Natural Language Processing

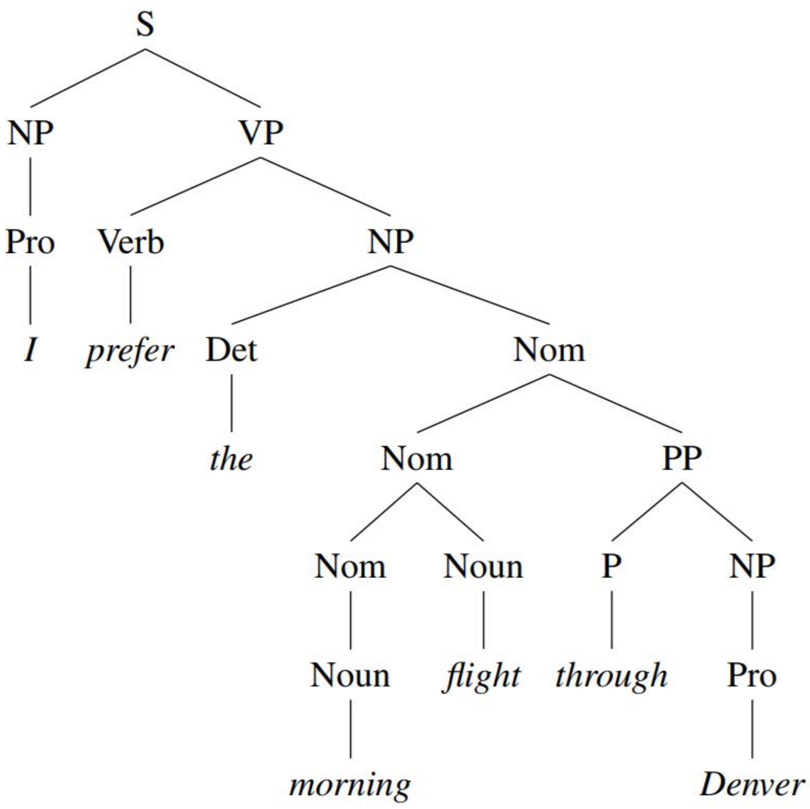
Syntactic parsing

Yulia Tsvetkov

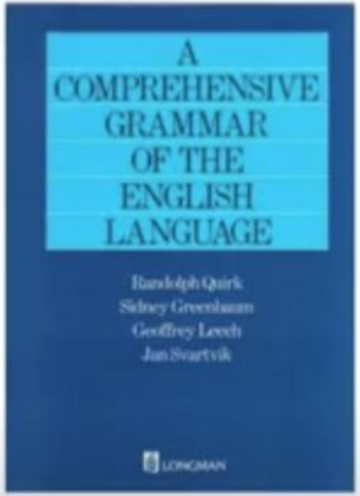
yuliats@cs.washington.edu

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Constituent (phrase-structure) representation



English grammar



Product Details (from Amazon)
Hardcover: 1779 pages
Publisher: Longman; 2nd Revised edition
Language: English
ISBN-10: 0582517346
ISBN-13: 978-0582517349
Product Dimensions: 8.4 x 2.4 x 10 inches
Shipping Weight: 4.6 pounds

Context Free Grammar (CFG)

Context Free Grammar (CFG)

Grammar (CFG)

ROOT → S
S → NP VP
NP → DT NN
NP → NN NNS
NP → NP PP
VP → VBP NP
VP → VBP NP PP
PP → IN NP

Lexicon

NN → interest
NNS → raises
VBP → interest
VBZ → raises
…

Other grammar formalisms: LFG, HPSG, TAG, CCG…

CFGs

S

$$S \rightarrow NP\ VP$$

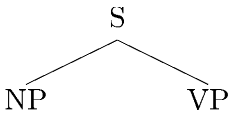
$$VP \rightarrow V$$
$$VP \rightarrow V\ NP$$
$$VP \rightarrow VP\ PP$$

$$NP \rightarrow NP\ PP$$
$$NP \rightarrow D\ N$$
$$NP \rightarrow PN$$

$$PP \rightarrow P\ NP$$

$$N \rightarrow girl$$
$$N \rightarrow telescope$$
$$N \rightarrow sandwich$$
$$PN \rightarrow I$$
$$V \rightarrow saw$$
$$V \rightarrow ate$$
$$P \rightarrow with$$
$$P \rightarrow in$$
$$D \rightarrow a$$
$$D \rightarrow the$$

CFGs



$$S \rightarrow NP\ VP$$

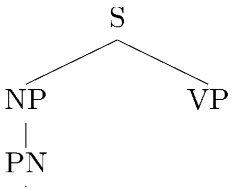
$$VP \rightarrow V$$
$$VP \rightarrow V\ NP$$
$$VP \rightarrow VP\ PP$$

$$NP \rightarrow NP\ PP$$
$$NP \rightarrow D\ N$$
$$NP \rightarrow PN$$

$$PP \rightarrow P\ NP$$

$$N \rightarrow girl$$
$$N \rightarrow telescope$$
$$N \rightarrow sandwich$$
$$PN \rightarrow I$$
$$V \rightarrow saw$$
$$V \rightarrow ate$$
$$P \rightarrow with$$
$$P \rightarrow in$$
$$D \rightarrow a$$
$$D \rightarrow the$$

CFGs



$$S \rightarrow NP \ \ VP$$

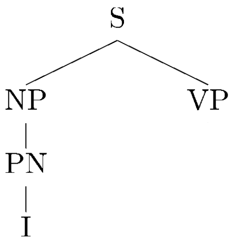
$$VP \rightarrow V$$
$$VP \rightarrow V \ \ NP$$
$$VP \rightarrow VP \ \ PP$$

$$NP \rightarrow NP \ \ PP$$
$$NP \rightarrow D \ \ N$$
$$NP \rightarrow PN$$

$$PP \rightarrow P \ \ NP$$

$$N \rightarrow girl$$
$$N \rightarrow telescope$$
$$N \rightarrow sandwich$$
$$PN \rightarrow I$$
$$V \rightarrow saw$$
$$V \rightarrow ate$$
$$P \rightarrow with$$
$$P \rightarrow in$$
$$D \rightarrow a$$
$$D \rightarrow the$$

CFGs



$$S \rightarrow NP \;\; VP$$

$$VP \rightarrow V$$

$$VP \rightarrow V \;\; NP$$

$$VP \rightarrow VP \;\; PP$$

$$NP \rightarrow NP \;\; PP$$

$$NP \rightarrow D \;\; N$$

$$NP \rightarrow PN$$

$$PP \rightarrow P \;\; NP$$

$$N \rightarrow girl$$

$$N \rightarrow telescope$$

$$N \rightarrow sandwich$$

$$PN \rightarrow I$$

$$V \rightarrow saw$$

$$V \rightarrow ate$$

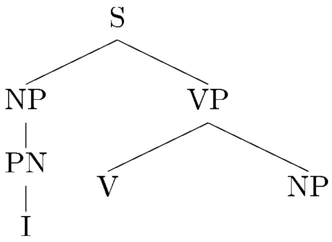
$$P \rightarrow with$$

$$P \rightarrow in$$

$$D \rightarrow a$$

$$D \rightarrow the$$

CFGs



$$S \rightarrow NP\ \ VP$$

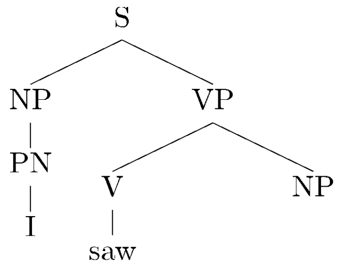
$$VP \rightarrow V$$
$$VP \rightarrow V\ \ NP$$
$$VP \rightarrow VP\ \ PP$$

$$NP \rightarrow NP\ \ PP$$
$$NP \rightarrow D\ \ N$$
$$NP \rightarrow PN$$

$$PP \rightarrow P\ \ NP$$

$$N \rightarrow girl$$
$$N \rightarrow telescope$$
$$N \rightarrow sandwich$$
$$PN \rightarrow I$$
$$V \rightarrow saw$$
$$V \rightarrow ate$$
$$P \rightarrow with$$
$$P \rightarrow in$$
$$D \rightarrow a$$
$$D \rightarrow the$$

CFGs



$$S \rightarrow NP\ \ VP$$

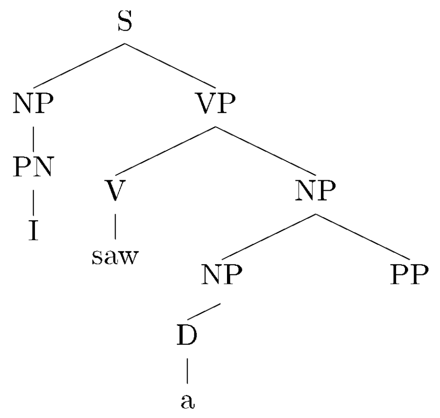
$$VP \rightarrow V$$
$$VP \rightarrow V\ \ NP$$
$$VP \rightarrow VP\ \ PP$$

$$NP \rightarrow NP\ \ PP$$
$$NP \rightarrow D\ \ N$$
$$NP \rightarrow PN$$

$$PP \rightarrow P\ \ NP$$

$$N \rightarrow girl$$
$$N \rightarrow telescope$$
$$N \rightarrow sandwich$$
$$PN \rightarrow I$$
$$V \rightarrow saw$$
$$V \rightarrow ate$$
$$P \rightarrow with$$
$$P \rightarrow in$$
$$D \rightarrow a$$
$$D \rightarrow the$$

CFGs



$$S \to NP\ VP$$

$$VP \to V$$

$$VP \to V\ NP$$

$$VP \to VP\ PP$$

$$NP \to NP\ PP$$

$$NP \to D\ N$$

$$NP \to PN$$

$$PP \to P\ NP$$

$$N \to girl$$

$$N \to telescope$$

$$N \to sandwich$$

$$PN \to I$$

$$V \to saw$$

$$V \to ate$$

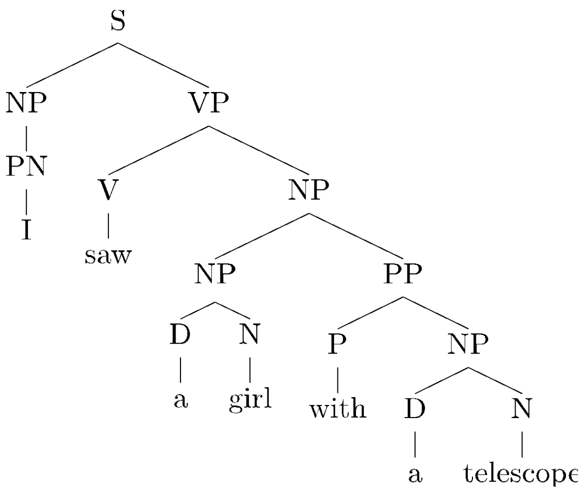
$$P \to with$$

$$P \to in$$

$$D \to a$$

$$D \to the$$

CFGs



$$S \rightarrow NP\ \ VP$$

$$VP \rightarrow V$$
$$VP \rightarrow V\ \ NP$$
$$VP \rightarrow VP\ \ PP$$

$$NP \rightarrow NP\ \ PP$$
$$NP \rightarrow D\ \ N$$
$$NP \rightarrow PN$$

$$PP \rightarrow P\ \ NP$$

$$N \rightarrow girl$$
$$N \rightarrow telescope$$
$$N \rightarrow sandwich$$
$$PN \rightarrow I$$
$$V \rightarrow saw$$
$$V \rightarrow ate$$
$$P \rightarrow with$$
$$P \rightarrow in$$
$$D \rightarrow a$$
$$D \rightarrow the$$

Treebank Sentences

```
( (S (NP-SBJ The move)
     (VP followed
         (NP (NP a round)
             (PP of
                 (NP (NP similar increases)
                     (PP by
                         (NP other lenders))
                     (PP against
                         (NP Arizona real estate loans)))))
         ,
         (S-ADV (NP-SBJ *)
                (VP reflecting
                    (NP (NP a continuing decline)
                        (PP-LOC in
                                (NP that market))))))
     .))
```

Context-Free Grammars

- A context-free grammar is a 4-tuple <N, T, S, R>
  - N : the set of non-terminals
    - **Phrasal categories**: S, NP, VP, ADJP, etc.
    - **Parts-of-speech** (pre-terminals): NN, JJ, DT, VB
  - T : the set of terminals (the words)
  - S : the start symbol
    - Often written as ROOT or TOP
    - Not usually the sentence non-terminal S
  - R : the set of rules
    - Of the form $X \rightarrow Y_1\ Y_2\ \ldots\ Y_k$, with $X, Y_i \in N$
    - Examples: $S \rightarrow NP\ VP$, $VP \rightarrow VP\ CC\ VP$
    - Also called rewrites, productions, or local trees

An example grammar

$N = \{S, VP, NP, PP, N, V, PN, P\}$

$T = \{girl, telescope, sandwich, I, saw, ate, with, in, a, the\}$

$S = \{S\}$

$R:$

Called **Inner rules**

$S \rightarrow NP\ VP$ (NP A girl) (VP ate a sandwich)

$VP \rightarrow V$

$VP \rightarrow V\ NP$ (V ate) (NP a sandwich)

$VP \rightarrow VP\ PP$ (VP saw a girl) (PP with a telescope)

$NP \rightarrow NP\ PP$ (NP a girl) (PP with a sandwich)

$NP \rightarrow D\ N$ (D a) (N sandwich)

$NP \rightarrow PN$

$PP \rightarrow P\ NP$ (P with) (NP with a sandwich)

Preterminal rules

$N \rightarrow girl$

$N \rightarrow telescope$

$N \rightarrow sandwich$

$PN \rightarrow I$

$V \rightarrow saw$

$V \rightarrow ate$

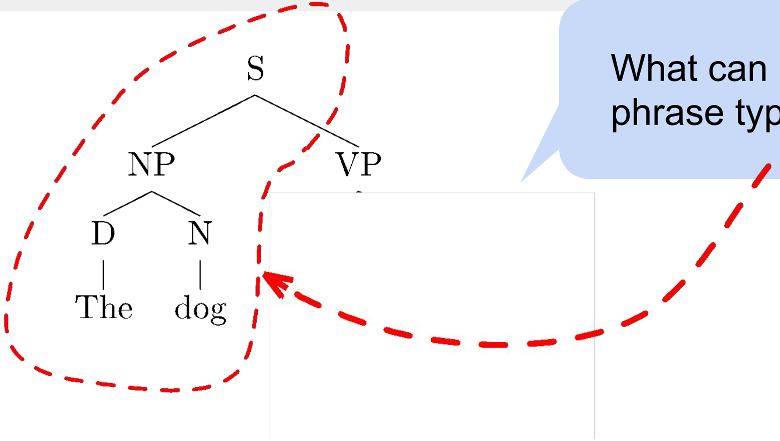
$P \rightarrow with$

$P \rightarrow in$

$D \rightarrow a$

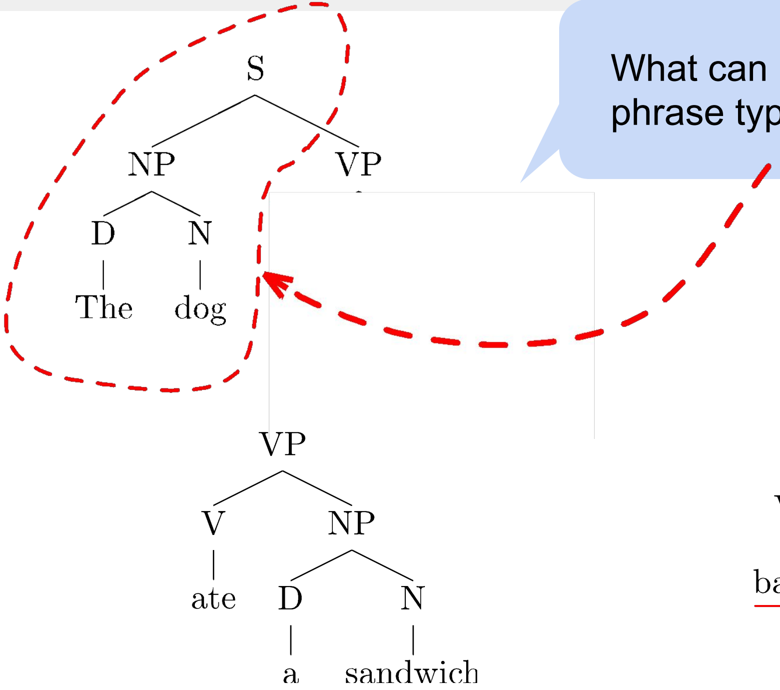
$D \rightarrow the$

Why context-free?

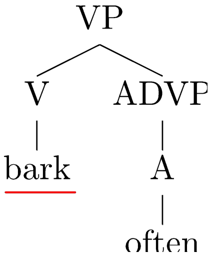


What can be a sub-tree is only affected by what the phrase type is (VP) but not the context

Why context-free?



What can be a sub-tree is only affected by what the phrase type is (VP) but not the context

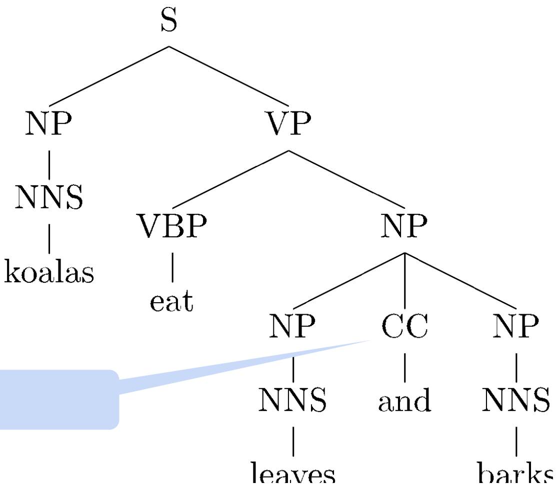


Not grammatical

Ambiguities

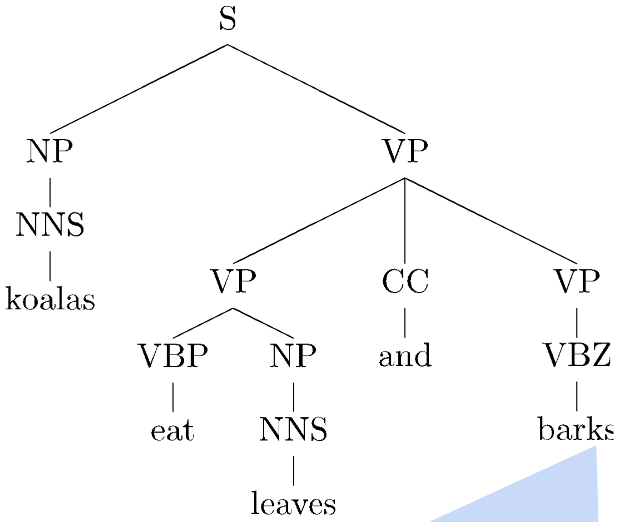
Coordination ambiguity

- Here, the coarse VP and NP categories cannot enforce subject-verb agreement in number resulting in the coordination ambiguity



Coordination

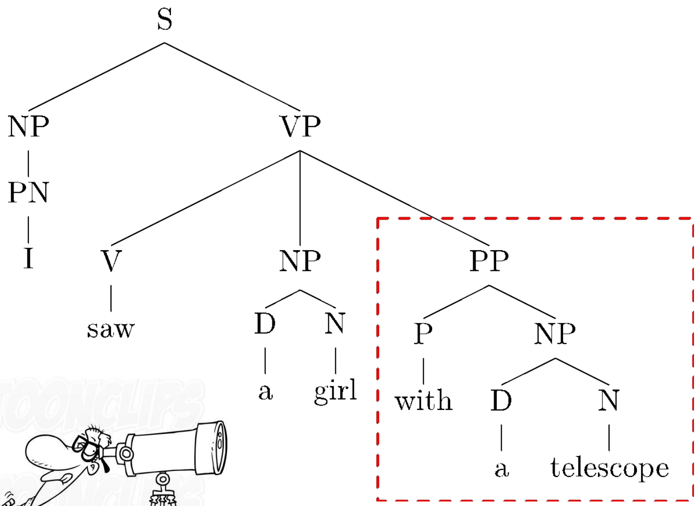
"Bark" can refer both to a noun or a verb

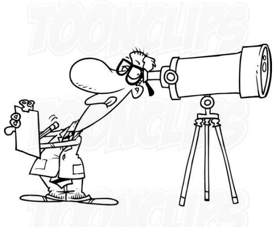


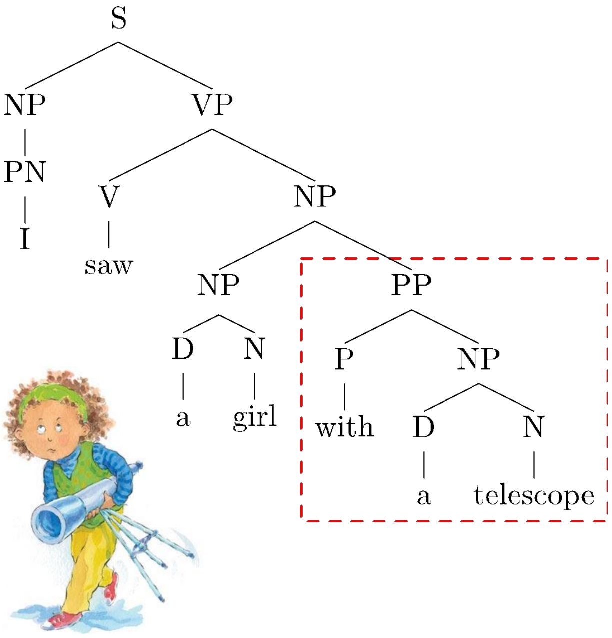
This tree would be ruled out if the context would be somehow captured (subject-verb agreement)

Why is parsing hard? Ambiguity

- Prepositional phrase attachment ambiguity







PP 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)

PP 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))
- …

A general case:

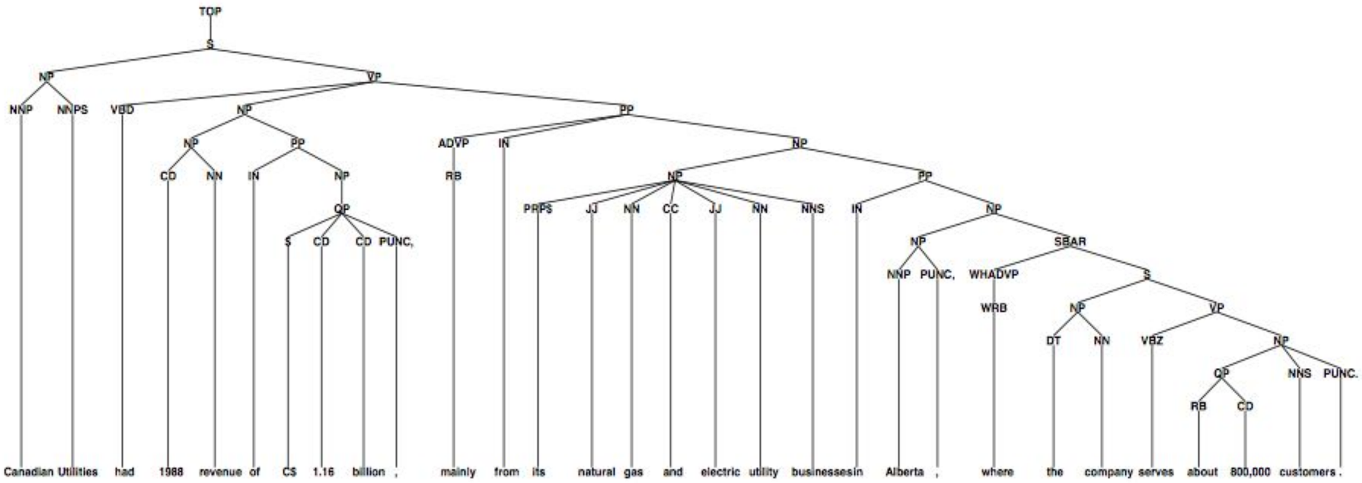
- ((()))   ()(())   ()()()   (())()   (()())

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

Catalan numbers

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

A typical tree from a standard dataset (Penn treebank WSJ)



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 .

[from Michael Collins slides]

Syntactic Ambiguities I

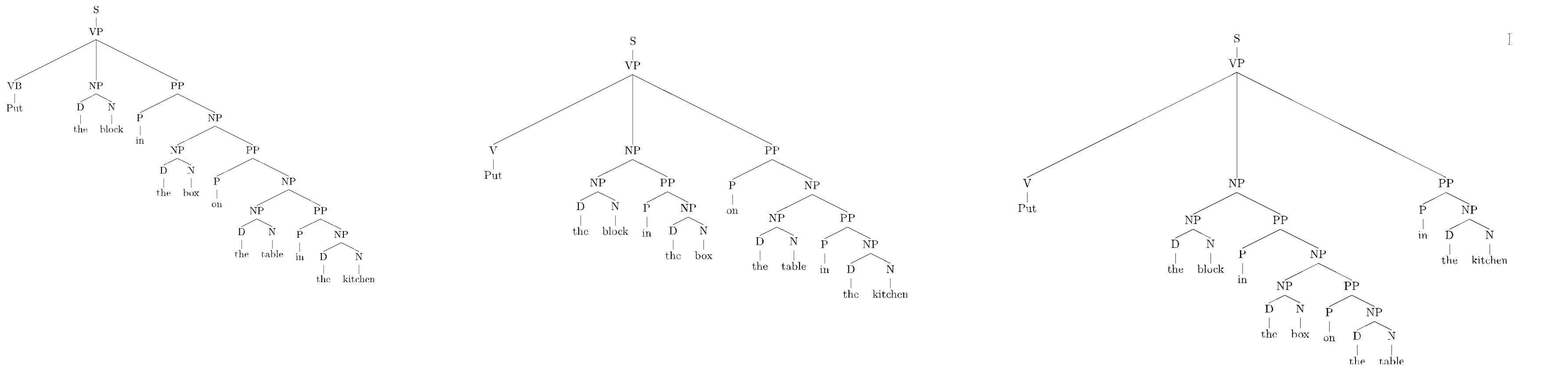
- Prepositional phrases:
  - They cooked the beans in the pot on the stove with handles.
- Particle vs. preposition:
  - The puppy tore up the staircase.
- Complement structures
  - The tourists objected to the guide that they couldn't hear.
    She knows you like the back of her hand.
- Gerund vs. participial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

Syntactic Ambiguities II

- Modifier scope within NPs
  - impractical design requirements
    plastic cup holder
- Multiple gap constructions
  - The chicken is ready to eat.
    The contractors are rich enough to sue.
- Coordination scope:
  - Small rats and mice can squeeze into holes or cracks in the wall.

How to Deal with Ambiguity?

- We want to score all the derivations to encode how plausible they are



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

Probabilistic Context Free Grammar (PCFG)

Probabilistic Context-Free Grammars

- A context-free grammar is a 4-tuple <N, T, S, R>
  - N : the set of non-terminals
    - **Phrasal categories**: S, NP, VP, ADJP, etc.
    - **Parts-of-speech** (pre-terminals): NN, JJ, DT, VB
  - T : the set of terminals (the words)
  - S : the start symbol
    - Often written as ROOT or TOP
    - Not usually the sentence non-terminal S
  - R : the set of rules
    - Of the form $X \rightarrow Y_1\ Y_2 \ldots Y_k$, with $X, Y_i \in N$
    - Examples: $S \rightarrow NP\ VP$, $VP \rightarrow VP\ CC\ VP$
    - Also called rewrites, productions, or local trees

- A PCFG adds:
  - A top-down production probability per rule $P(Y_1\ Y_2 \ldots Y_k \mid X)$

PCFGs

Associate probabilities with the rules : $p(X \rightarrow \alpha)$

$$\forall \;\; X \rightarrow \alpha \in R : \quad 0 \leq p(X \rightarrow \alpha) \leq 1$$

$$\forall X \in N : \quad \sum_{\alpha : X \rightarrow \alpha \in R} p(X \rightarrow \alpha) = 1$$

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

$S \rightarrow NP \;\; VP$	1.0	(NP A girl) (VP ate a sandwich)
$VP \rightarrow V$	0.2	
$VP \rightarrow V \;\; NP$	0.4	(VP ate) (NP a sandwich)
$VP \rightarrow VP \;\; PP$	0.4	(VP saw a girl) (PP with …)
$NP \rightarrow NP \;\; PP$	0.3	(NP a girl) (PP with ….)
$NP \rightarrow D \;\; N$	0.5	(D a) (N sandwich)
$NP \rightarrow PN$	0.2	
$PP \rightarrow P \;\; NP$	1.0	(P with) (NP with a sandwich)

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

PCFGs

$S \rightarrow NP\ \ VP$ 1.0

$VP \rightarrow V$ 0.2
$VP \rightarrow V\ \ NP$ 0.4
$VP \rightarrow VP\ \ PP$ 0.4

$NP \rightarrow NP\ \ PP$ 0.3
$NP \rightarrow D\ \ N$ 0.5
$NP \rightarrow PN$ 0.2

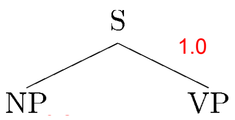
$PP \rightarrow P\ \ NP$ 1.0

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

S

$p(T) =$

PCFGs



$S \rightarrow NP\ \ VP$ 1.0

$VP \rightarrow V$ 0.2
$VP \rightarrow V\ \ NP$ 0.4
$VP \rightarrow VP\ \ PP$ 0.4

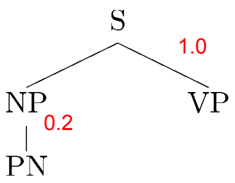
$NP \rightarrow NP\ \ PP$ 0.3
$NP \rightarrow D\ \ N$ 0.5
$NP \rightarrow PN$ 0.2

$PP \rightarrow P\ \ NP$ 1.0

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

$$p(T) = 1.0 \times$$

PCFGs



$S \to NP\ \ VP$ 1.0

$VP \to V$ 0.2
$VP \to V\ \ NP$ 0.4
$VP \to VP\ \ PP$ 0.4

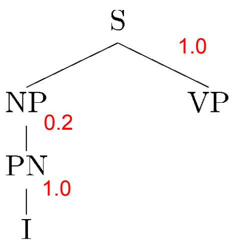
$NP \to NP\ \ PP$ 0.3
$NP \to D\ \ N$ 0.5
$NP \to PN$ 0.2

$PP \to P\ \ NP$ 1.0

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

$$p(T) = 1.0 \times 0.2 \times$$

PCFGs



$S \rightarrow NP \;\; VP$ 1.0

$VP \rightarrow V$ 0.2

$VP \rightarrow V \;\; NP$ 0.4

$VP \rightarrow VP \;\; PP$ 0.4

$NP \rightarrow NP \;\; PP$ 0.3

$NP \rightarrow D \;\; N$ 0.5

$NP \rightarrow PN$ 0.2

$PP \rightarrow P \;\; NP$ 1.0

$N \rightarrow girl$ 0.2

$N \rightarrow telescope$ 0.7

$N \rightarrow sandwich$ 0.1

$PN \rightarrow I$ 1.0

$V \rightarrow saw$ 0.5

$V \rightarrow ate$ 0.5

$P \rightarrow with$ 0.6

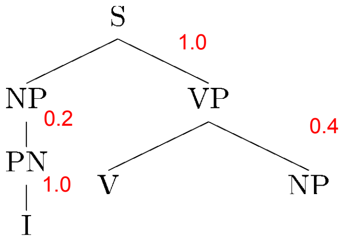
$P \rightarrow in$ 0.4

$D \rightarrow a$ 0.3

$D \rightarrow the$ 0.7

$$p(T) = 1.0 \times 0.2 \times 1.0 \times$$

PCFGs



$S \rightarrow NP \ \ VP$ 1.0

$VP \rightarrow V$ 0.2

$VP \rightarrow V \ \ NP$ 0.4

$VP \rightarrow VP \ \ PP$ 0.4

$NP \rightarrow NP \ \ PP$ 0.3

$NP \rightarrow D \ \ N$ 0.5

$NP \rightarrow PN$ 0.2

$PP \rightarrow P \ \ NP$ 1.0

$N \rightarrow girl$ 0.2

$N \rightarrow telescope$ 0.7

$N \rightarrow sandwich$ 0.1

$PN \rightarrow I$ 1.0

$V \rightarrow saw$ 0.5

$V \rightarrow ate$ 0.5

$P \rightarrow with$ 0.6

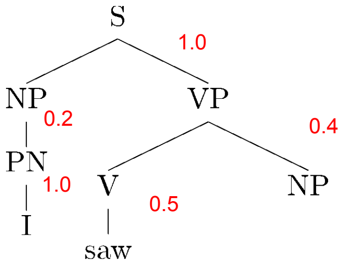
$P \rightarrow in$ 0.4

$D \rightarrow a$ 0.3

$D \rightarrow the$ 0.7

$$p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times$$

PCFGs



$S \rightarrow NP\ \ VP$ 1.0

$VP \rightarrow V$ 0.2

$VP \rightarrow V\ \ NP$ 0.4

$VP \rightarrow VP\ \ PP$ 0.4

$NP \rightarrow NP\ \ PP$ 0.3

$NP \rightarrow D\ \ N$ 0.5

$NP \rightarrow PN$ 0.2

$PP \rightarrow P\ \ NP$ 1.0

$N \rightarrow girl$ 0.2

$N \rightarrow telescope$ 0.7

$N \rightarrow sandwich$ 0.1

$PN \rightarrow I$ 1.0

$V \rightarrow saw$ 0.5

$V \rightarrow ate$ 0.5

$P \rightarrow with$ 0.6

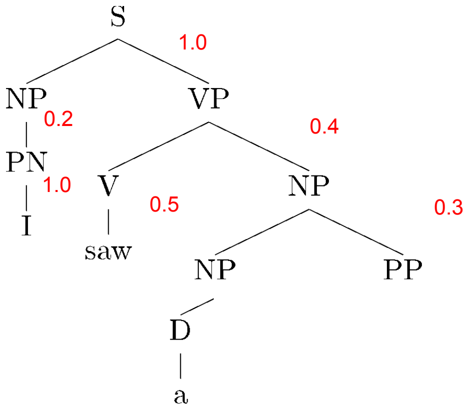
$P \rightarrow in$ 0.4

$D \rightarrow a$ 0.3

$D \rightarrow the$ 0.7

$$p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times$$

PCFGs



$S \rightarrow NP\ \ VP$ 1.0

$VP \rightarrow V$ 0.2

$VP \rightarrow V\ \ NP$ 0.4

$VP \rightarrow VP\ \ PP$ 0.4

$NP \rightarrow NP\ \ PP$ 0.3

$NP \rightarrow D\ \ N$ 0.5

$NP \rightarrow PN$ 0.2

$PP \rightarrow P\ \ NP$ 1.0

$N \rightarrow girl$ 0.2

$N \rightarrow telescope$ 0.7

$N \rightarrow sandwich$ 0.1

$PN \rightarrow I$ 1.0

$V \rightarrow saw$ 0.5

$V \rightarrow ate$ 0.5

$P \rightarrow with$ 0.6

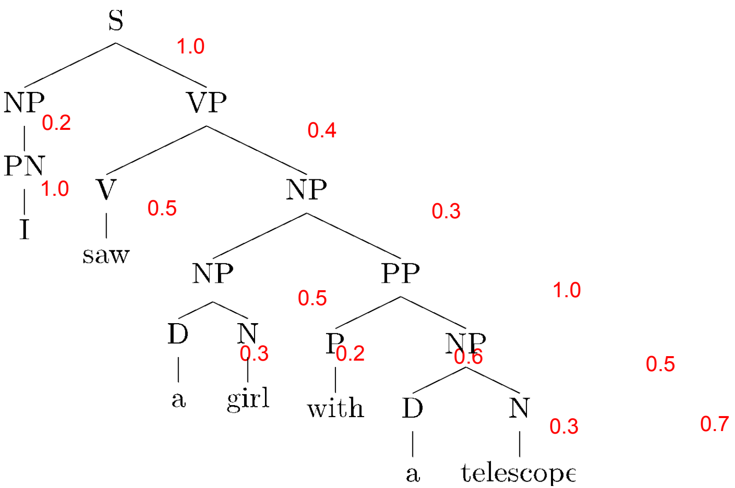
$P \rightarrow in$ 0.4

$D \rightarrow a$ 0.3

$D \rightarrow the$ 0.7

$$p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times$$

PCFGs



$S \rightarrow NP \;\; VP$ 1.0

$VP \rightarrow V$ 0.2

$VP \rightarrow V \;\; NP$ 0.4

$VP \rightarrow VP \;\; PP$ 0.4

$NP \rightarrow NP \;\; PP$ 0.3

$NP \rightarrow D \;\; N$ 0.5

$NP \rightarrow PN$ 0.2

$PP \rightarrow P \;\; NP$ 1.0

$N \rightarrow girl$ 0.2

$N \rightarrow telescope$ 0.7

$N \rightarrow sandwich$ 0.1

$PN \rightarrow I$ 1.0

$V \rightarrow saw$ 0.5

$V \rightarrow ate$ 0.5

$P \rightarrow with$ 0.6

$P \rightarrow in$ 0.4

$D \rightarrow a$ 0.3

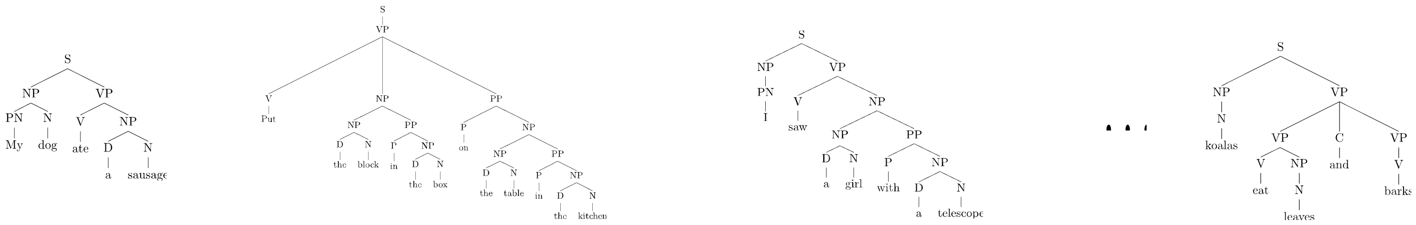
$D \rightarrow the$ 0.7

$$p(T) = 1.0 \times 0.2 \times 1.0 \times 0.4 \times 0.5 \times 0.3 \times$$
$$0.5 \times 0.3 \times 0.2 \times 1.0 \times 0.6 \times 0.5 \times 0.3 \times 0.7 \;\; = 2.26 \times 10^{-5}$$

PCFG Estimation

ML estimation

- A treebank: a collection sentences annotated with constituent trees



- An estimated probability of a rule (maximum likelihood estimates)

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

The number of times the rule used in the corpus

The number of times the nonterminal X appears in the treebank

- Smoothing is helpful
  - Especially important for preterminal rules

Parsing evaluation

Parsing evaluation

- Intrinsic evaluation:
  - Automatic: evaluate against annotation provided by human experts (gold standard) according to some predefined measure
  - Manual: … according to human judgment

- Extrinsic evaluation: score syntactic representation by comparing how well a system using this representation performs on some task
  - E.g., use syntactic representation as input for a semantic analyzer and compare results of the analyzer using syntax predicted by different parsers.

Standard evaluation setting in parsing

- Automatic intrinsic evaluation is used: parsers are evaluated against gold standard by provided by linguists
  - There is a standard split into the parts:
    - training set: used for estimation of model parameters
    - development set: used for tuning the model (initial experiments)
    - test set: final experiments to compare against previous work

Automatic evaluation of constituent parsers

- Exact match: percentage of trees predicted correctly
- Bracket score: scores how well individual phrases (and their boundaries) are identified

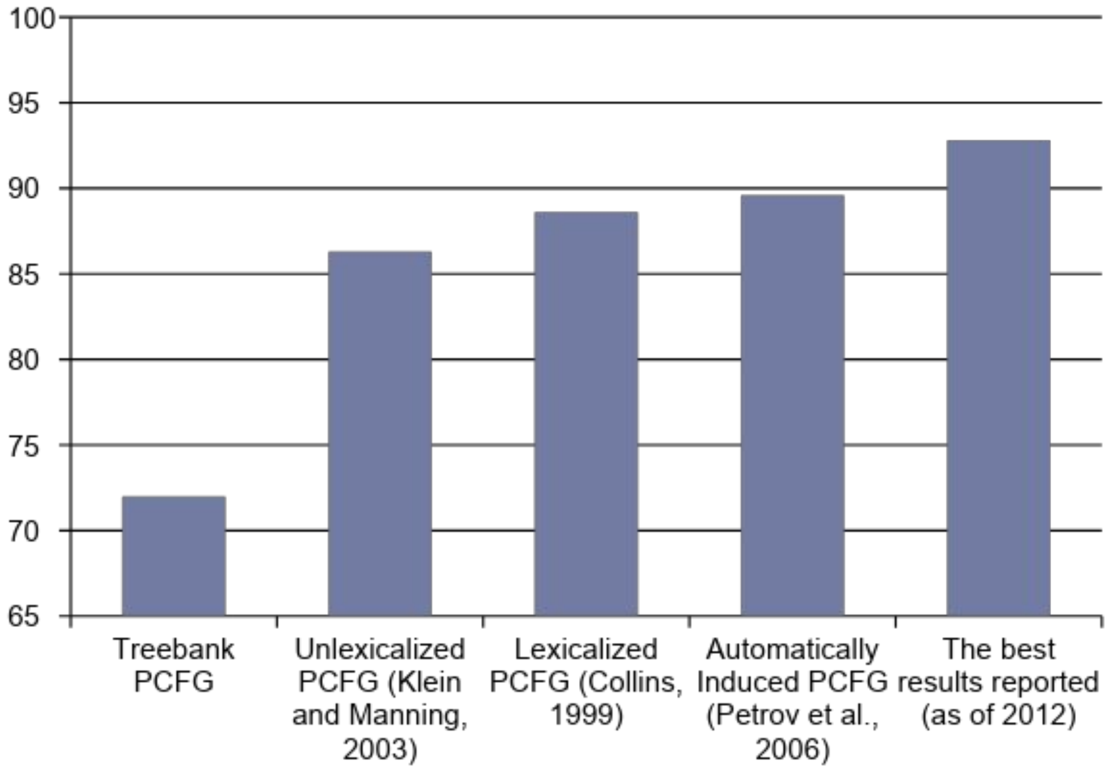
The most standard measure; we will focus on it

Brackets scores

Subtree signatures for CKY

- The most standard score is bracket score
- It regards a tree as a collection of brackets: $[min, max, C]$
- The set of brackets predicted by a parser is compared against the set of brackets in the tree annotated by a linguist
- Precision, recall and F1 are used as scores

Preview: F1 bracket score



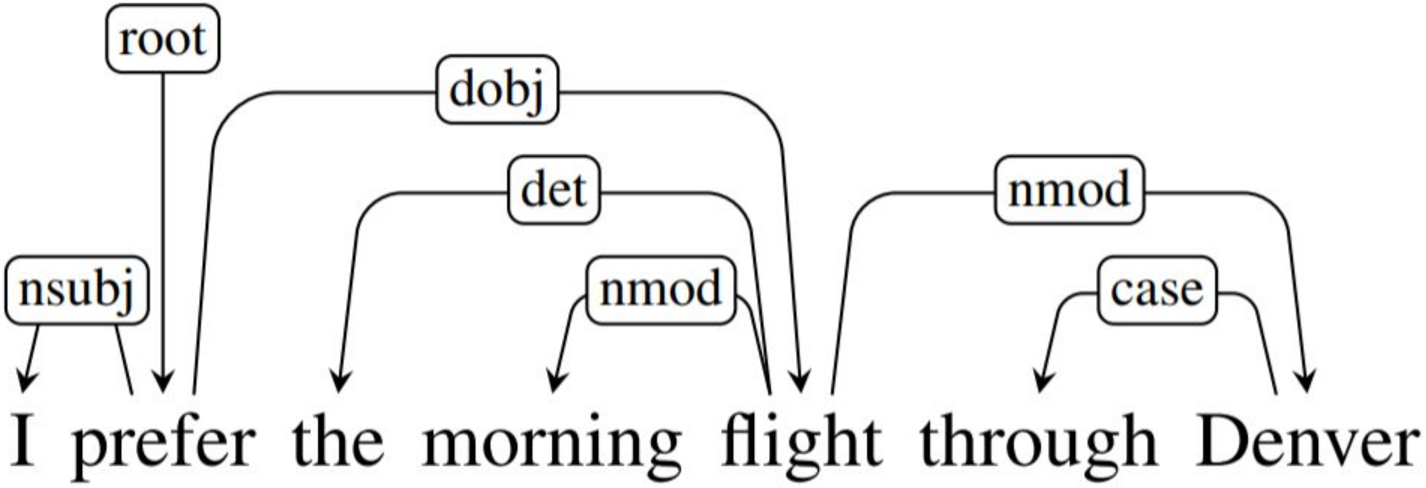
CKY Parsing

- Dynamic programming algorithm
- Not covered in lectures but see slides from the previous lecture if you are interested in learning more

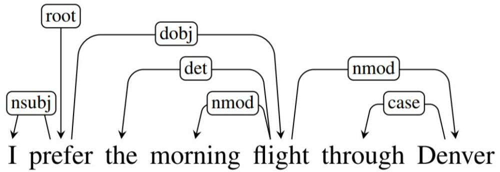
Dependency Treebanks

- the major English dependency treebanks converted from the WSJ sections of the PTB (Marcus et al., 1993)
- OntoNotes project (Hovy et al. 2006, Weischedel et al. 2011) adds conversational telephone speech, weblogs, usenet newsgroups, broadcast, and talk shows in English, Chinese and Arabic
- annotated dependency treebanks created for morphologically rich languages such as Czech, Hindi and Finnish, eg Prague Dependency Treebank (Bejcek et al., 2013)
- http://universaldependencies.org/
  - 150 treebanks, 90 languages

Dependency representation



Dependency representation



- A dependency structure can be defined as a directed graph G, consisting of
  - a set V of nodes – vertices, *words, punctuation, morphemes*
  - a set A of arcs – directed edges,
  - a linear precedence order < on V (word order).
- Labeled graphs
  - nodes in V are labeled with word forms (and annotation).
  - arcs in A are labeled with dependency types
  - $L = \{l_1, \ldots, l_{|L|}\}$ is the set of permissible arc labels;
  - Every arc in A is a triple (i,j,k), representing a dependency from w_i to w_j with label l_k.

Conversion from constituency to dependency

- Xia and Palmer (2001)
  - mark the head child of each node in a phrase structure, using the appropriate head rules
  - make the head of each non-head child depend on the head of the head-child

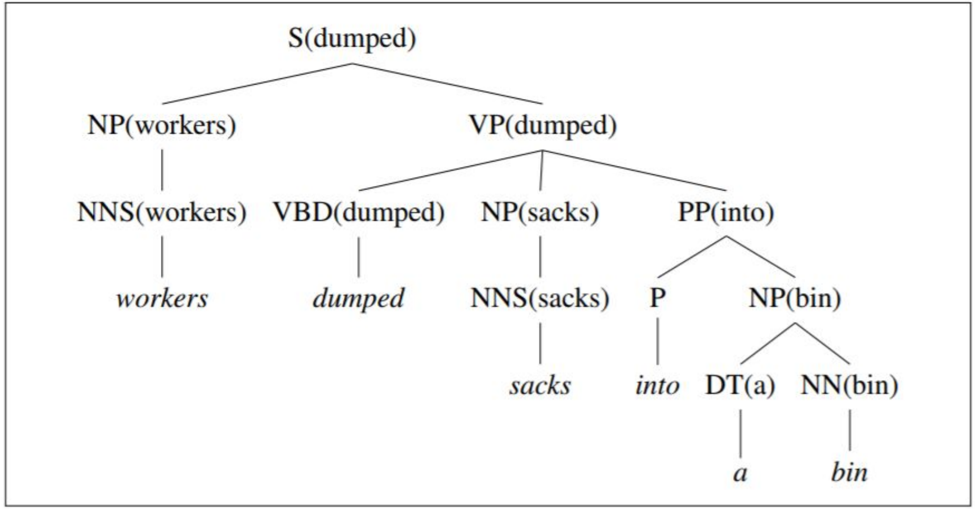


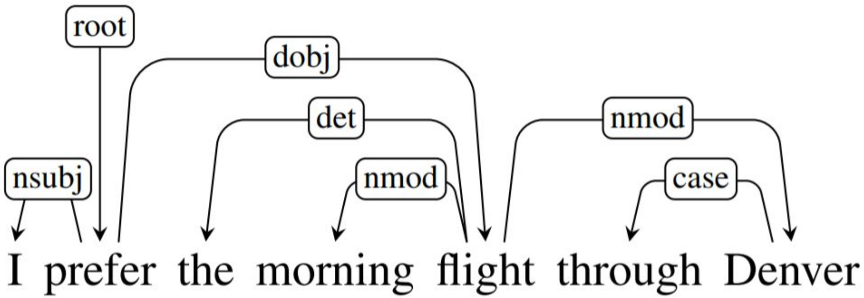
Figure 10.11 A lexicalized tree from Collins (1999).

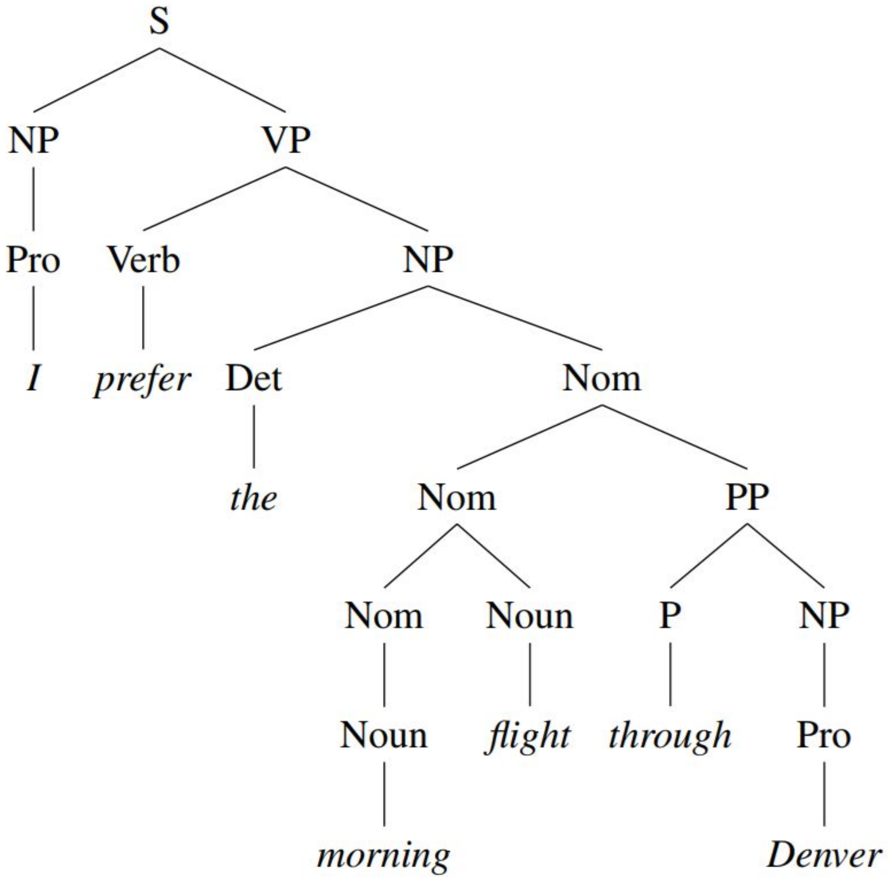
Dependency vs Constituency

- Dependency structures explicitly represent
  - head-dependent relations (directed arcs),
  - functional categories (arc labels)
  - possibly some structural categories (parts of speech)

- Phrase (aka constituent) structures explicitly represent
  - phrases (nonterminal nodes),
  - structural categories (nonterminal labels)

Dependency vs Constituency trees





Parsing Languages with Flexible Word Order

I prefer the morning flight through Denver

Я предпочитаю утренний перелет через Денвер

Languages with free word order

I prefer the morning flight through Denver

Я предпочитаю утренний перелет через Денвер

Я предпочитаю через Денвер утренний перелет

Утренний перелет я предпочитаю через Денвер

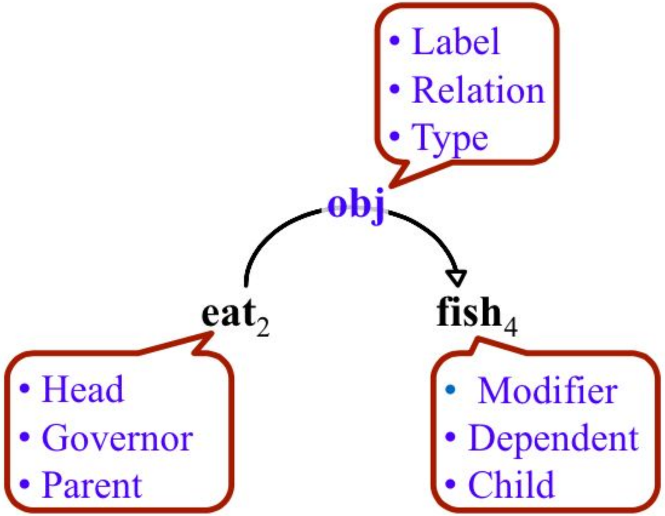
Перелет утренний я предпочитаю через Денвер

Через Денвер я предпочитаю утренний перелет

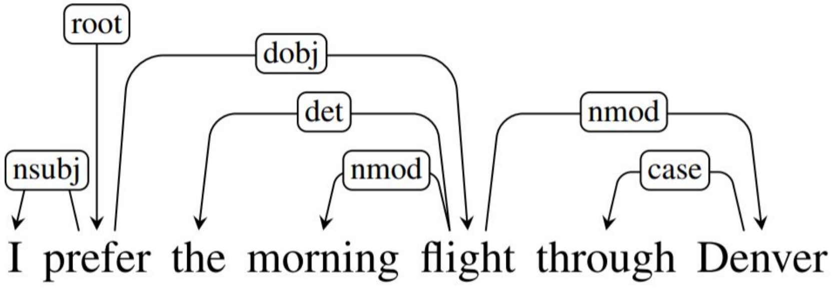
Я через Денвер предпочитаю утренний перелет

...

Dependency relations



Types of relationships



- The clausal relations NSUBJ and DOBJ identify the arguments: the subject and direct object of the predicate *cancel*

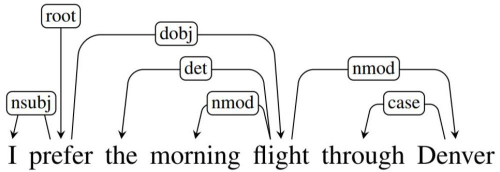
- The NMOD, DET, and CASE relations denote modifiers of the nouns *flights* and *Houston*.

Grammatical functions

Clausal Argument Relations	Description
NSUBJ	Nominal subject
DOBJ	Direct object
IOBJ	Indirect object
CCOMP	Clausal complement
XCOMP	Open clausal complement
Nominal Modifier Relations	**Description**
NMOD	Nominal modifier
AMOD	Adjectival modifier
NUMMOD	Numeric modifier
APPOS	Appositional modifier
DET	Determiner
CASE	Prepositions, postpositions and other case markers
Other Notable Relations	**Description**
CONJ	Conjunct
CC	Coordinating conjunction

Figure 13.2 Selected dependency relations from the Universal Dependency set. (de Marneffe et al., 2014)

Dependency Constraints

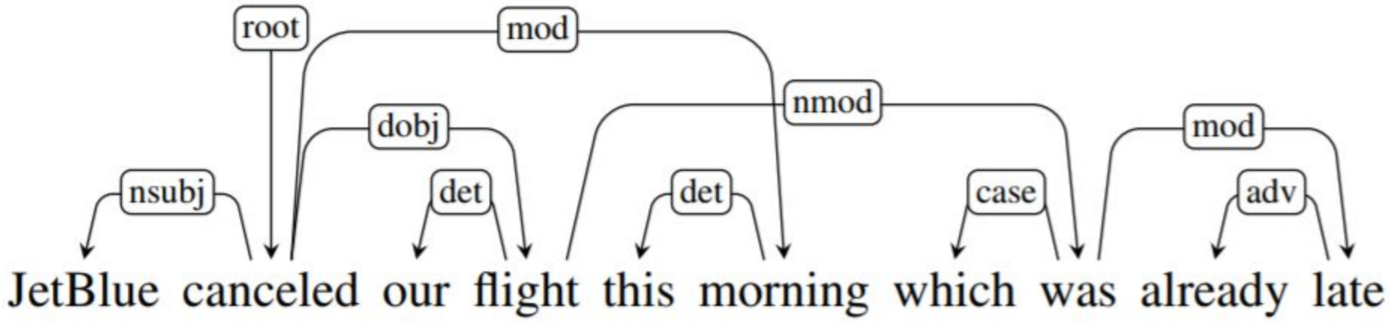


- Syntactic structure is complete (connectedness)
  - connectedness can be enforced by adding a special root node
- Syntactic structure is hierarchical (acyclicity)
  - there is a unique pass from the root to each vertex
- Every word has at most one syntactic head (single-head constraint)
  - except root that does not have incoming arcs

This makes the dependencies a tree

Projectivity

- Projective parse
  - arcs don't cross each other
  - mostly true for English
- Non-projective structures are needed to account for
  - long-distance dependencies
  - flexible word order



Projectivity

- Dependency grammars do not normally assume that all dependency-trees are projective, because some linguistic phenomena can only be achieved using non-projective trees.
- But a lot of parsers assume that the output trees are projective
- Reasons
  - conversion from constituency to dependency
  - the most widely used families of parsing algorithms impose projectivity

Non-Projective Statistics

Arabic: 11.2 %

Bulgarian: 5.4 %

Chinese: 0.0 %

Czech: 23.2 %

Danish: 15.6 %

Dutch: 36.4 %

German: 27.8 %

Japanese: 5.3 %

Polish: 18.9 %

Slovene: 22.2 %

Spanish 1.7 %

Swedish: 9.8 %

Turkish: 11.6 %

English: 0.0% (SD: 0.1%)

Percentage of non-projective trees for some treebanks of the CoNLL-X Shared Task and English.

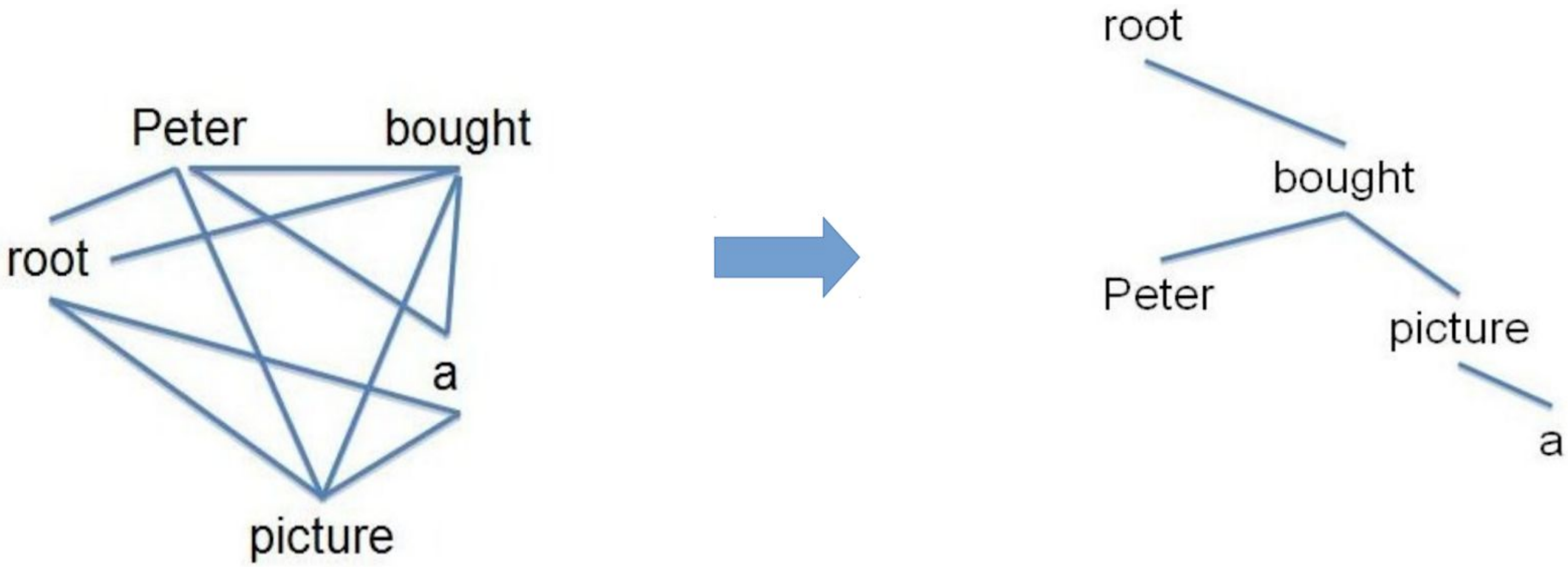
Parsing problem

The parsing problem for a dependency parser is to find the optimal dependency tree ***y*** given an input sentence ***x***

This amounts to assigning a syntactic head ***i*** and a label ***l*** to every node ***j*** corresponding to a word x_j in such a way that the resulting graph is a tree rooted at the node 0

Parsing problem

- This is equivalent to finding a spanning tree in the complete graph containing all possible arcs



Parsing algorithms

- Transition based
  - greedy choice of local transitions guided by a good classifier
  - deterministic
  - MaltParser (Nivre et al. 2008)
- Graph based
  - Minimum Spanning Tree for a sentence
  - McDonald et al.'s (2005) MSTParser
  - Martins et al.'s (2009) Turbo Parser