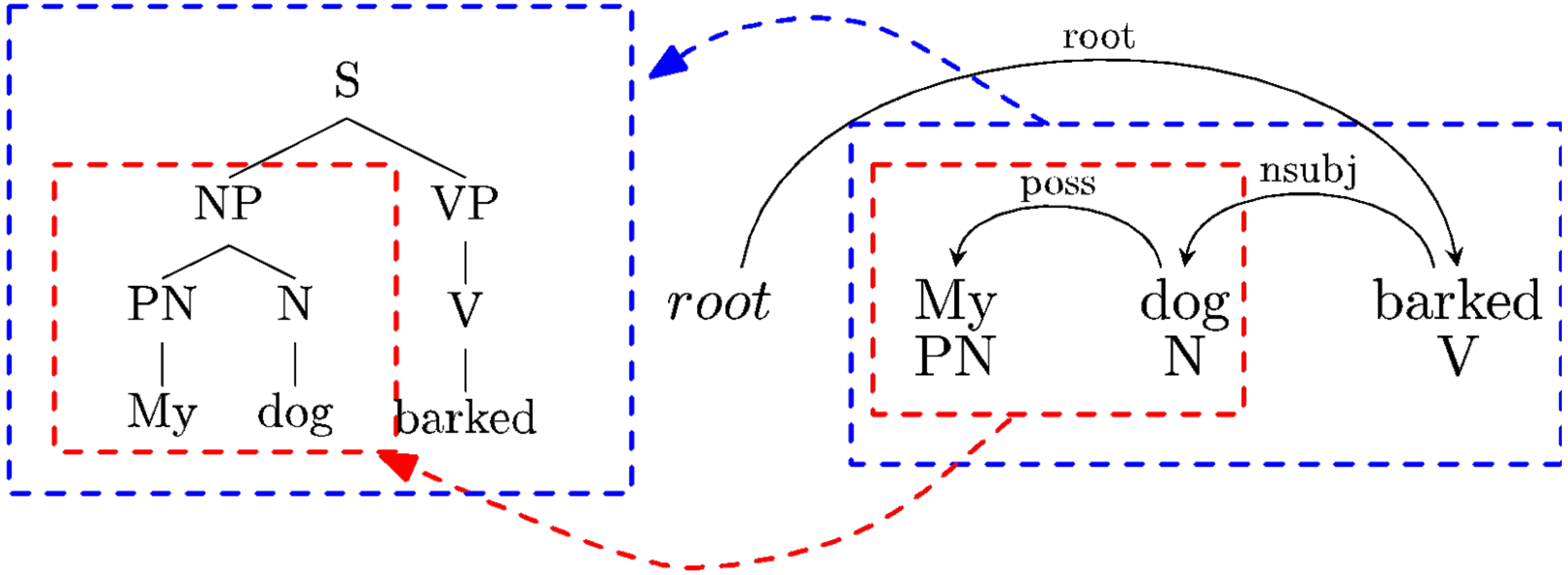
- Some semantic information can be (approximately) derived from syntactic information
  - Subjects (nsubj) are (often) **agents**: *initiators / doers of an action*
  - Direct objects (dobj) are (often) **patients**: *affected entities*
- Even for agents and patients, consider:
  - Mary is baking a cake in the oven
  - A cake is baking in the oven
- In general, it is not trivial even for the most shallow forms of semantics
  - e.g. prepositions: *in* can encode direction, position, temporal information, …

Constituency and dependency representations

- Constituency trees can (potentially) be converted to dependency trees.



- Dependency trees can (potentially) be converted to constituency trees.



Context-free grammars (CFGs)

- **Context-free grammars (CFGs)**: a formalism for parsing.

Grammar (CFG)		Lexicon
ROOT → S	NP → NP PP	NN → interest
S → NP VP	VP → VBP NP	NNS → raises
NP → DT NN	VP → VBP NP PP	VBP → interest
NP → NN NNS	PP → IN NP	VBP → raises
		…

- Other (non-CF) grammar formalisms: LFG, HPSG, TAG, CCG, …

Context-free grammars (CFGs)

Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)

S

Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

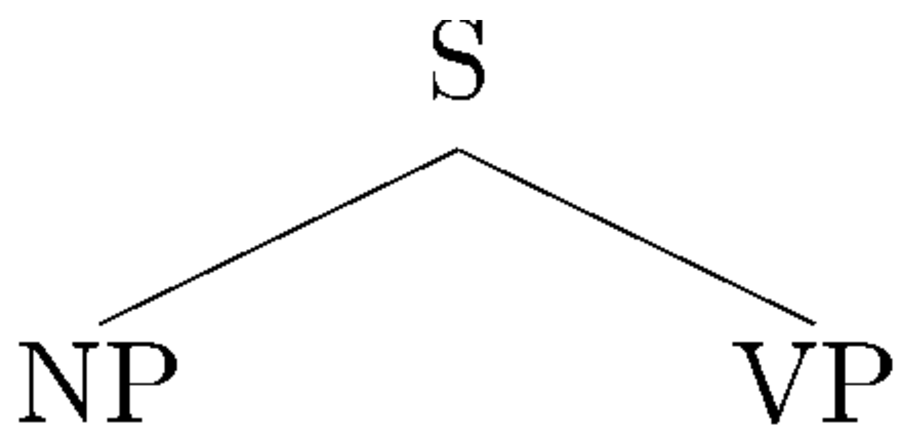
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

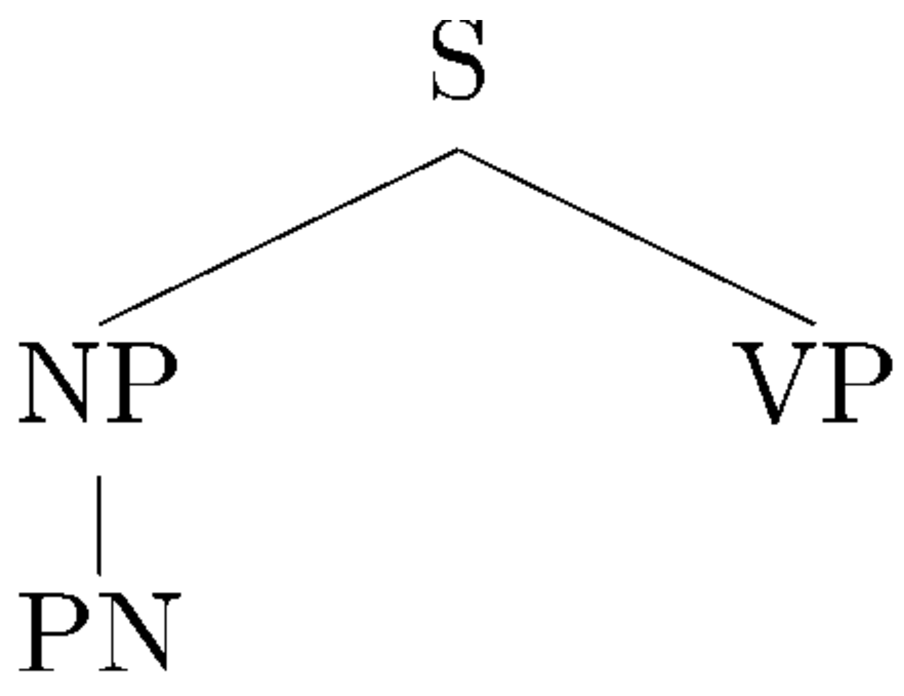
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

$S \rightarrow NP\ VP$

$VP \rightarrow V$

$VP \rightarrow V\ NP$

$VP \rightarrow VP\ PP$

$NP \rightarrow NP\ PP$

$NP \rightarrow D\ N$

$NP \rightarrow PN$

$PP \rightarrow P\ NP$

Lexicon

N → girl

N → telescope

N → sandwich

PN → I

V → saw

V → ate

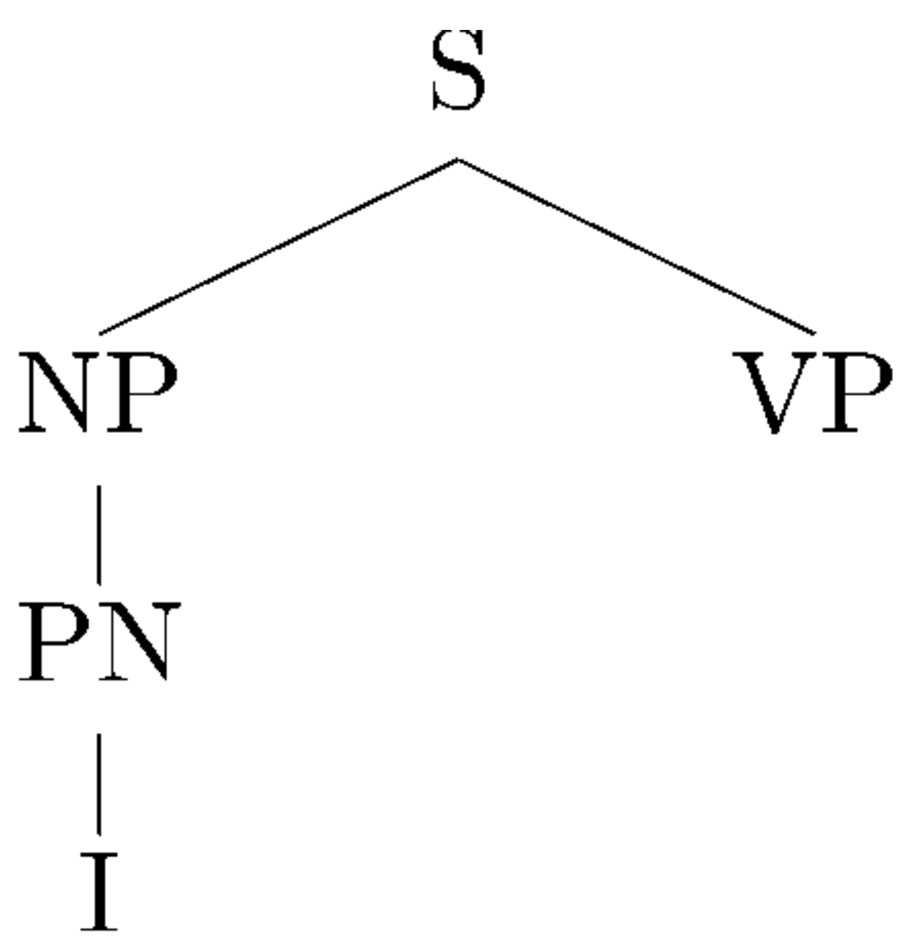
P → with

P → in

D → a

D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

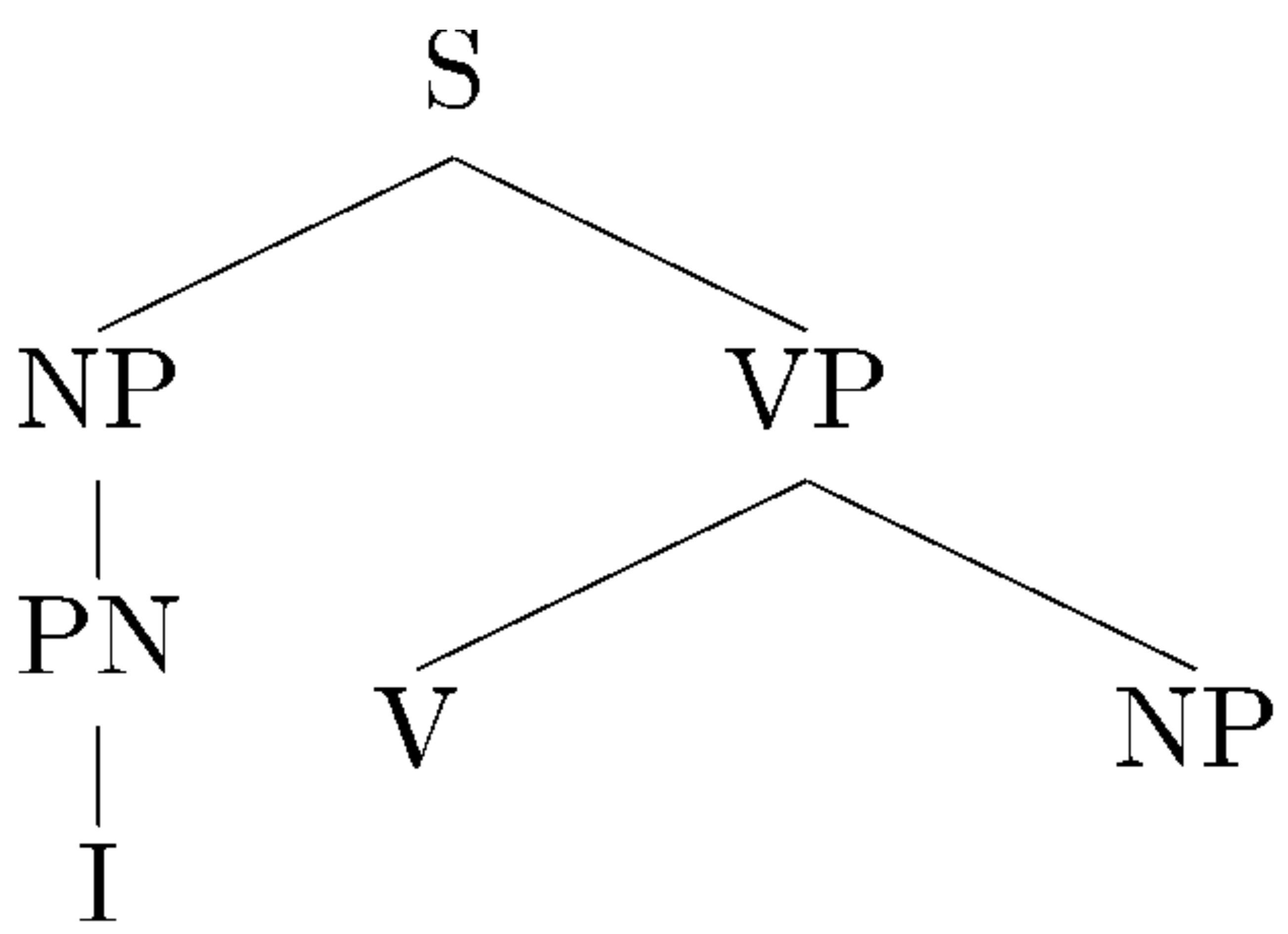
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

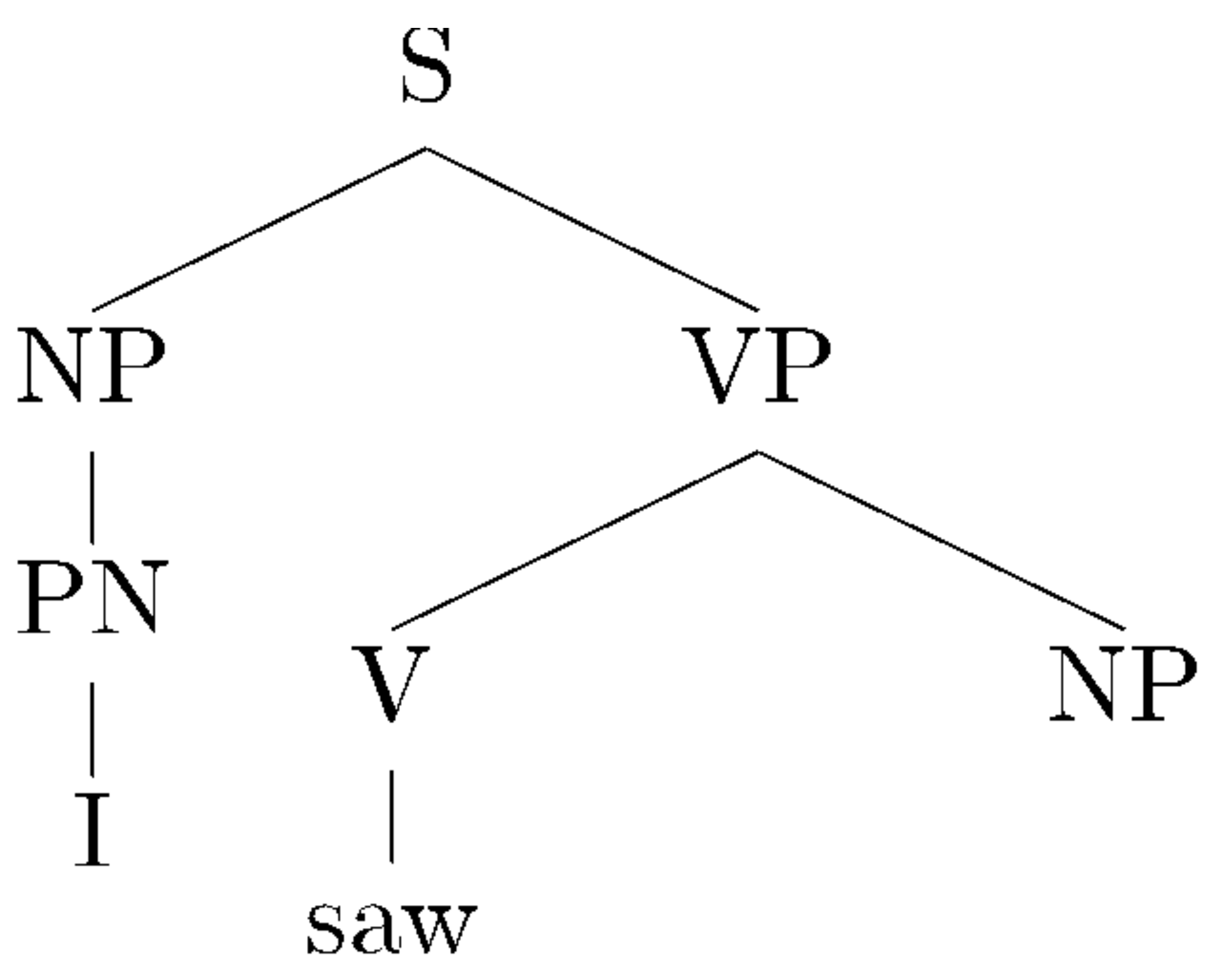
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

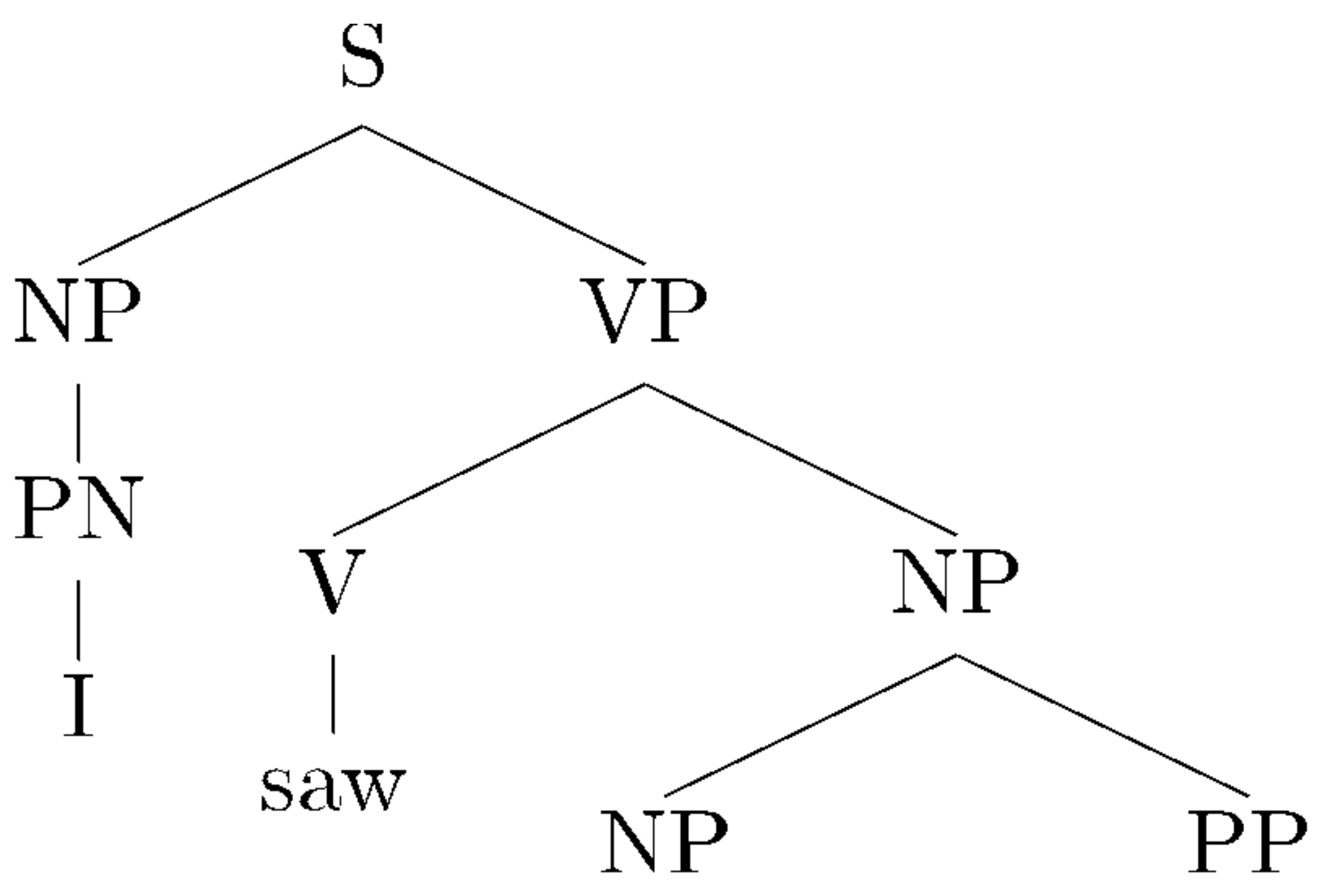
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V

VP → V NP

VP → VP PP

NP → NP PP

NP → D N

NP → PN

PP → P NP

Lexicon

N → girl

N → telescope

N → sandwich

PN → I

V → saw

V → ate

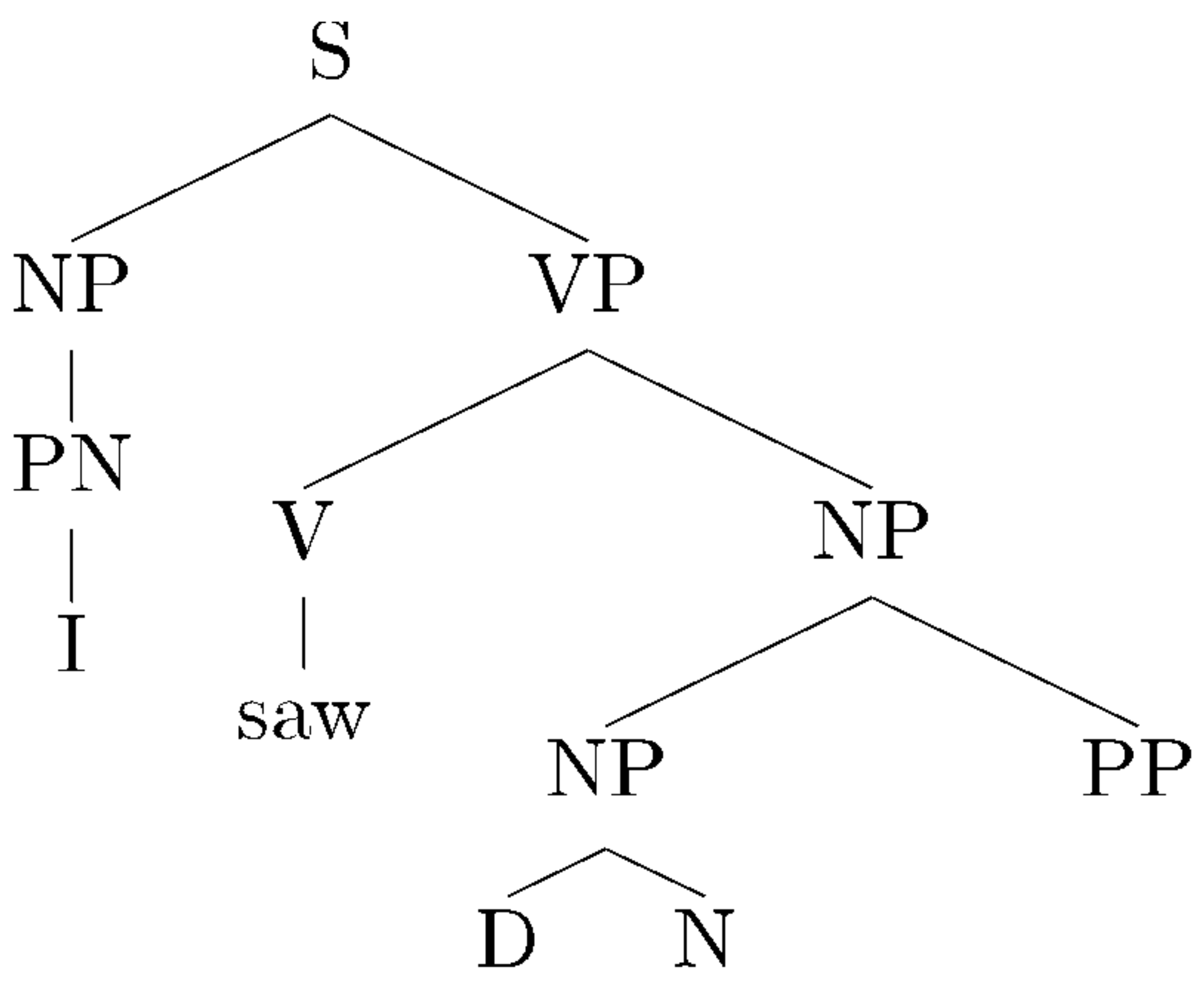
P → with

P → in

D → a

D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

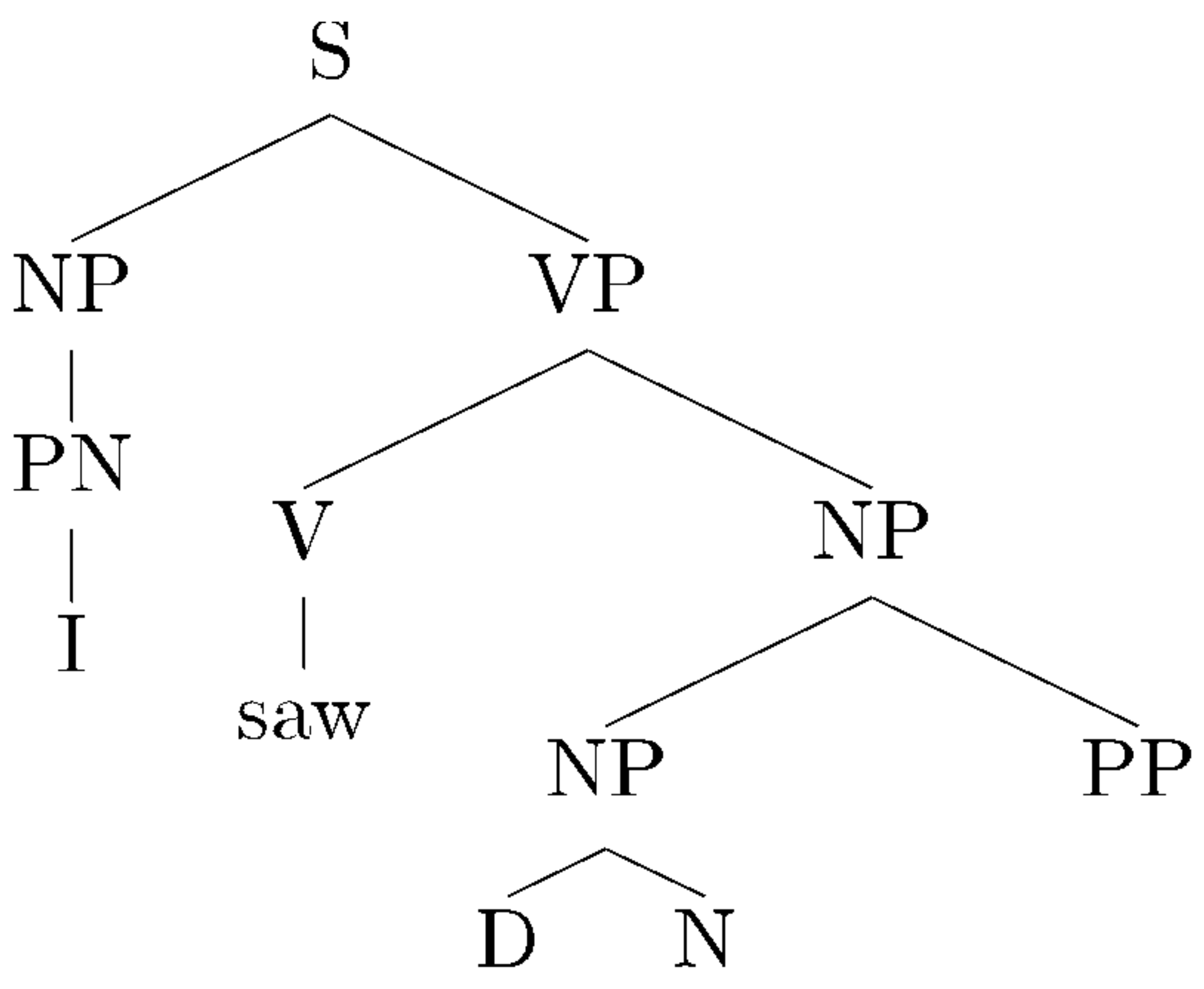
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

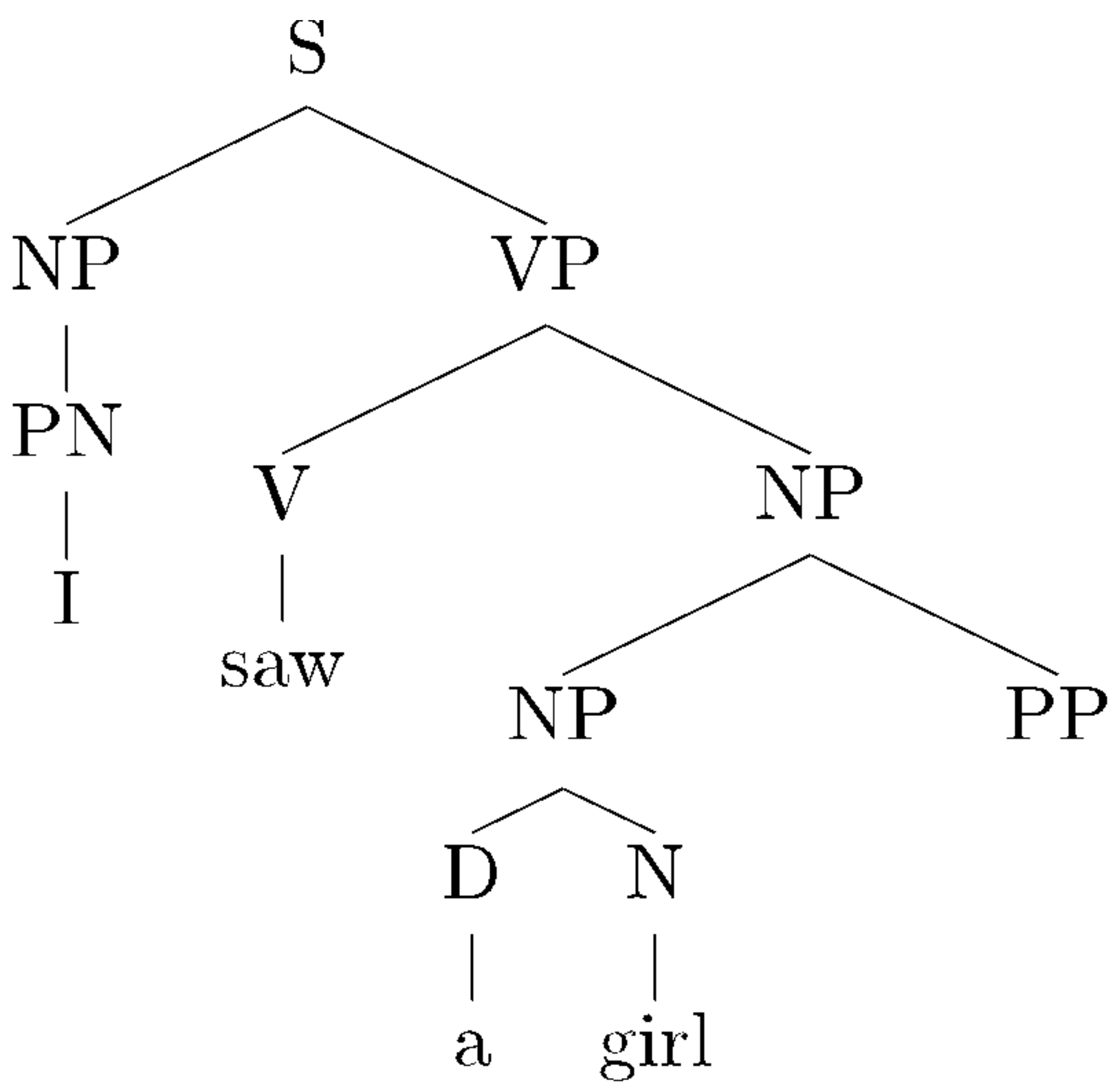
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

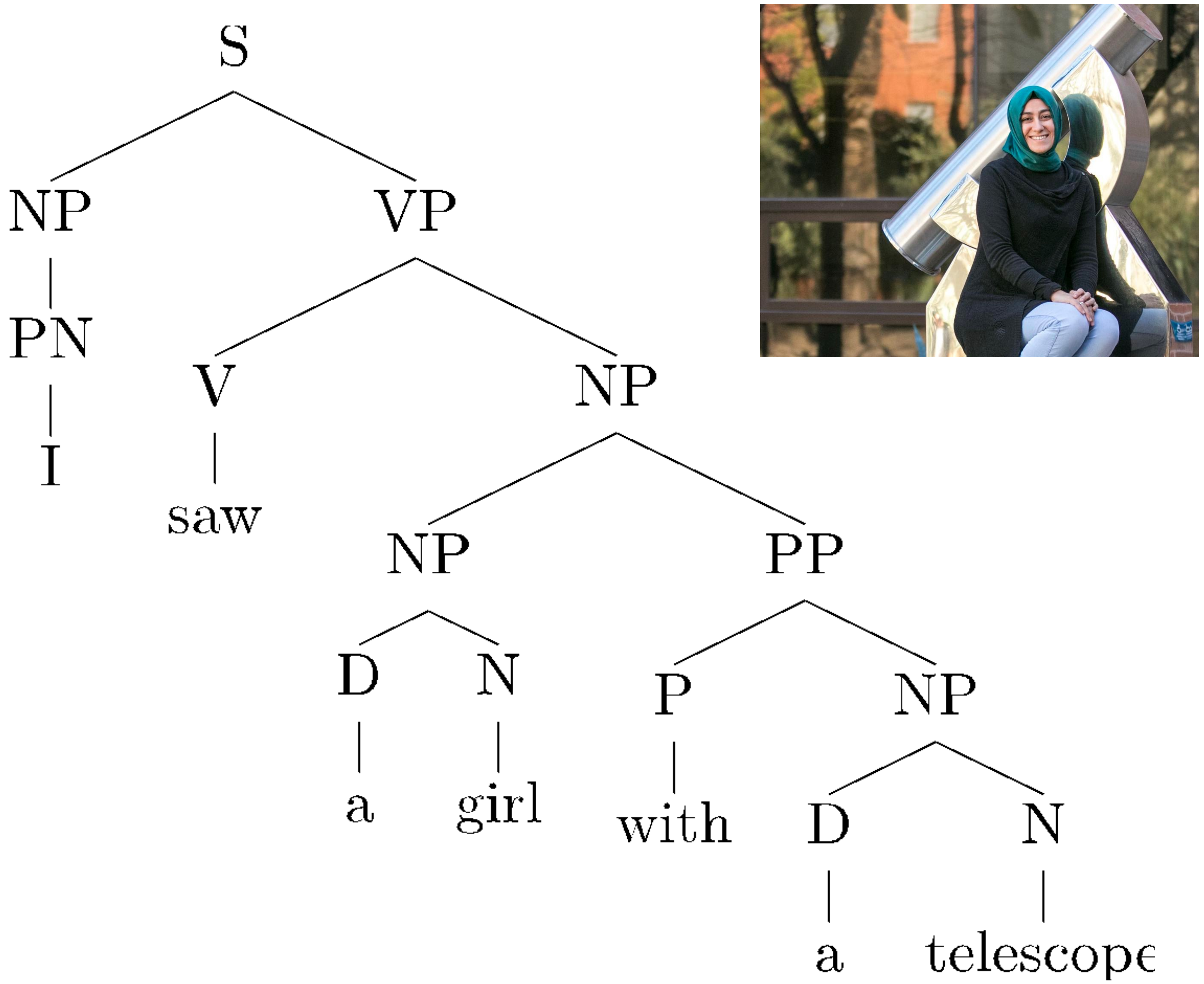
NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)



Grammar (CFG)

S → NP VP

VP → V
VP → V NP
VP → VP PP

NP → NP PP
NP → D N
NP → PN

PP → P NP

Lexicon

N → girl
N → telescope
N → sandwich
PN → I
V → saw
V → ate
P → with
P → in
D → a
D → the

Context-free grammars (CFGs)

- **CFG:** Formal definition. A 4-tuple (N, Σ, R, S):

N	a set of **non-terminal symbols** (or **variables**)
Σ	a set of **terminal symbols** (disjoint from N)
R	a set of **rules** or productions, each of the form $A \rightarrow \beta$,
	where A is a non-terminal,
	β is a string of symbols from the infinite set of strings $(\Sigma \cup N)*$
S	a designated **start symbol** and a member of N

VP, NP, S, PP, …
V, N, P…

saw, telescope, the, girl, …

NP → NP PP, …

ROOT, TOP

An example grammar

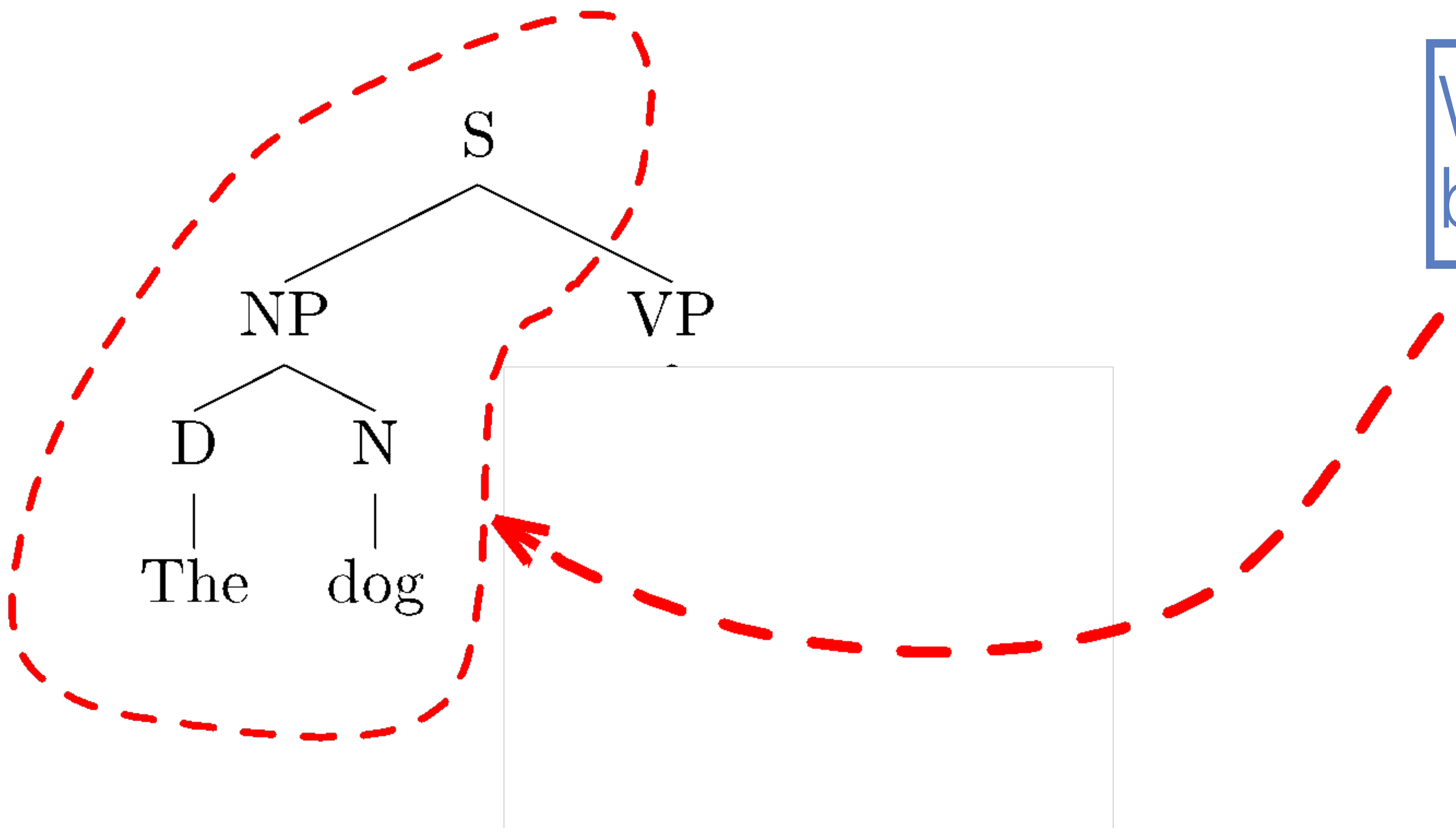
- N = {S, VP, NP, PP, N, V, PN, P}
- Σ = {*girl*, *telescope*, *sandwich*, *I*, *saw*, *ate*, *with*, *in*, *a*, *the*}
- S = {S}
- R =

S → NP VP	(NP a girl) (VP ate a sandwich)
	inner rules
VP → V	
VP → V NP	(V ate) (NP a sandwich)
VP → VP PP	(VP saw a girl) (PP with a telescope)
NP → NP PP	(NP a girl) (PP with a sandwich)
NP → D N	(D a) (N sandwich)
NP → PN	
PP → P NP	(P with) (NP a sandwich)

preterminal rules

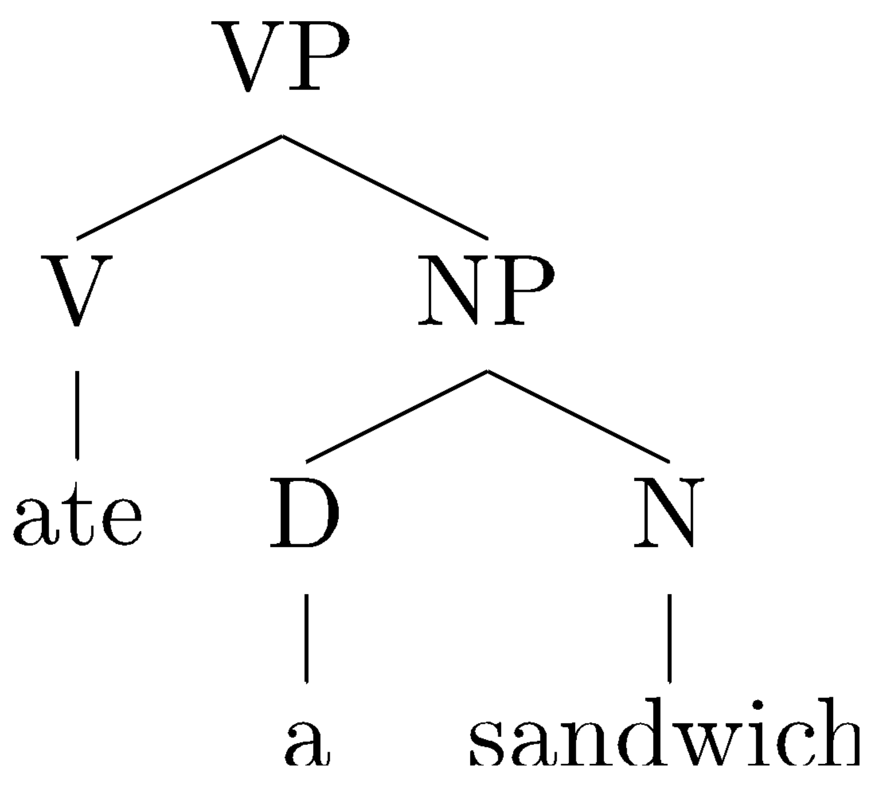
N → *girl*
N → *telescope*
N → *sandwich*
PN → *I*
V → *saw*
V → *ate*
P → *with*
P → *in*
D → *a*
D → *the*

Why "context-free"?

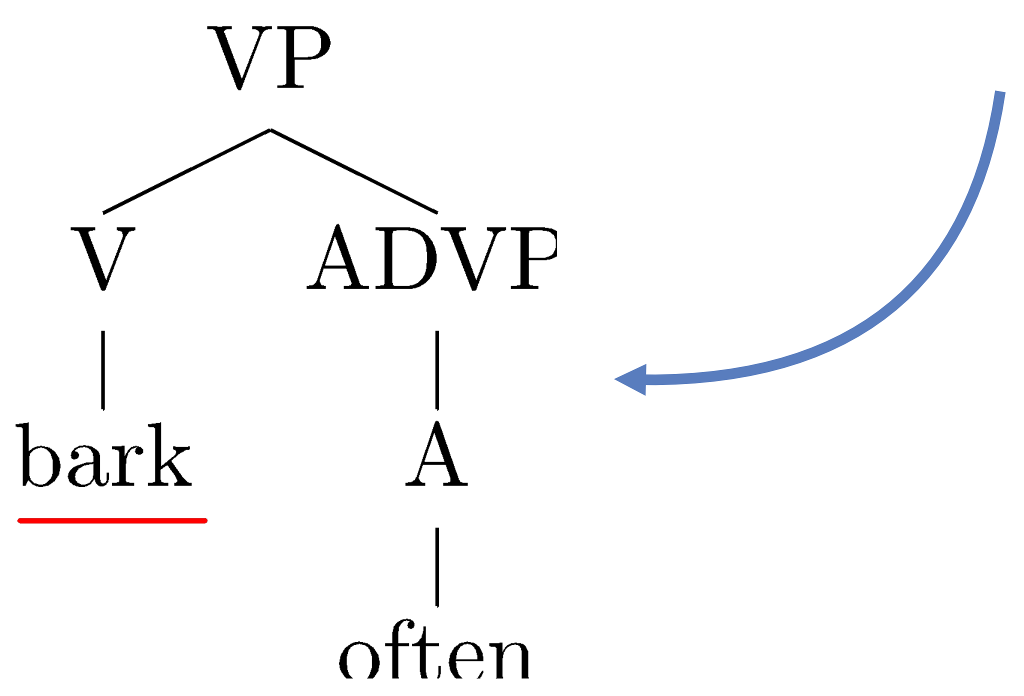


What can be a valid subtree is only effected by the phrase type (VP) but not the **context**.

Example contexts:



not grammatical



Formal Language Theory

Formal Language Theory

Two main classes of models

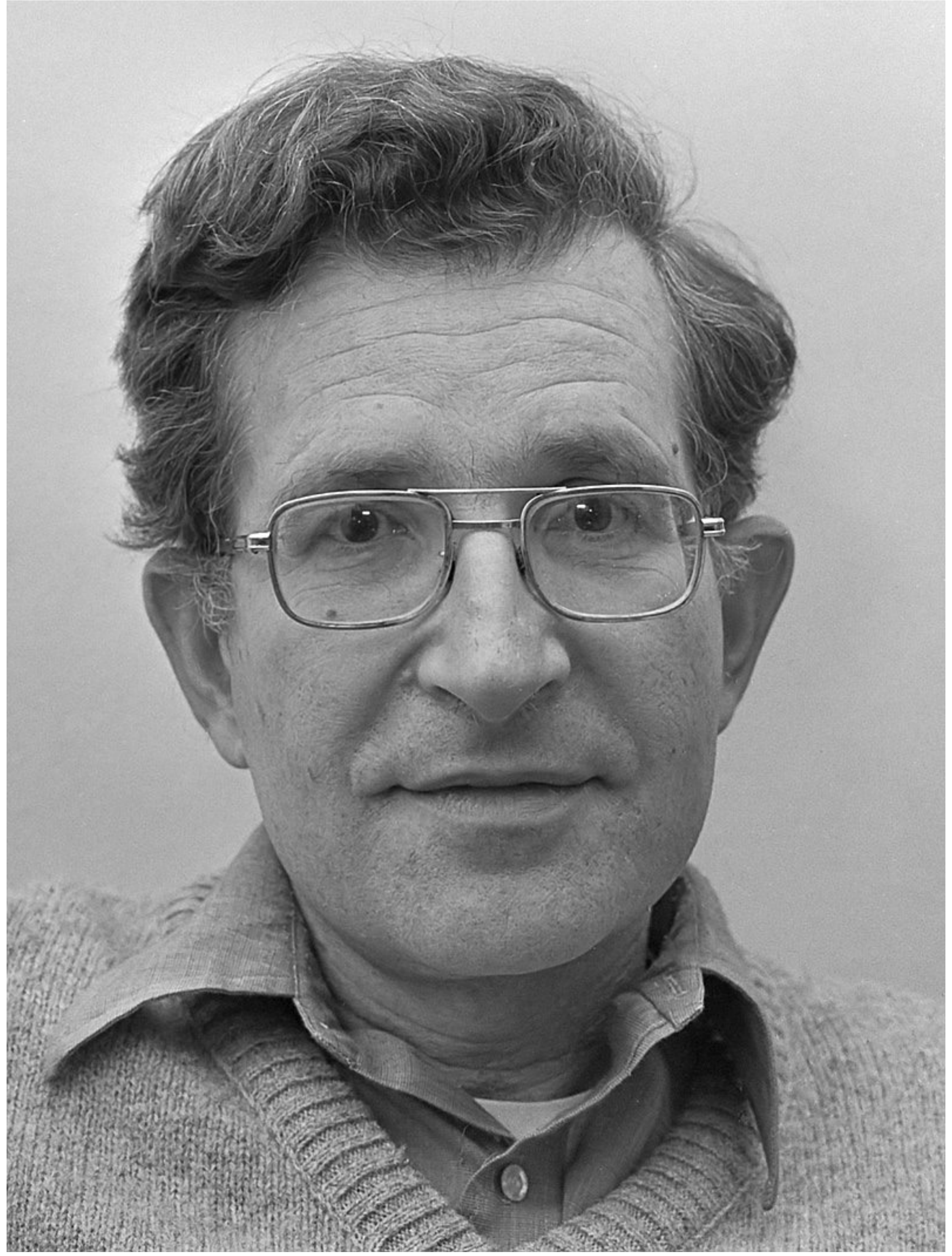
- **Automata**
  - Machines, like Finite-State Automata
- **Grammars**
  - Rule sets, like we have been using to parse

- We can formally prove complexity-class relations between these formal models

Chomsky Hierarchy

- Type 3: Finite State Machines/Regular Expressions/Regular Grammars
  - $A \rightarrow Bw$ or $A \rightarrow w$
- Type 2: Push Down Automata/Context Free Grammars
  - $A \rightarrow \gamma$ where γ is any sequence of terminals/non-terminals
- Type 1: Linear-Bounded Automata/Context Sensitive Grammars
  - $\alpha A \beta \rightarrow \alpha \gamma \beta$ where γ is not empty
- Type 0: Turing Machines/Unrestricted Grammars
  - $aAb \rightarrow aab$ but $bAb \rightarrow bb$

Noam Chomsky, very famous person



1970s version

Most cited living author:

- Linguist
- CS theoretician
- Leftist politics

Might not *always* be right.

Mildly Context-Sensitive Grammars

- We really like CFGs, but are they in fact expressive enough to capture all human grammar?
- Many approaches start with a “CF backbone”, and add registers, equations, or hacks, that are ***not*** CF.
- Several non-hack extensions (CCG, TAG, etc.) turn out to be weakly equivalent!
    - “Mildly context sensitive”
        - So CSFs get even less respect…
        - And so much for the Chomsky Hierarchy being such a big deal

Similarly hard English examples (Center Embedding)

The cat likes tuna fish

The cat the dog chased likes tuna fish

The cat the dog the mouse scared chased likes tuna fish

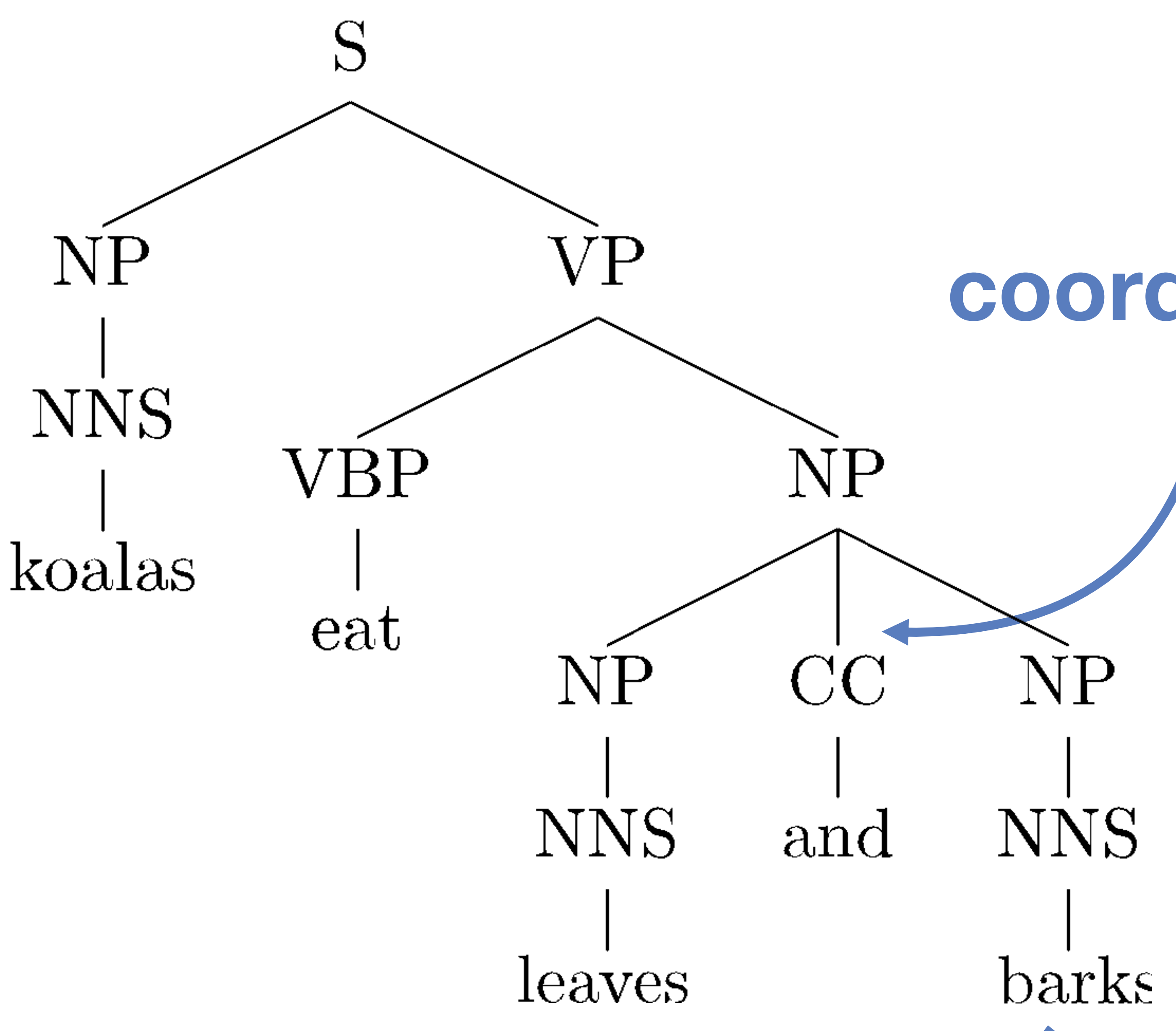
The cat the dog the mouse the elephant squashed scared chased likes tuna fish

The cat the dog the mouse the elephant the flea bit squashed scared chased likes tuna fish

The cat the dog the mouse the elephant the flea the virus infected bit squashed scared chased likes tuna fish

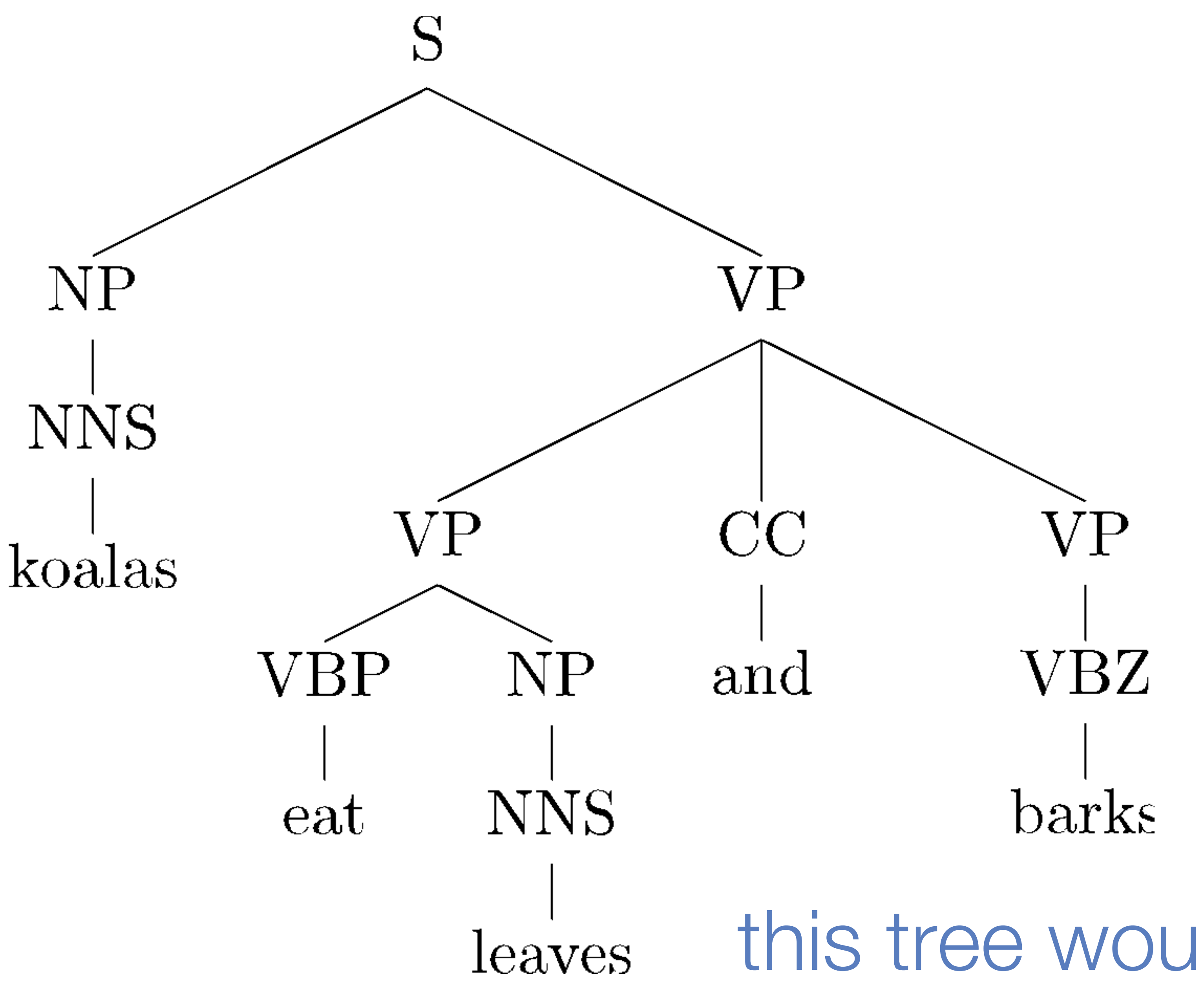
Ambiguity

- Ambiguity makes parsing hard.
- Example: **coordination ambiguity**
  - For example: coarse VP and NP categories can't enforce subject-verb agreement in number, resulting in this coordination ambiguity.



coordination

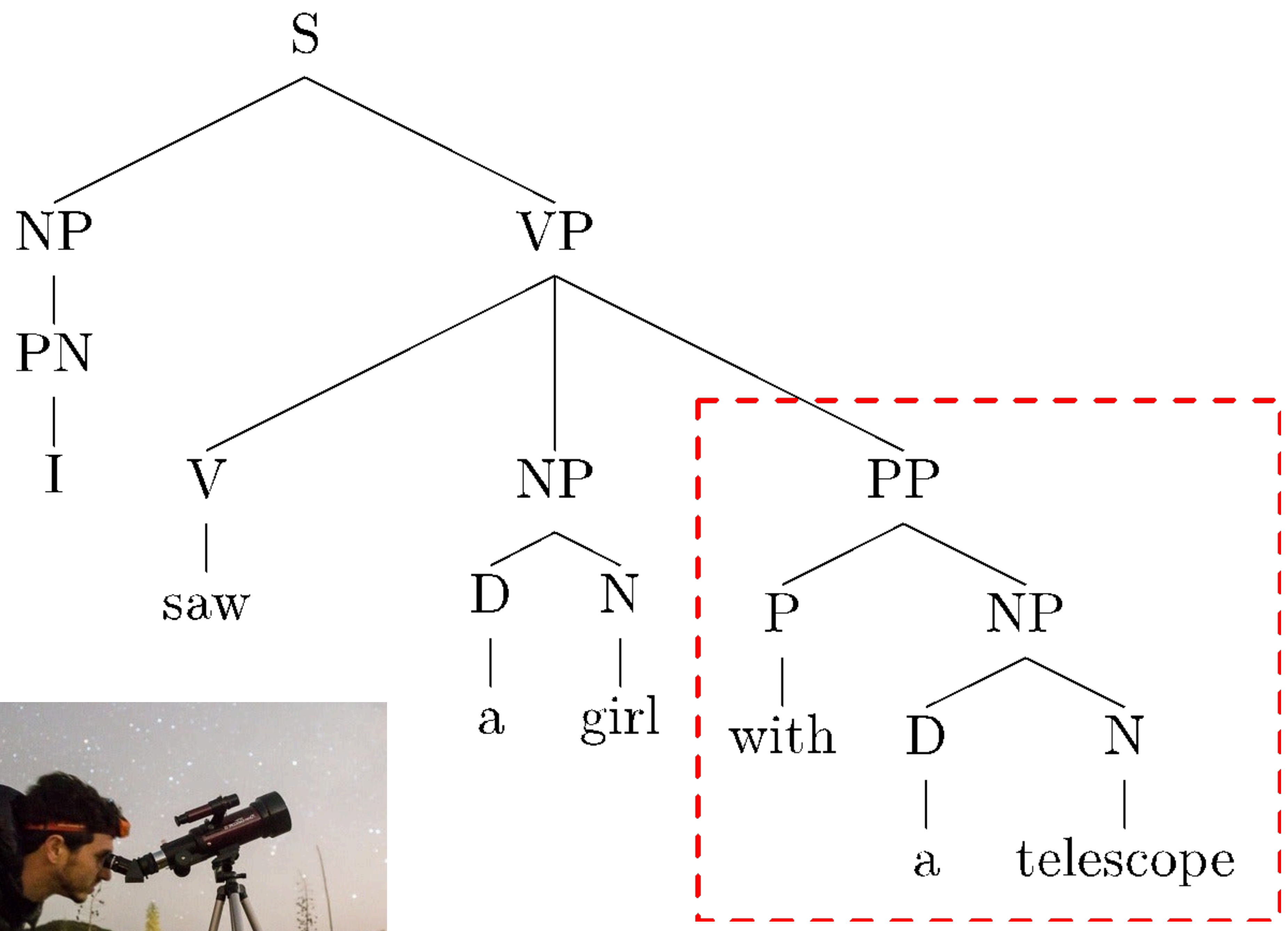
bark may be a noun or a verb

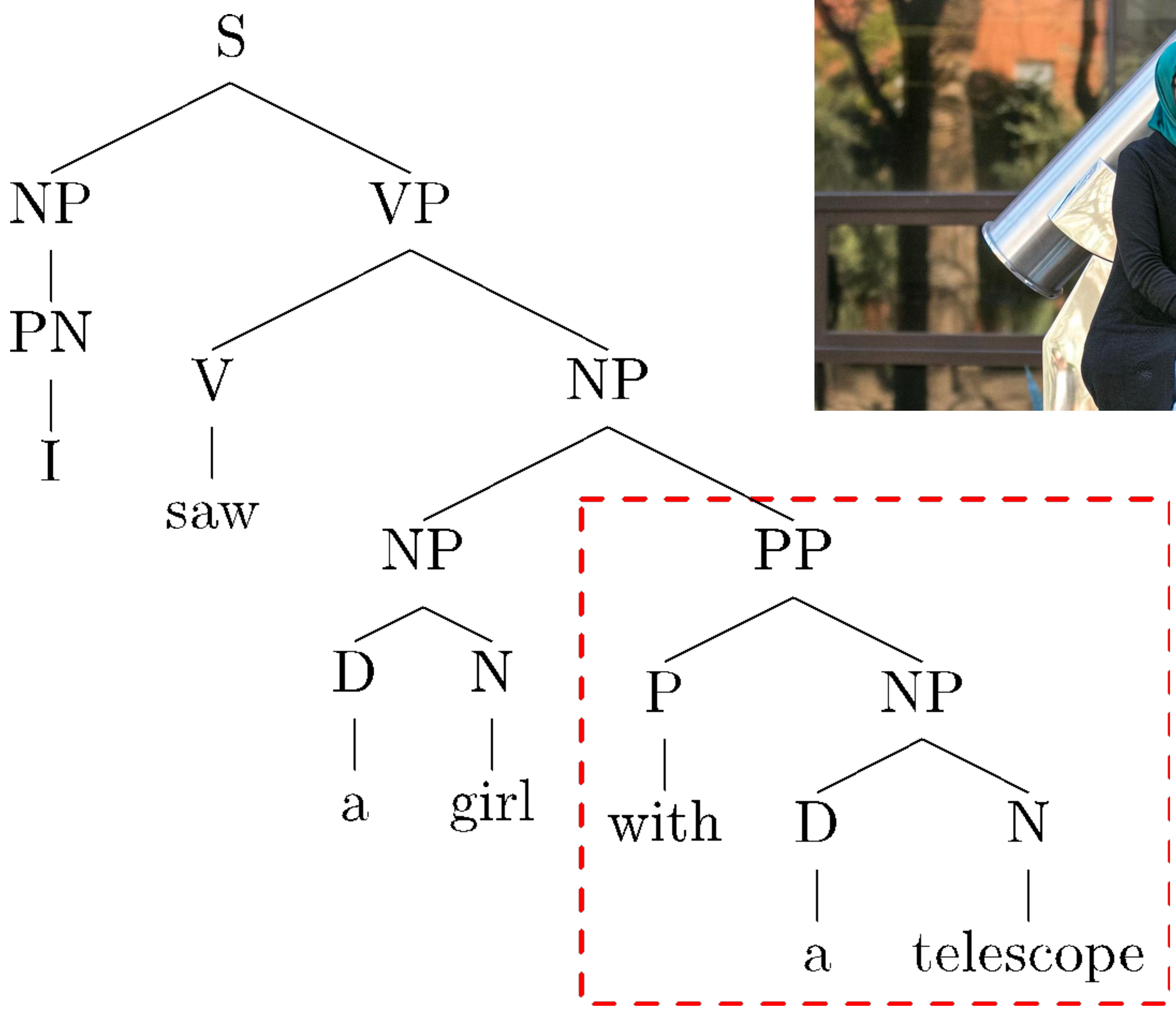


this tree would be ruled out if the context could be captured (subject-verb agreement)

Ambiguity

- Ambiguity makes parsing hard.
- Example: **prepositional phrase attachment ambiguity**

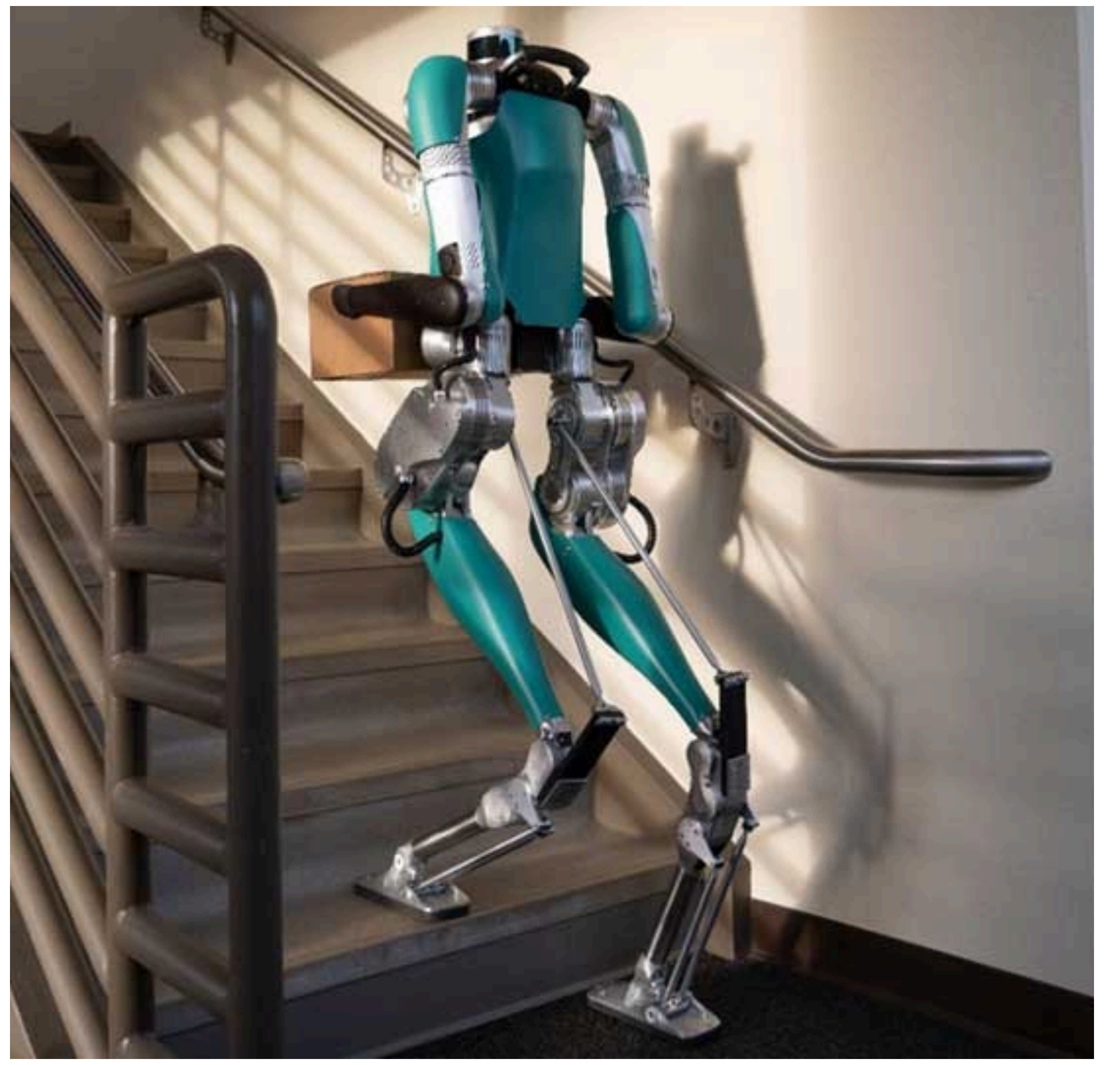




Prepositional phrase ambiguity

"Put the block in the box on the table in the kitchen."

- 3 prepositional phrases, 5 interpretations:
  - Put the block **((in the box on the table) in the kitchen.)**
  - Put the block (in the box (on the table in the kitchen.)
  - Put ((the block in the box) on the table) in the kitchen.
  - Put (the block (in the box on the table)) in the kitchen.
  - Put **(the block in the box) (on the table in the kitchen.)**



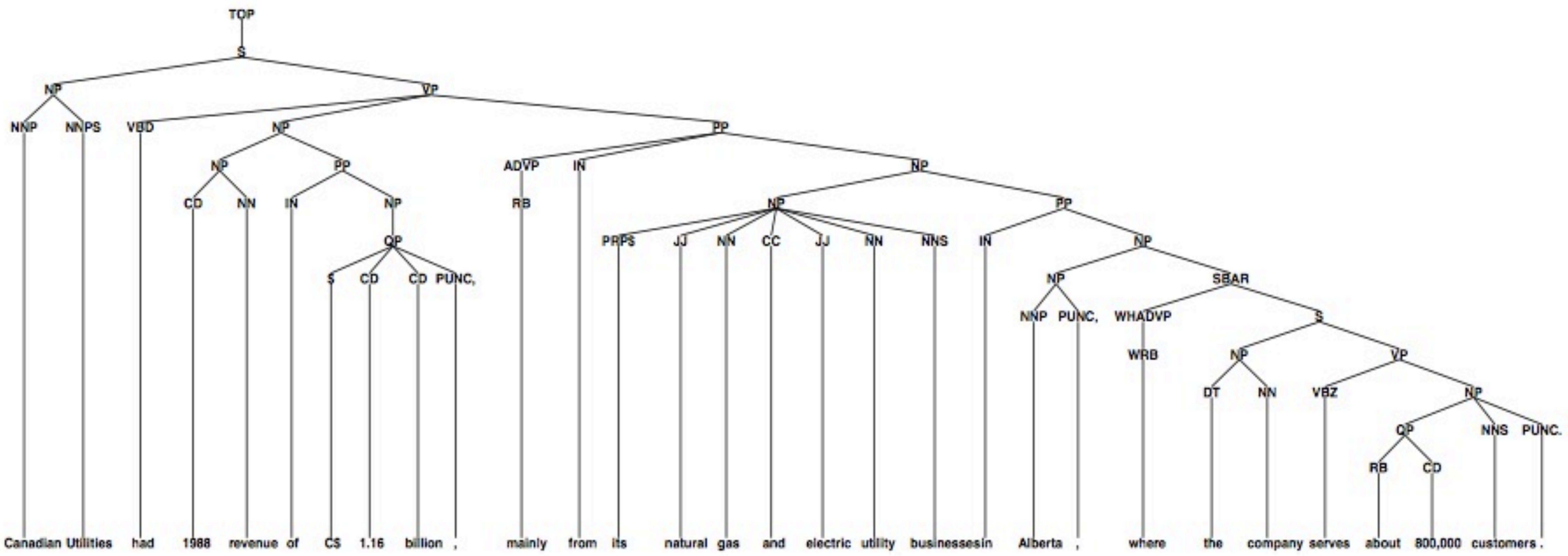
- General case:
  - ((())) ()(()) ()()() (())() (()())

Catalan numbers:

$$Cat_n = \binom{2n}{n} - \binom{2n}{n-1} \sim \frac{4^n}{n^{3/2}\sqrt{\pi}}$$

$$1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, \ldots$$

Typical tree



Canadian Utilities had 1988 revenue of $ 1.16 billion , mainly from its natural gas and electric utility businesses in Alberta , where the company serves about 800,000 customers .

More syntactic ambiguities

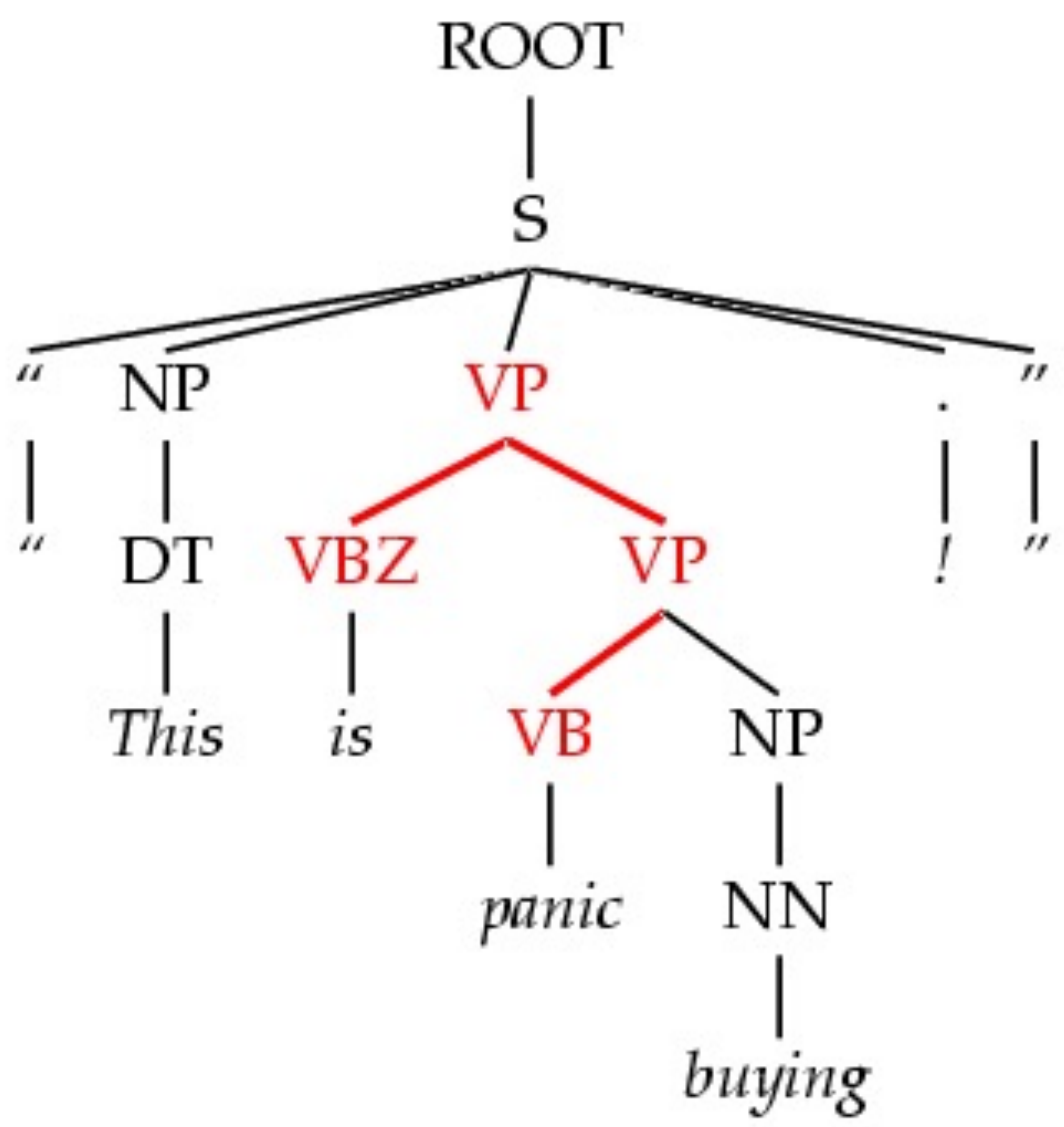
- **Prepositional phrases:**
  They cooked the beans in the pot on the stove with handles.
- **Particle vs. preposition:**
  The puppy tore up the staircase
- **Complement structures:**
  The tourists objected to the guide that they couldn't hear.
  She knows you like the back of her hand.
- **Gerund vs. participal adjective:**
  Visiting relatives can be boring.
  Changing schedules frequently confused passengers. the chicken

Dark ambiguities

- **Dark ambiguities:** most analyses are shockingly bad (meaning, they don't have an interpretation you can get your mind around.)

This analysis corresponds to the correct parse of:

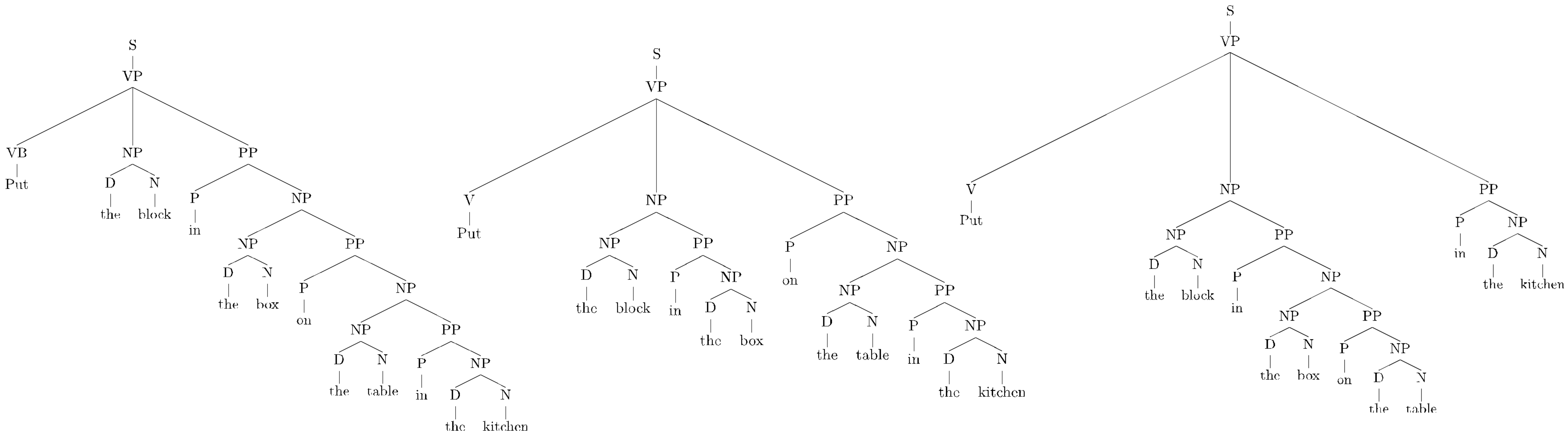
"This is panic buying!"



- Unknown words and new usages

- Solution: need mechanisms to focus attention on the best ones… probabilistic techniques do this.

How to deal with ambiguity?



Put the block in the box on the table in the kitchen.

- Want to score all derivations to encode how plausible they are.

Probabilistic context-free grammars (PCFGs)

- **CFG:** A 4-tuple (N, Σ, R, S):

N	a set of **non-terminal symbols** (or **variables**)
Σ	a set of **terminal symbols** (disjoint from N)
R	a set of **rules** or productions, each of the form $A \rightarrow \beta$,
	where A is a non-terminal,
	β is a string of symbols from the infinite set of strings $(\Sigma \cup N)*$
S	a designated **start symbol** and a member of N

- A **PCFG** adds: a top-down production probability per rule.
  - If each rule is of the form $X \rightarrow Y_1 Y_2 \ldots Y_k$
  - Model its probability: $P(Y_1 Y_2 \ldots Y_k \mid X)$

An example PCFG

- Associate probabilities with the rules: $P(X \to \alpha)$ $\quad \forall X \to \alpha \in R: 0 \leq P(X \to \alpha) \leq 1$

$$\forall X \in N: \sum_{\alpha : X \to \alpha \in R} P(X \to \alpha) = 1$$

S → NP VP	1.0	(NP a girl) (VP ate a sandwich)
VP → V	0.2	
VP → V NP	0.4	(V ate) (NP a sandwich)
VP → VP PP	0.4	(VP saw a girl) (PP with a telescope)
NP → NP PP	0.3	(NP a girl) (PP with a sandwich)
NP → D N	0.5	(D a) (N sandwich)
NP → PN	0.2	
PP → P NP	1.0	(P with) (NP a sandwich)

N → *girl*	0.2
N → *telescope*	0.7
N → *sandwich*	0.1
PN → *I*	1.0
V → *saw*	0.5
V → *ate*	0.5
P → *with*	0.6
P → *in*	0.4
D → *a*	0.3
D → *the*	0.7

Now we can score a tree as a product of probabilities corresponding to the used rules!

PCFGs

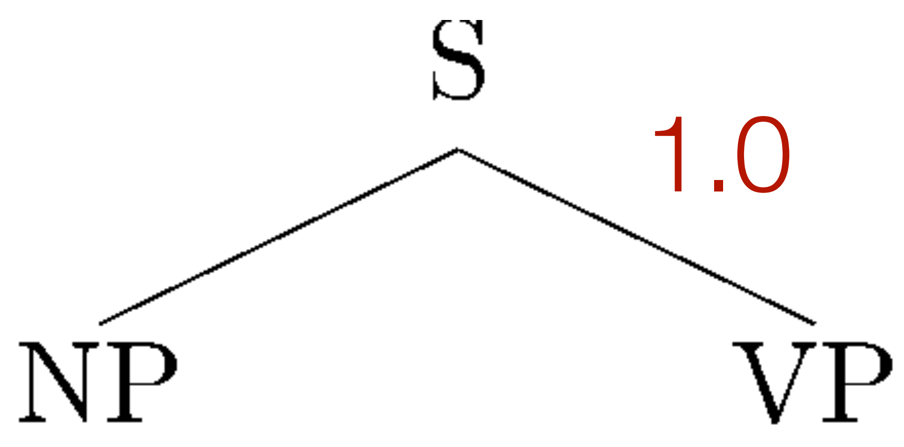
S

S → NP VP	1.0
VP → V	0.2
VP → V NP	0.4
VP → VP PP	0.4
NP → NP PP	0.3
NP → D N	0.5
NP → PN	0.2
PP → P NP	1.0

N → *girl*	0.2
N → *telescope*	0.7
N → *sandwich*	0.1
PN → *I*	1.0
V → *saw*	0.5
V → *ate*	0.5
P → *with*	0.6
P → *in*	0.4
D → *a*	0.3
D → *the*	0.7

$P(T) =$

PCFGs

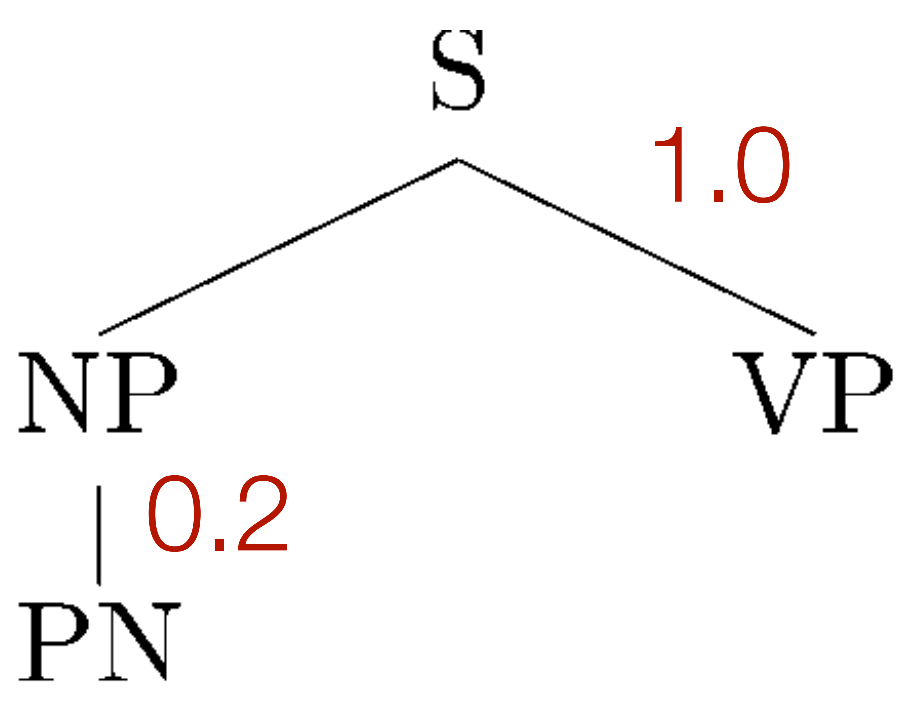


S → NP VP	1.0
VP → V	0.2
VP → V NP	0.4
VP → VP PP	0.4
NP → NP PP	0.3
NP → D N	0.5
NP → PN	0.2
PP → P NP	1.0

N → *girl*	0.2
N → *telescope*	0.7
N → *sandwich*	0.1
PN → *I*	1.0
V → *saw*	0.5
V → *ate*	0.5
P → *with*	0.6
P → *in*	0.4
D → *a*	0.3
D → *the*	0.7

$P(T) = 1.0 *$

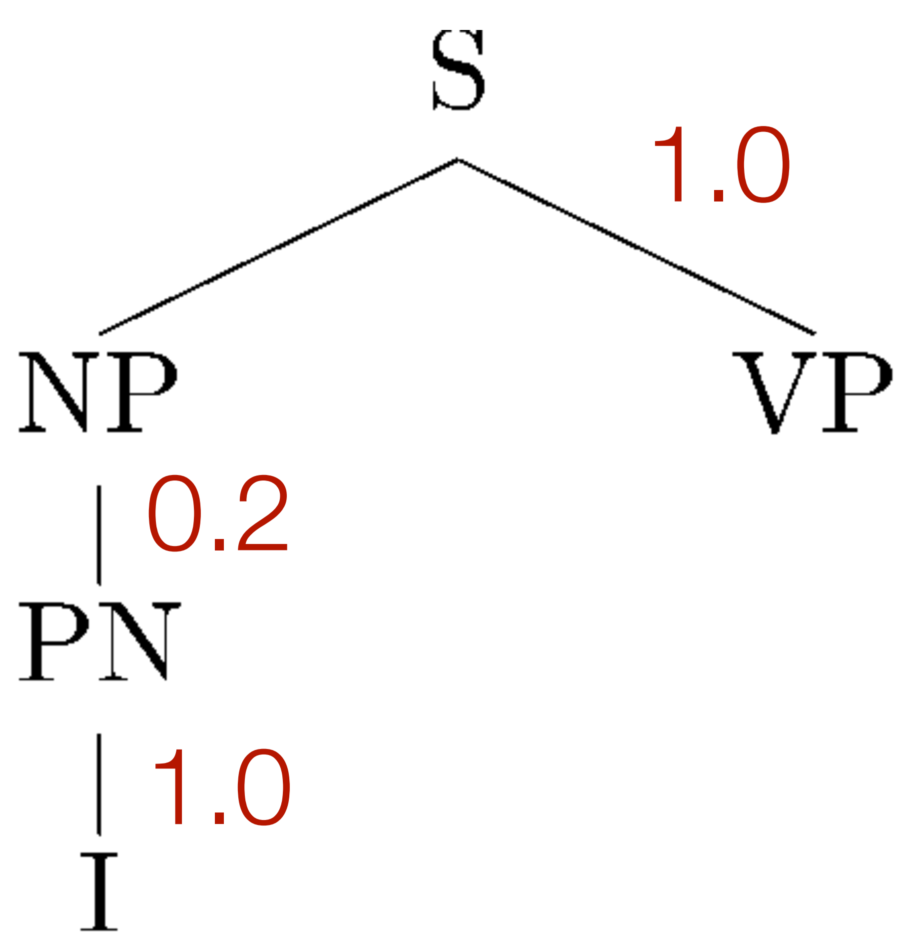
PCFGs



S → NP VP	1.0	N → *girl*	0.2
VP → V	0.2	N → *telescope*	0.7
VP → V NP	0.4	N → *sandwich*	0.1
VP → VP PP	0.4	PN → *I*	1.0
NP → NP PP	0.3	V → *saw*	0.5
NP → D N	0.5	V → *ate*	0.5
NP → PN	0.2	P → *with*	0.6
PP → P NP	1.0	P → *in*	0.4
		D → *a*	0.3
		D → *the*	0.7

$P(T) = 1.0 * 0.2 *$

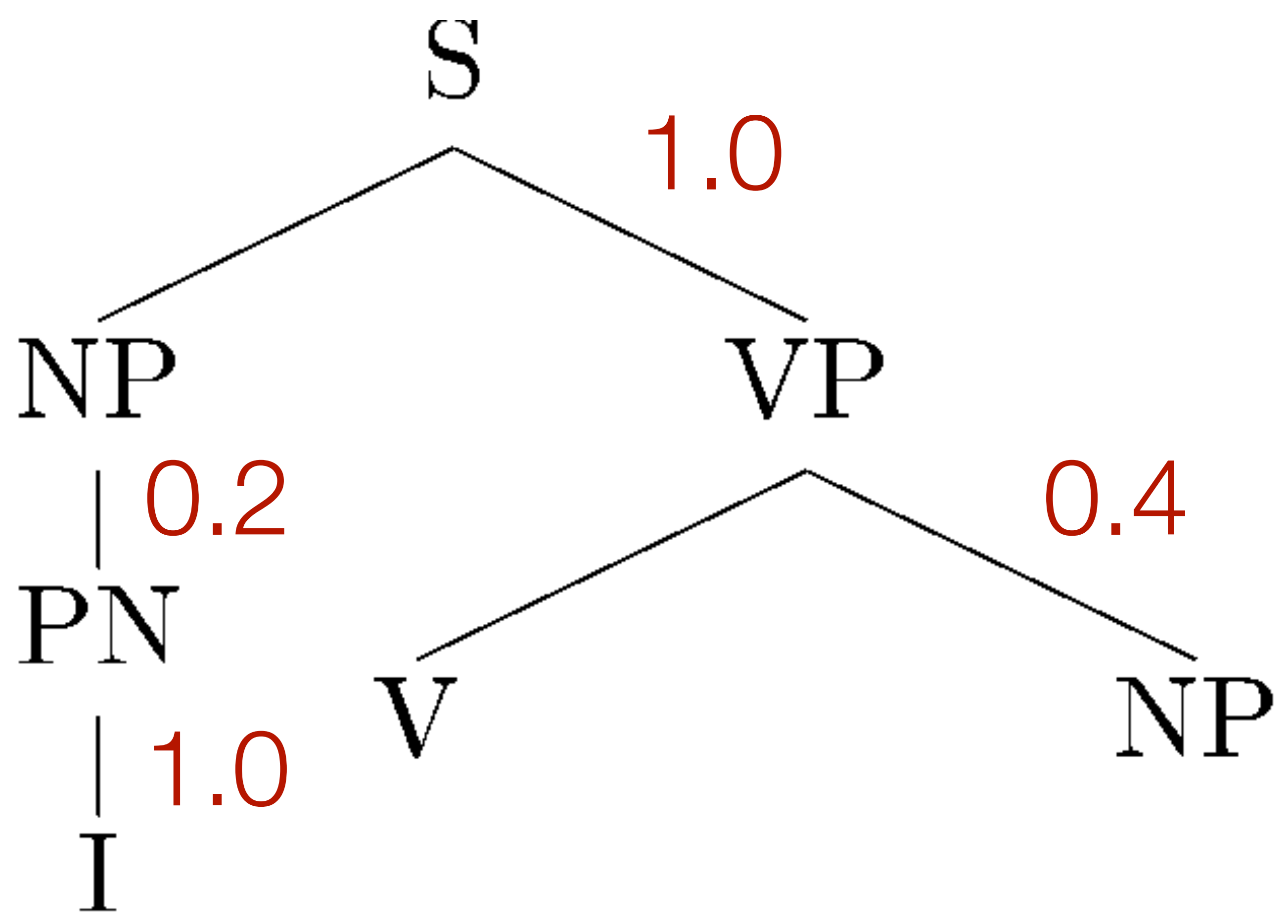
PCFGs



S → NP VP	1.0	N → *girl*	0.2
VP → V	0.2	N → *telescope*	0.7
VP → V NP	0.4	N → *sandwich*	0.1
VP → VP PP	0.4	PN → *I*	1.0
NP → NP PP	0.3	V → *saw*	0.5
NP → D N	0.5	V → *ate*	0.5
NP → PN	0.2	P → *with*	0.6
PP → P NP	1.0	P → *in*	0.4
		D → *a*	0.3
		D → *the*	0.7

$P(T)$ = 1.0 * 0.2 * 1.0 *

PCFGs

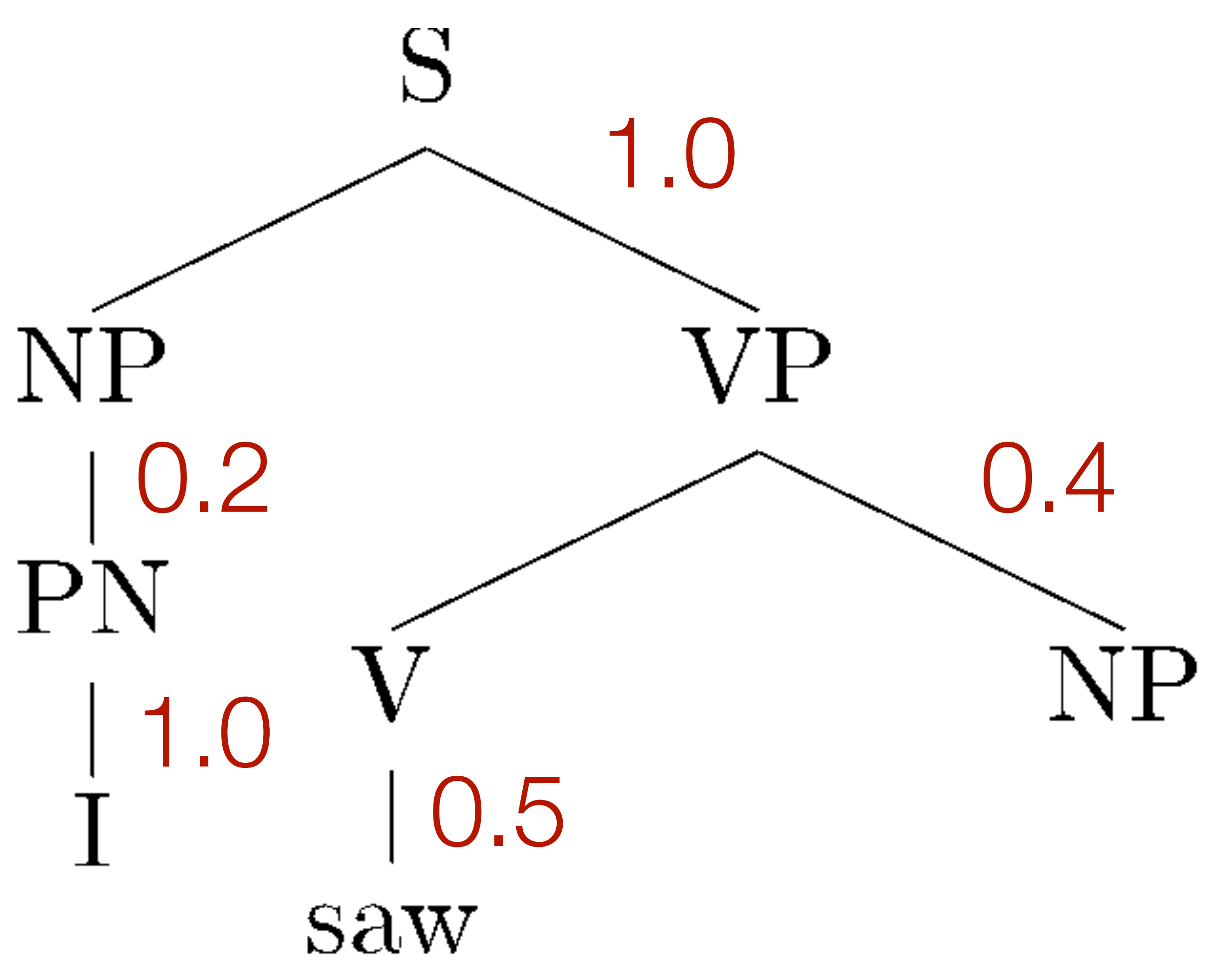


S → NP VP	1.0
VP → V	0.2
VP → V NP	0.4
VP → VP PP	0.4
NP → NP PP	0.3
NP → D N	0.5
NP → PN	0.2
PP → P NP	1.0

N → *girl*	0.2
N → *telescope*	0.7
N → *sandwich*	0.1
PN → *I*	1.0
V → *saw*	0.5
V → *ate*	0.5
P → *with*	0.6
P → *in*	0.4
D → *a*	0.3
D → *the*	0.7

P(*T*) = 1.0 * 0.2 * 1.0 * 0.4 *

PCFGs

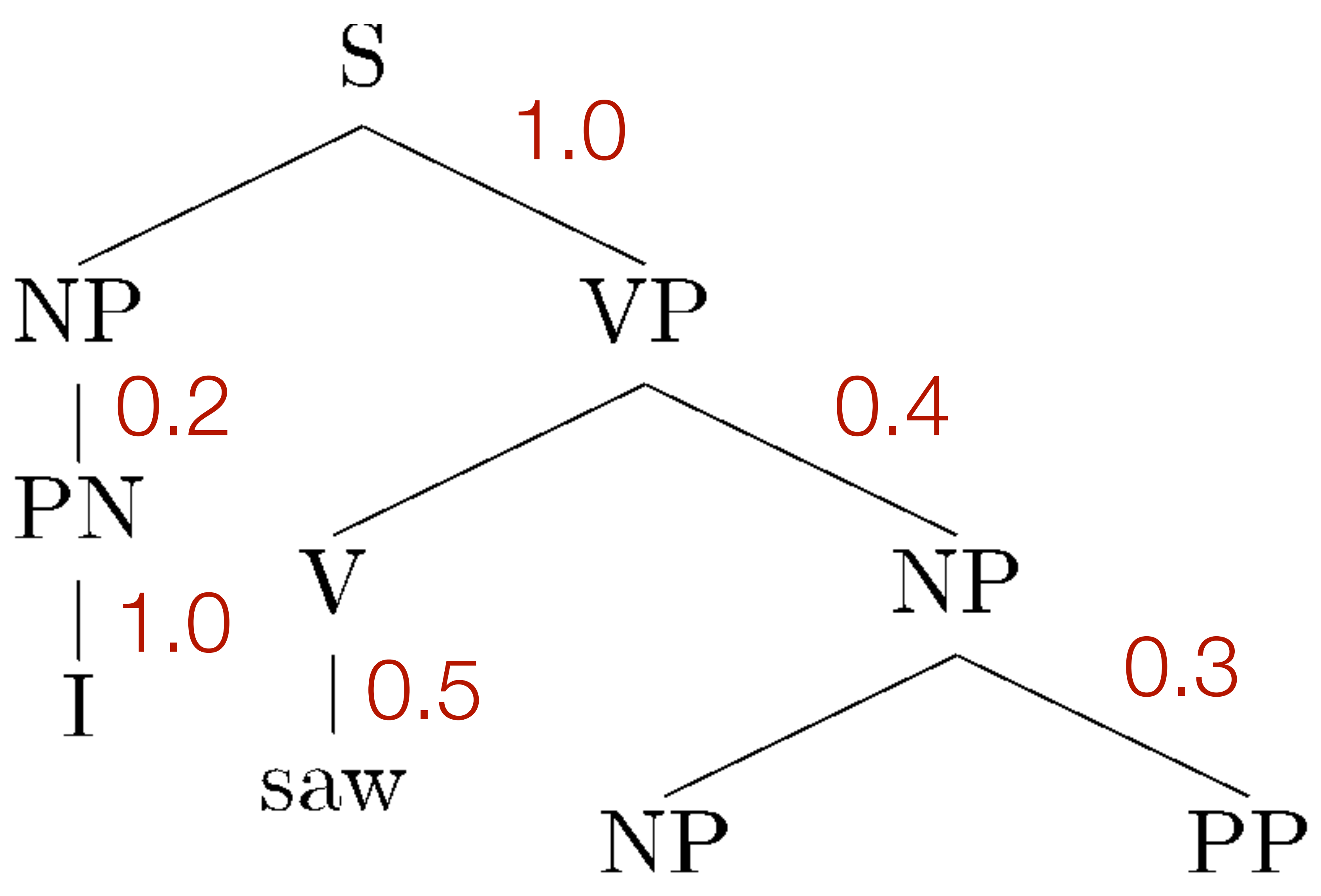


S → NP VP	1.0
VP → V	0.2
VP → V NP	0.4
VP → VP PP	0.4
NP → NP PP	0.3
NP → D N	0.5
NP → PN	0.2
PP → P NP	1.0

N → *girl*	0.2
N → *telescope*	0.7
N → *sandwich*	0.1
PN → *I*	1.0
V → *saw*	0.5
V → *ate*	0.5
P → *with*	0.6
P → *in*	0.4
D → *a*	0.3
D → *the*	0.7

P(*T*) = 1.0 * 0.2 * 1.0 * 0.4 * 0.5 *

PCFGs

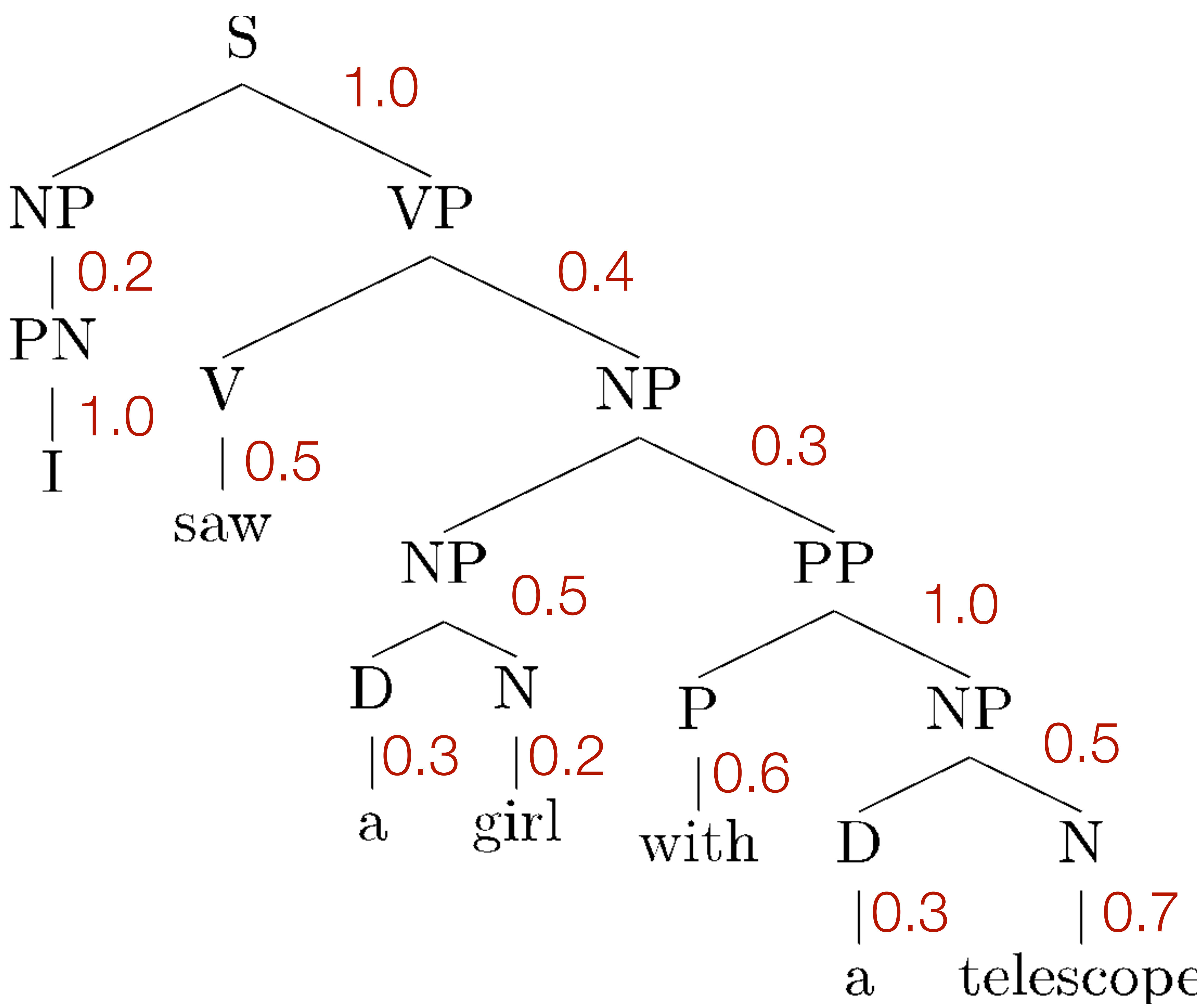


S → NP VP	1.0
VP → V	0.2
VP → V NP	0.4
VP → VP PP	0.4
NP → NP PP	0.3
NP → D N	0.5
NP → PN	0.2
PP → P NP	1.0

N → *girl*	0.2
N → *telescope*	0.7
N → *sandwich*	0.1
PN → *I*	1.0
V → *saw*	0.5
V → *ate*	0.5
P → *with*	0.6
P → *in*	0.4
D → *a*	0.3
D → *the*	0.7

$P(T) = 1.0 * 0.2 * 1.0 * 0.4 * 0.5 * 0.3 *$

PCFGs



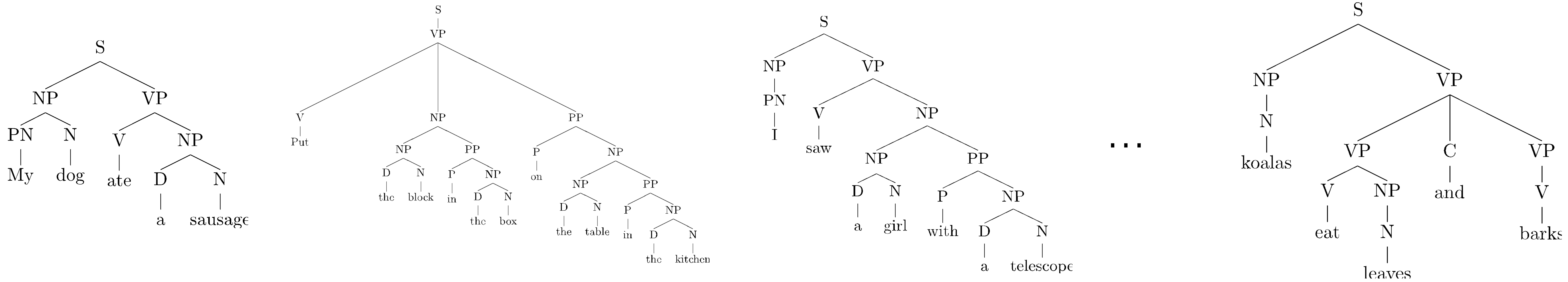
S → NP VP	1.0
VP → V	0.2
VP → V NP	0.4
VP → VP PP	0.4
NP → NP PP	0.3
NP → D N	0.5
NP → PN	0.2
PP → P NP	1.0

N → *girl*	0.2
N → *telescope*	0.7
N → *sandwich*	0.1
PN → *I*	1.0
V → *saw*	0.5
V → *ate*	0.5
P → *with*	0.6
P → *in*	0.4
D → *a*	0.3
D → *the*	0.7

$P(T) = 1.0 * 0.2 * 1.0 * 0.4 * 0.5 * 0.3 * 0.5 * 0.3 * 0.2 * 1.0 * 0.6 * 0.5 * 0.3 * 0.7 = 2.26\text{e-}5$

PCFG estimation

- A treebank: a collection of sentences annotated with constituency trees



- Estimated probability of a rule (maximum likelihood estimate):

$$P(X \to \alpha) = \frac{C(X \to \alpha)}{C(X)}$$

$C(X \to \alpha)$: # times the rule was used in the corpus

$C(X)$: # times nonterminal X appeared in the treebank

- Smoothing is helpful (especially for preterminal rules).

Distribution over trees

- We defined a distribution over production rules for each nonterminal.
- Our goal was to define a distribution over parse trees.
    - Unfortunately, not all PCFGs result in a proper distribution over trees, i.e. the sum over probabilities of all trees in the grammar may be less than 1.
- Fortunately: any PCFG estimated by maximum likelihood is always proper [Chi and Geman, 1998].

[Hao: CFG/PCFG parsing]