Dependency Parsing

Feb 3, 2026

RIT Rochester Institute of Technology

Outline

RIT Rochester Institute of Technology

Outline

RIT Rochester Institute of Technology

Understanding linguistic structure

- So far we have focused on how NLP deals with word meanings.
- But language goes beyond individual words.
- **Linguistic structure** is equally important for capturing how words combine together to create meaning.

Understanding linguistic structure

A **grammar** is the system of rules that defines how linguistic structures are formed and how words relate to each other within a sentence.

Understanding linguistic structure

A **grammar** is the system of rules that defines how linguistic structures are formed and how words relate to each other within a sentence.

1. Part of Speech (POS)
2. Dependency grammar

1. Part-of-Speech (POS)

- A word's POS determines how words fit into a sentence.
- POS is also called lexical category.
- Main POS categories (in English):
  - **N**___: reindeer, game, government
  - **V**___: play, run, believe
  - **A**___: fun, beautiful
  - **A**___: well, heavily
  - **P**___: on, into
  - **A**___/**D**___: a, the, some
  - **C**___: and, or

How do we identify POS?

- Examples:
  - *This car is very interesting.*
  - *This car mooked fast.*
  - *This nony car mooked fast.*
- We usually identify POS by:
  - **Morphology:** how a word changes form (e.g., verbs mark tense: *play* $\rightarrow$ *play***ed**, sometimes irregularly: *go* $\rightarrow$ *went*)
  - **Distribution:** where a word appears in a sentence (e.g., nouns after articles, verbs after subjects)

RIT Rochester Institute of Technology

2. From words to phrases

- Words combine into constituents based on POS:
  - **the reindeer** = article + noun = noun phrase
  - **play games** = verb + noun phrase = verb phrase
- Constituents combine based on **phrasal category**:
  - **Noun Phrase** + **Verb Phrase** = Sentence

Structure over meaning

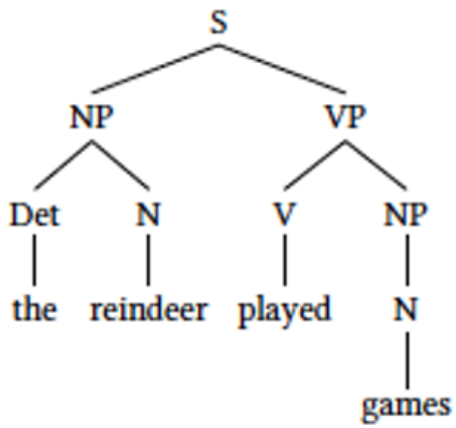
- Chomsky (1957): *"Colorless green ideas sleep furiously"*
- Nonsensical meaning, but:
  - Correct lexical and phrasal categories
  - Grammatically well-formed
- Syntax is about **structure**, not always meaning.

Understanding linguistic structure 1: Constituency

- **Constituency grammar**
- A linguistic theory that analyzes sentences as nested constituents (e.g., noun phrases, verb phrases).
- Also known as phrase structure grammar

Lexicon:
N → *reindeer, dragon, lunch, game, evening, morning*
V(trans) → *play, eat*
V(intrans) → *run, swim, dance*
Adj → *fun, beautiful, interesting*
Det → *the, a, some, many*
P → *for, in, to, at*

Phrase structure rules:
S → NP VP
VP → V(trans) NP
VP → V(intrans)
NP → Det (A*) N
NP → N
NP → NP PP
PP → P NP



RIT Rochester Institute of Technology

Frameworks for analyzing grammar

- Linguists formalize sentence structure using grammar frameworks:
  - Phrase Structure Grammar

Frameworks for analyzing grammar

- Linguists formalize sentence structure using grammar frameworks:
  - Phrase Structure Grammar (linguistics)

Frameworks for analyzing grammar

- Linguists formalize sentence structure using grammar frameworks:
  - Phrase Structure Grammar (linguistics)
  - **Dependency Grammar (widely used in NLP)**

Understanding linguistic structure 2: Dependency

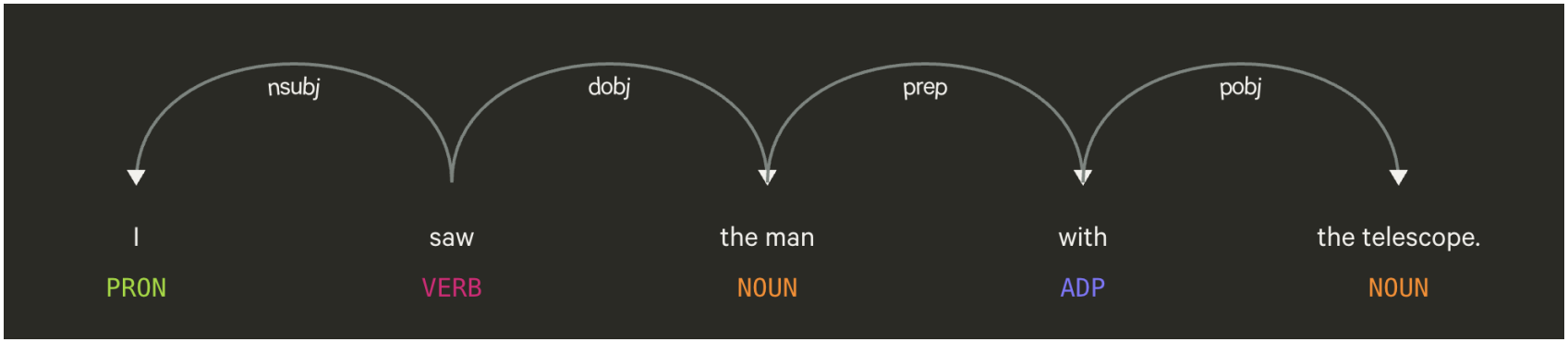
- Dependency grammar shows that syntactic structure consists of relationships between lexical items, normally binary asymmetric relations ("arrows") called **dependencies**.
- Dependency structure shows which words depend on (modify, attach to, or are arguments of) which other words.

Why do we need dependency structure?

e.g., *I saw the man with the telescope.*

Why do we need dependency structure?

e.g., *I saw the man with the telescope.*



- Readers/Listeners/NLP models need to work out what modifies (attaches to) what.

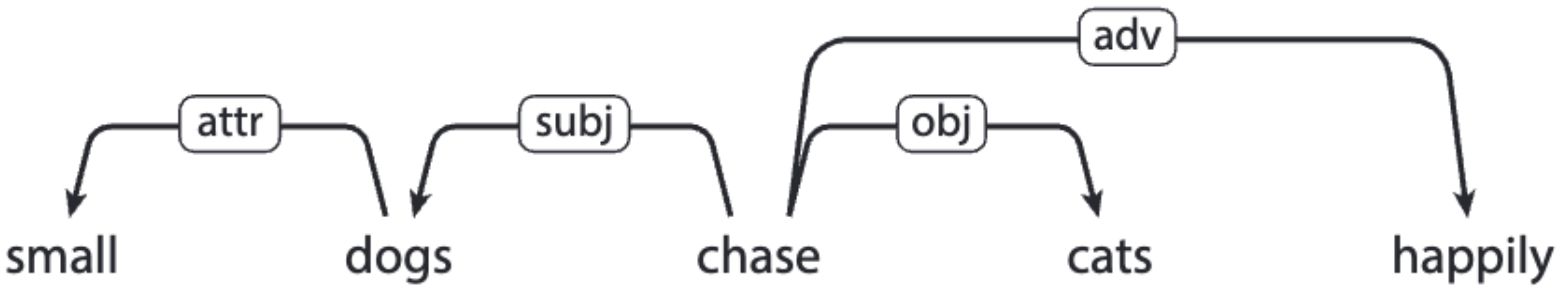
RIT Rochester Institute of Technology

Outline

RIT Rochester Institute of Technology

Dependency grammar and dependency structure

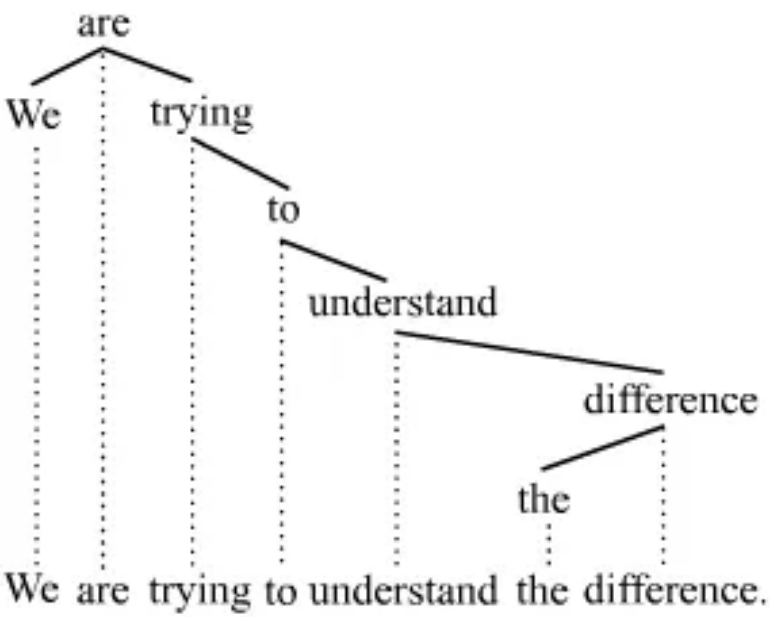
Dependency grammar shows that syntactic structure (of a sentence) consists of relations between lexical items, normally binary asymmetric relations ("arrows") called dependencies.



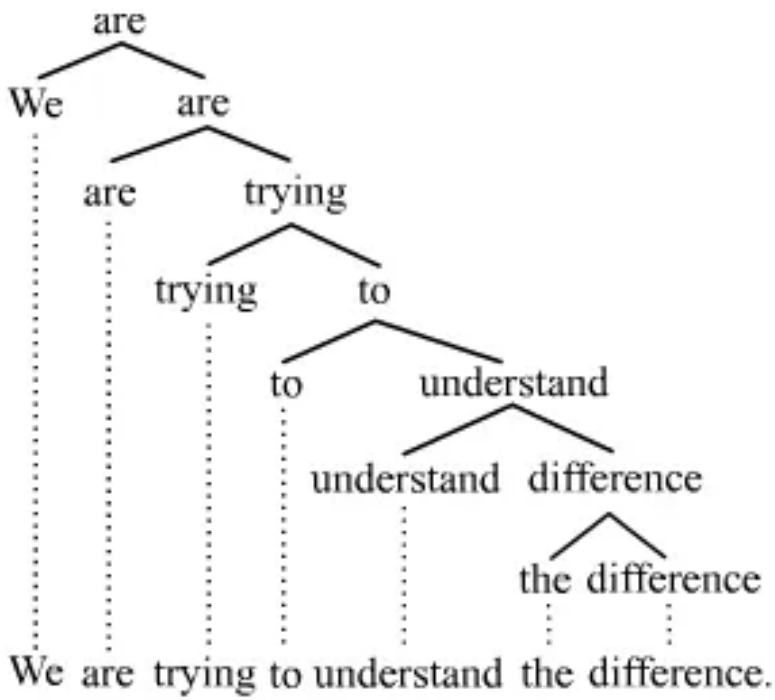
Sourced from: De Marneffe, M. C., & Nivre, J. (2019). *Dependency grammar*. *Annual Review of Linguistics*, 5(1), 197–218. Figure 1.

RIT Rochester Institute of Technology

Dependency grammar vs. Constituency parsing



Dependency



Constituency (BPS)

RIT Rochester Institute of Technology

The arise of annotated data & Universal Dependencies tree

What is a treebank? An annotated corpus that includes syntactic or morphological structure, often in the form of parse trees.

The arise of annotated data & Universal Dependencies tree

What is a treebank? An annotated corpus that includes syntactic or morphological structure, often in the form of parse trees.

Milestones in treebank development:

- *Brown corpus* (1967): First general-purpose corpus; part-of-speech (PoS) tagged in 1979
- *Lancaster-IBM treebank* (Late 1980s): One of the first syntactically annotated corpora
- *The Penn treebank* (Marcus et al., 1993): Influential constituency-based treebank for English
- **Universal Dependencies (UD)**: A multilingual, cross-linguistically consistent treebank project using dependency grammar

The rise of annotated data

Starting off, building a treebank seems a lot slower and less useful than writing a grammar (by hand)

But a treebank gives us many things:

- Reusability of the labor
  - Many parser, POS taggers, and built on it
  - Valuable resource for linguistics
- Broad coverage, not just a few intuitions
- Frequencies and distributional information
- A way to evaluate NLP systems (work as a benchmark for empirical science)

Presentation 1

Matthew & Valbona: de Marneffe et al. (2021). Universal Dependencies.

Outline

RIT Rochester Institute of Technology

Sources of information for dependency parsing

How do we build a **parser**, once we get the dependency information?

Methods of dependency parsing

There are several ways (including dynamic programming, graph algorithms, etc.) but we'll focus on the one method.

Presentation 2

Duvarakanath: Chen & Manning (2014). A Fast and Accurate Neural Dependency Parser.

Outline

RIT Rochester Institute of Technology

Evaluation

Gold Standard: Hand-annotated syntactic structure used for evaluating parser output.

Evaluation

Gold Standard: Hand-annotated syntactic structure used for evaluating parser output.
Metrics: (1) **UAS (Unlabeled Attachment Score):** Correct head only; (2) **LAS (Labeled Attachment Score):** Correct head and label

Evaluation

Gold Standard: Hand-annotated syntactic structure used for evaluating parser output.
Metrics: (1) **UAS (Unlabeled Attachment Score):** Correct head only; (2) **LAS (Labeled Attachment Score):** Correct head and label

Example:

Word	Gold Head	Gold Label	Pred Head	Pred Label
She	2	nsubj	2	nsubj
likes	0	root	0	root
chocolate	2	obj	2	nmod
very	4	advmod	4	advmod
much	2	advmod	4	advmod

Example:

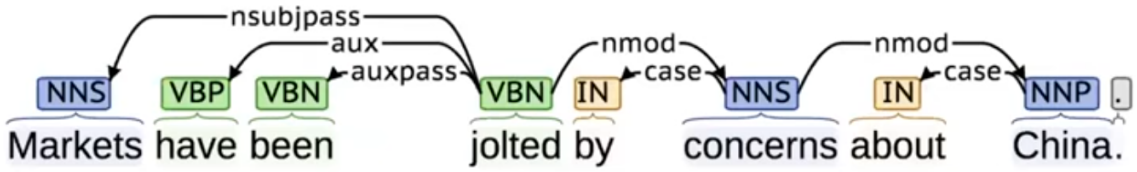
Word	Gold Head	Gold Label	Pred Head	Pred Label
She	2	nsubj	2	nsubj
likes	0	root	0	root
chocolate	2	obj	2	nmod
very	4	advmod	4	advmod
much	2	advmod	4	advmod

Evaluation:

- **Total dependencies:** 5
- **Correct heads (UAS):** 4 $\rightarrow$ **UAS = 4/5 = 80%**
- **Correct heads + labels (LAS):** 3 $\rightarrow$ **LAS = 3/5 = 60%**

Dependency parsing for sentence structure

- Chen and Manning (2014) showed that neural networks can accurately determine the structure of sentences, supporting meaning interpretation.



- It was the first simple, successful neural dependency parser.
- The dense representations (and non-linear classifier) let it outperform other greedy parsers in both accuracy and speed.
- This work was further developed and improved by others.

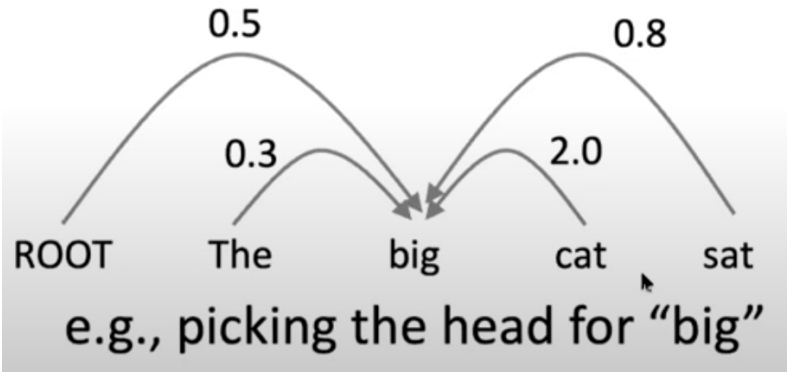
RIT Rochester Institute of Technology

Further developments

This work was further developed and improved by others, including in particular at Google.

Graph-based dependency parsers

- Compute a score for every possible dependency (choice of head) for each word
  - Doing this well requires more than just knowing two words
  - We need good "contextual" representations of each word token
- Repeat the same process for each other word; find the best parse



RIT Rochester Institute of Technology

A neural graph-based dependency parser

- Dozat and Manning (2017); Dozat, Qi, and Manning (2017) - This paper revived interest in graph-based dependency parsing in a neural world
  - Designed a new scoring model (i.e., biaffine) for neural dependency parsing
- Really great results!

Method	UAS	LAS (PTB WSJ SD 3.3
Chen & Manning 2014	92.0	89.7
Weiss et al. 2015	93.99	92.05
Andor et al. 2016	94.61	92.79
Dozat & Manning 2017	**95.74**	**94.08**

- But, slower than the simple neural transition-based parsers.

RIT Rochester Institute of Technology

Outline

RIT Rochester Institute of Technology

Preview

- Thursday (2/5): Lab 4 - Work with a small dataset to train dependency parser (No in-person class; **Check mycourses on time - there'll be a link for a guided video**)
- Friday (2/6): Submission deadline for **Identifying team members**
- Tuesday (2/10) RNNs, LSTMs; Presenter: Shubh Sudan