COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

Sequence Labeling & Syntax

CMSC 470

Marine Carpuat

Recap: We know how to perform POS tagging with structured perceptron

- An example of sequence labeling tasks
- Requires a predefined set of POS tags
  - Penn Treebank commonly used for English
  - Encodes some distinctions and not others
- Given annotated examples, we can address sequence labeling with multiclass perceptron
  - but computing the argmax naively is expensive
  - constraints on the feature definition make efficient algorithms possible
    - E.g, Viterbi algorithm

Sequence labeling tasks

Beyond POS tagging

Many NLP tasks can be framed as sequence labeling

- Information Extraction: detecting named entities
  - E.g., names of people, organizations, locations

"Brendan Iribe, a co-founder of Oculus VR and a prominent University of Maryland donor, is leaving Facebook four years after it purchased his company."

http://www.dbknews.com/2018/10/24/brendan-iribe-facebook-leaves-oculus-vr-umd-computer-science/

Many NLP tasks can be framed as sequence labeling

x = [Brendan, Iribe, “,”, a, co-founder, of, Oculus, VR, and, a, prominent, University, of, Maryland, donor, “,”, is, leaving, Facebook, four, years, after, it, purchased, his, company, “.”]

y = [B-PER, I-PER, O, O, O, O, B-ORG, I-ORG, O, O, O,B-ORG, I-ORG, I-ORG, O, O, O,B-ORG, O, O, O, O, O, O, O, O]

“BIO” labeling scheme for named entity recognition

Many NLP tasks can be framed as sequence labeling

- The same kind of BIO scheme can be used to tag other spans of text
    - Syntactic analysis: detecting noun phrase and verb phrases
    - Semantic roles: detecting semantic roles (who did what to whom)

Many NLP tasks can be framed as sequence labeling

- Other sequence labeling tasks
  - Language identification in code-switched text
    "Ulikuwa ukiongea a lot of nonsense." (Swahili/English)
  - Metaphor detection
    "he **swam** in a **sea** of diamonds"
    "authority is a **chair**, it needs **legs** to **stand**"
    "in Washington, people change **dance partners** frequently, but not the **dance**"
  - …

Other algorithms for solving the argmax problem

Structured perceptron can be used for other structures than sequences

- The Viterbi algorithm we’ve seen is specific to sequences
  - Other argmax algorithms necessary for other structures (e.g. trees)

- Integer Linear Programming provides a general framework for solving the argmax problem

Argmax problem as an Integer Linear Program

- An integer linear program (ILP) is an optimization problem of the form

$$\max_{\boldsymbol{z}} \quad \boldsymbol{a} \cdot \boldsymbol{z} \quad \text{subj. to} \quad \text{linear constraints on } \boldsymbol{z}$$

  - For a fixed vector a
  - Example of integer constraint: $z_3 \in \{0, 1\}$

- Well-engineered solvers exist
  - e.g, Gurobi
  - Useful for prototyping
  - But general not as efficient as dynamic programming

Casting sequence labeling with Markov features as an ILP

- Step 1: Define variables z as binary indicator variables which encode an output sequence y

$$z_{l,k',k} = \mathbf{1}[\text{label } l \text{ is } k \text{ and label } l-1 \text{ is } k']$$

- Step 2: Construct the linear objective function

$$a_{l,k',k} = \boldsymbol{w} \cdot \phi_l(\boldsymbol{x}, \langle \ldots, k', k \rangle)$$

Casting sequence labeling with Markov features as an ILP

- Step 3: Define constraints to ensure a well-formed solution
  - Z's should be binary: for all l, k', k

$$z_{l,k',k} \in \{0,1\}$$

  - For a given position l, there is exactly one active z

$$\sum_k \sum_{k'} z_{l,k',k} = 1 \text{ for all } l$$

  - The z's are internally consistent

$$\sum_{k'} z_{l,k',k} = \sum_{k''} z_{l+1,k,k''} \text{ for all } l, k$$

What you should know

- POS tagging as an example of sequence labeling task
- Requires a predefined set of POS tags
  - Penn Treebank commonly used for English
  - Encodes some distinctions and not others
- How to train and predict with the structured perceptron
  - constraints on feature structure make efficient algorithms possible
  - Unary and markov features => Viterbi algorithm
- Extensions:
  - How to frame other problems as sequence labeling tasks
  - Viterbi is not the only way to solve the argmax: Integer Linear Programming is a more general solution

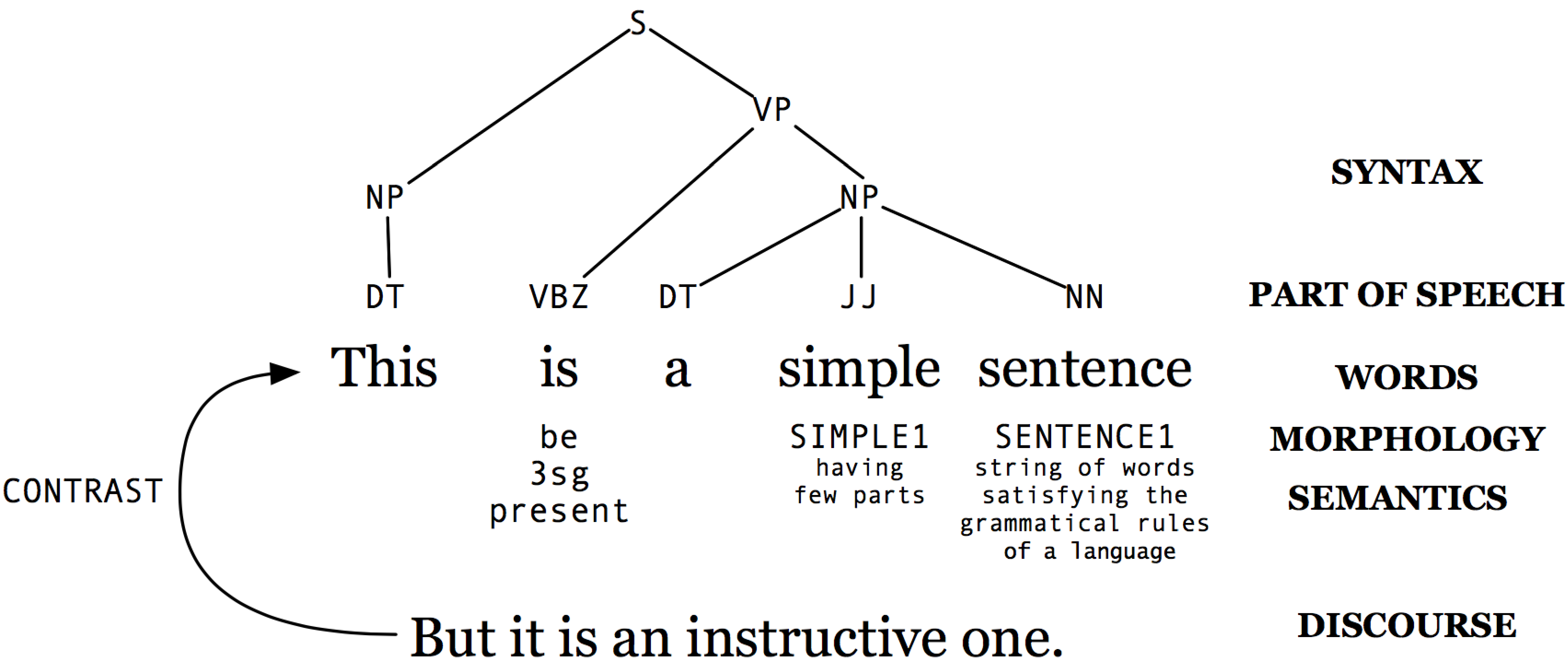
COMPUTER SCIENCE
UNIVERSITY OF MARYLAND

Syntax, Grammars & Parsing

CMSC 470

Marine Carpuat

Fig credits: Joakim Nivre, Dan Jurafsky & James Martin


S
VP
NP
NP
DT
VBZ
DT
JJ
NN
This is a simple sentence
be
3sg
present
SIMPLE1
having few parts
SENTENCE1
string of words satisfying the grammatical rules of a language
CONTRAST
But it is an instructive one.
SYNTAX
PART OF SPEECH
WORDS
MORPHOLOGY
SEMANTICS
DISCOURSE

Syntax & Grammar

- Syntax
  - From Greek syntaxis, meaning “setting out together”
  - refers to the way words are arranged together.

- Grammar
  - Set of structural rules governing composition of clauses, phrases, and words in any given natural language
  - Descriptive, not prescriptive
  - Panini’s grammar of Sanskrit ~2000 years ago

Syntax and Grammar

- Goal of syntactic theory
  - “explain how people combine words to form sentences and how children attain knowledge of sentence structure”

- Grammar
  - implicit knowledge of a native speaker
  - acquired without explicit instruction
  - minimally able to generate all and only the possible sentences of the language

[Philips, 2003]

Two views of syntactic structure

- Constituency (phrase structure)
  - Phrase structure organizes words in nested constituents

- Dependency structure
  - Shows which words depend on (modify or are arguments of) which on other words

Constituency

- Basic idea: groups of words act as a single unit

- Constituents form coherent classes that behave similarly
  - With respect to their internal structure: e.g., at the core of a noun phrase is a noun
  - With respect to other constituents: e.g., noun phrases generally occur before verbs

Constituency: Example

- The following are all noun phrases in English...

Harry the Horse	a high-class spot such as Mindy's
the Broadway coppers	the reason he comes into the Hot Box
they	three parties from Brooklyn

- Why?
  - They can all precede verbs
  - They can all be preposed/postposed
  - …

Grammars and Constituency

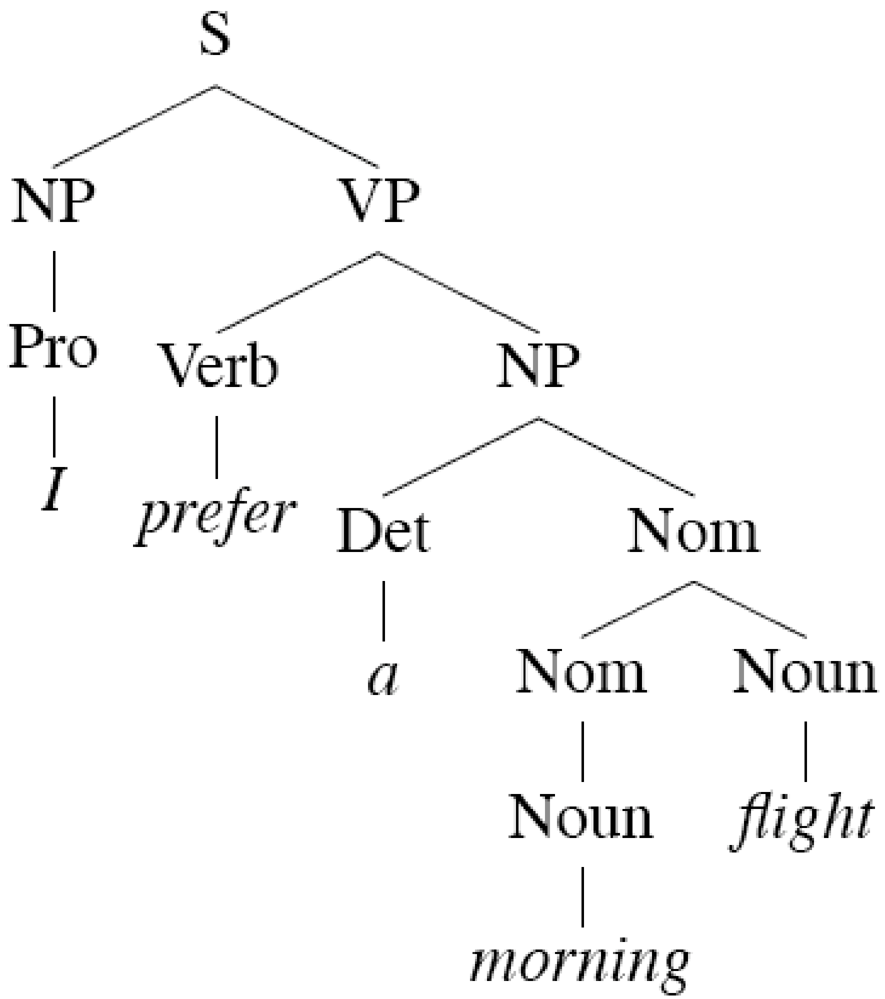
- For a particular language:
  - What are the “right” set of constituents?
  - What rules govern how they combine?

- Answer: not obvious and difficult
  - There are many different theories of grammar and competing analyses of the same data!

An Example Context-Free Grammar

Grammar Rules			Examples
S	$\rightarrow$	$NP\ VP$	I + want a morning flight
NP	$\rightarrow$	*Pronoun*	I
	|	*Proper-Noun*	Los Angeles
	|	*Det Nominal*	a + flight
Nominal	$\rightarrow$	*Nominal Noun*	morning + flight
	|	*Noun*	flights
VP	$\rightarrow$	*Verb*	do
	|	*Verb NP*	want + a flight
	|	*Verb NP PP*	leave + Boston + in the morning
	|	*Verb PP*	leaving + on Thursday
PP	$\rightarrow$	*Preposition NP*	from + Los Angeles

Parse Tree: Example

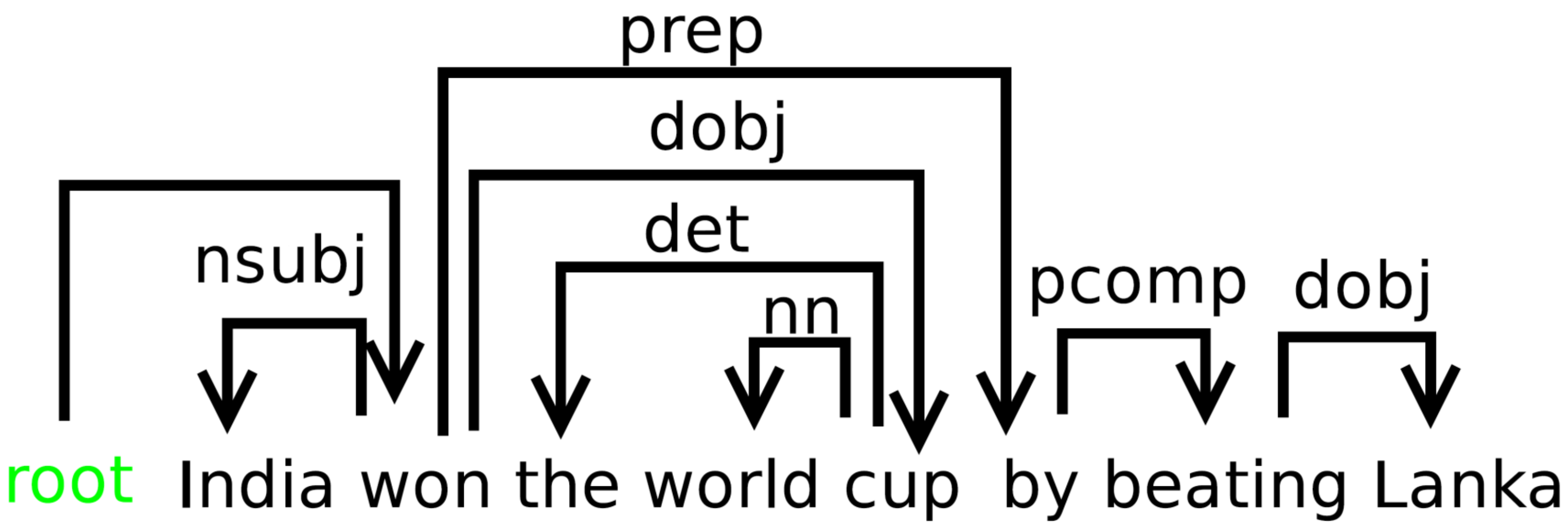


Dependency Grammars

- Context-Free Grammars focus on constituents
  - Non-terminals don't actually appear in the sentence

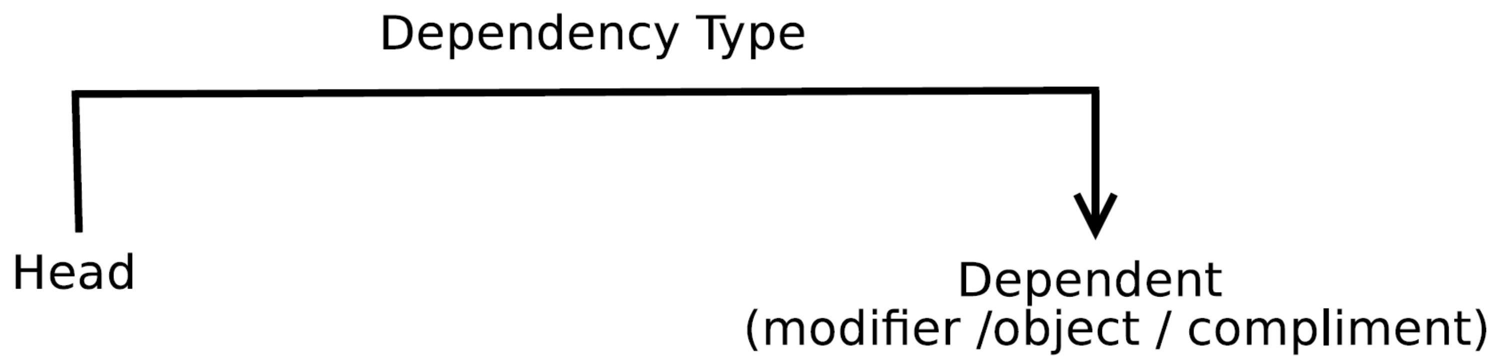
- In dependency grammar, a parse is a graph (usually a tree) where:
  - Nodes represent words
  - Edges represent dependency relations between words (typed or untyped, directed or undirected)

Example Dependency Parse

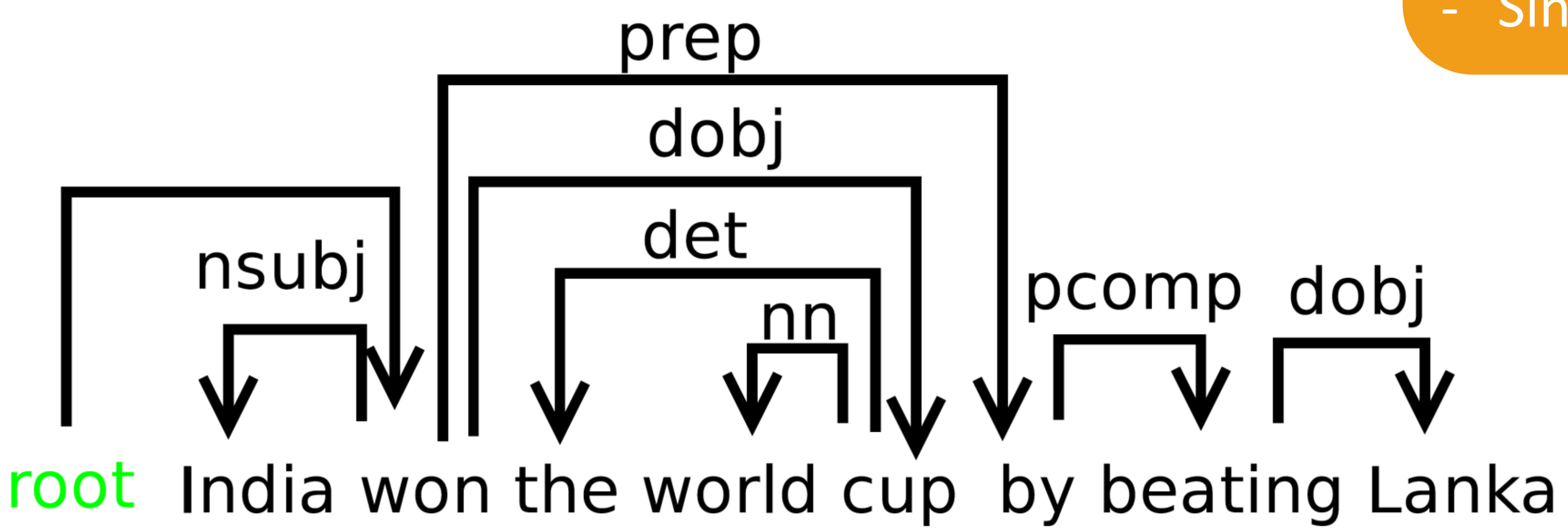


Dependency Grammars

- Syntactic structure = lexical items linked by binary asymmetrical relations called dependencies



Example Dependency Parse



Dependencies (usually) form a tree:
- Connected
- Acyclic
- Single-head

Dependency Relations

Argument Dependencies	Description
nsubj	nominal subject
csubj	clausal subject
dobj	direct object
iobj	indirect object
pobj	object of preposition
Modifier Dependencies	**Description**
tmod	temporal modifier
appos	appositional modifier
det	determiner
prep	prepositional modifier

Relation	Examples with *head* and **dependent**
NSUBJ	**United** *canceled* the flight.
DOBJ	United *diverted* the **flight** to Reno.
	We *booked* her the first **flight** to Miami.
IOBJ	We *booked* **her** the flight to Miami.
NMOD	We took the **morning** *flight*.
AMOD	Book the **cheapest** *flight*.
NUMMOD	Before the storm JetBlue canceled **1000** *flights*.
APPOS	*United*, a **unit** of UAL, matched the fares.
DET	**The** *flight* was canceled.
	Which *flight* was delayed?
CONJ	We *flew* to Denver and **drove** to Steamboat.
CC	We flew to Denver **and** *drove* to Steamboat.
CASE	Book the flight **through** *Houston*.

Figure 14.3 Examples of core Universal Dependency relations.

Universal Dependencies project

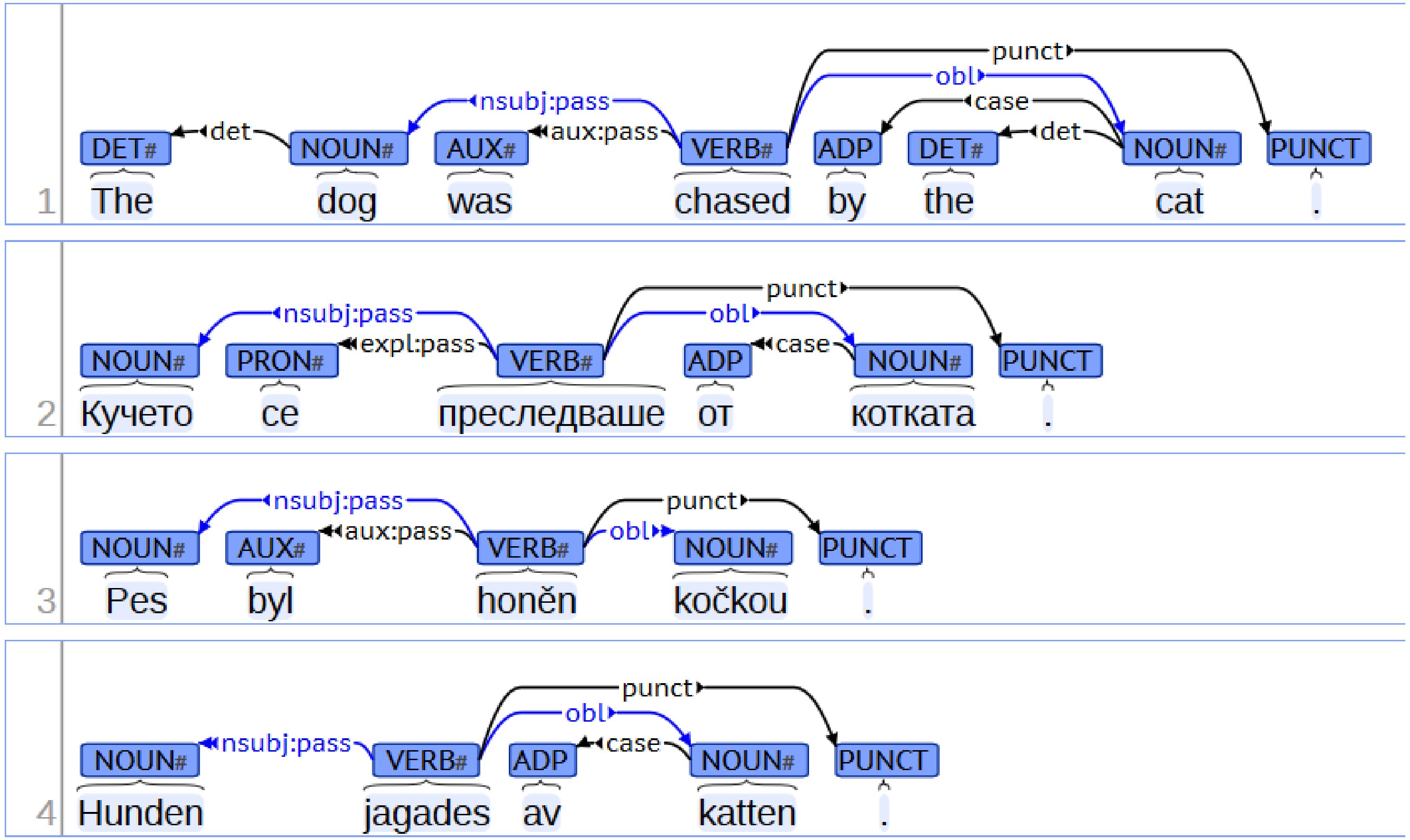
- Set of dependency relations that are
  - Linguistically motivated
  - Computationally useful
  - Cross-linguistically applicable

  [Nivre et al. 2016]

universaldependencies.org

Universal Dependencies Illustrated

Parallel examples for English, Bulgarian, Czech & Swedish



https://universaldependencies.org/introduction.html

What you should know

- Syntax vs. Grammar
- Two views of syntactic structures
  - Context-Free Grammar vs. Dependency grammars
  - Can be used to capture various facts about the structure of language (but not all!)
- Dependency grammars
  - Definition of dependency links: head, dependent
  - Annotate an example given a set of dependency types
- How syntactic analysis can be used to define NLP tasks or features

- Next: how can we predict syntactic parses?