

# Dependency Parsing



# Outline

Two formalisms for syntactic structure: Phrase structure and dependencies

Two algorithms for dependency parsing

- Transition based dependency parsing
- Graph based dependency parsing

Evaluating dependencies

This lecture has English examples. **But...**

...the ideas here are generally applicable to other languages as well.

**A useful exercise:** If you know another language, try to map what you see here to that language

# Outline

Two formalisms for syntactic structure: Phrase structure and dependencies

Two algorithms for dependency parsing

- Transition based dependency parsing
- Graph based dependency parsing

Evaluating dependencies

# Language is compositional

Words compose to construct complex meaning

Readers (and listeners) of language must unpack the compositional structure of language to understand it

Any system that seeks to “understand” language should also do so

# Language is compositional

Words compose to construct complex meaning

Readers (and listeners) of language must unpack the compositional structure of language to understand it

Any system that seeks to “understand” language should also do so

high school book shop returns desk

# Language is compositional

Words compose to construct complex meaning

Readers (and listeners) of language must unpack the compositional structure of language to understand it

Any system that seeks to “understand” language should also do so

high school book shop returns desk



high school book shop returns desk

# Language is compositional

Words compose to construct complex meaning

Readers (and listeners) of language must unpack the compositional structure of language to understand it

Any system that seeks to “understand” language should also do so

high school book shop returns desk



high school book shop returns desk

high school book shop returns desk

high school book shop returns desk

Or a different grouping?



# Linguistic syntax

How do words combine with each other to construct valid phrases, clauses and sentences?

Governs word order and grammatical relations between words such as subject verb agreement and preposition attachment

One of the more commonly seen structures in the NLP and linguistics literature

Other kinds of linguistic structure: phonological, morphological, semantic, discourse

# The New York Times

---

## *Paper Details Sites on Mars With Plumes of Methane*

By Kenneth Chang

Jan. 15, 2009

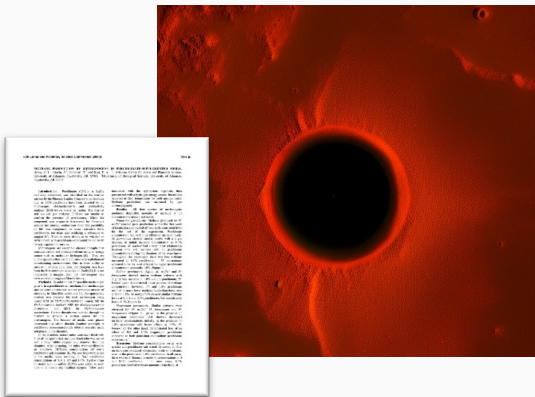
# The New York Times

## *Paper Details Sites on Mars With Plumes of Methane*

By Kenneth Chang

Jan. 15, 2009

Paper Details [[Sites on Mars with Plumes of Methane](#)]



# The New York Times

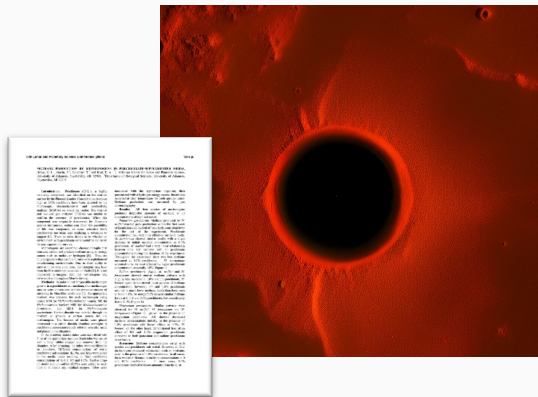
## *Paper Details Sites on Mars With Plumes of Methane*

By Kenneth Chang

Jan. 15, 2009

Paper Details [[Sites on Mars with Plumes of Methane](#)]

Paper Details [[Sites on Mars](#)] [[with Plumes of Methane](#)]



# The New York Times

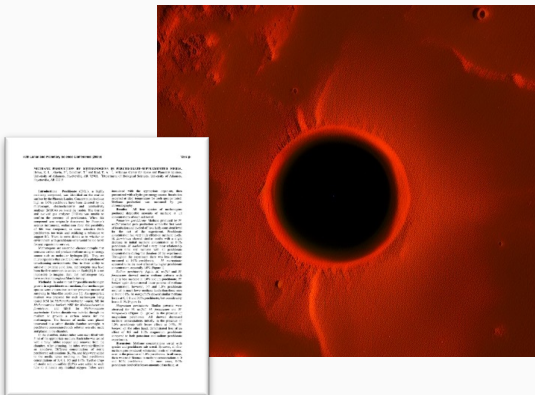
## *Paper Details Sites on Mars With Plumes of Methane*

By Kenneth Chang

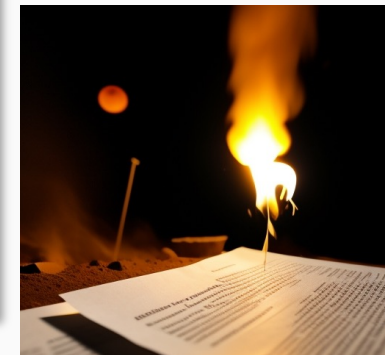
Jan. 15, 2009

Paper Details [[Sites on Mars with Plumes of Methane](#)]

Paper Details [[Sites on Mars](#)] [[with Plumes of Methane](#)]



Figuring out the the right syntactic structure resolves such ambiguities



# Linguistic syntax

How do words combine with each other to construct valid phrases, clauses and sentences?

Governs word order and grammatical relations between words such as subject verb agreement and preposition attachment

One of the more commonly seen structures in the NLP and linguistics literature

Other kinds of linguistic structure: phonological, morphological, semantic, discourse

Two types of syntactic structure are commonly seen in the literature:

1. Constituency structure or phrase structure
2. Dependency structure

# Linguistic syntax

How do words combine with each other to construct valid phrases, clauses and sentences?

Governs word order and grammatical relations between words such as subject verb agreement and preposition attachment

One of the more commonly seen structures in the NLP and linguistics literature

Other kinds of linguistic structure: phonological, morphological, semantic, discourse

Two types of syntactic structure are commonly seen in the literature:

1. Constituency structure or phrase structure
2. Dependency structure

Other theories of syntax also show up in the NLP literature somewhat less frequently, such as, [Combinatory Categorical Grammar \(CCG\)](#), [Tree Adjoining grammar](#)

# Linguistic syntax

How do words combine with each other to construct valid phrases, clauses and sentences?

Governs word order and grammatical relations between words such as subject verb agreement and preposition attachment

One of the more commonly seen structures in the NLP and linguistics literature

Other kinds of linguistic structure: phonological, morphological, semantic, discourse

Two types of syntactic structure are commonly seen in the literature:

1. Constituency structure or phrase structure
2. Dependency structure



# Phrase structure

Words in a sentence are grouped into nested constituents. Together, these form a tree

# Phrase structure

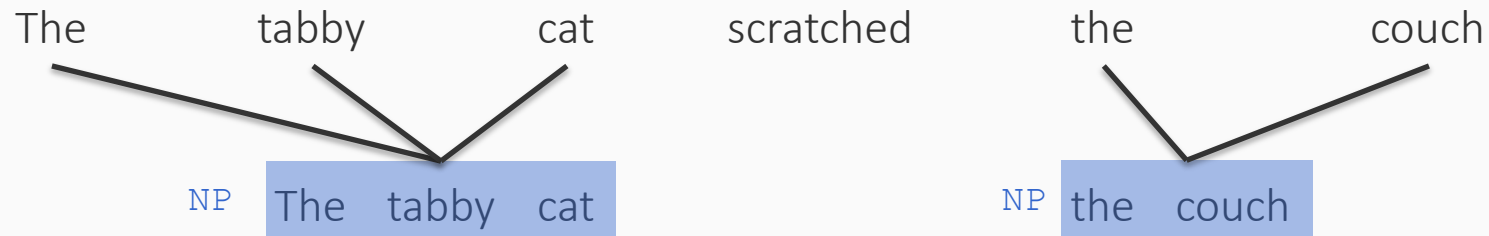
Words in a sentence are grouped into nested constituents. Together, these form a tree

The	tabby	cat	scratched	the	couch
DET	ADJ	NOUN	VERB	DET	NOUN

The smallest unit in the constituency tree is the word. Each word also has a *part of speech*

# Phrase structure

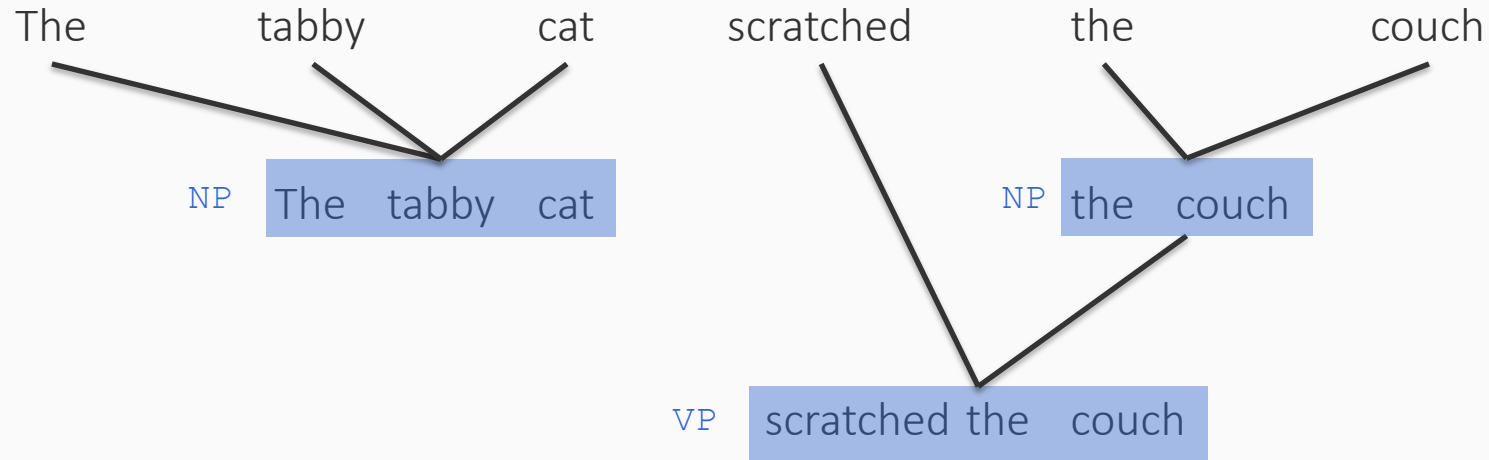
Words in a sentence are grouped into nested constituents. Together, these form a tree



Words combine to form constituents. Here we have two *noun phrases*

# Phrase structure

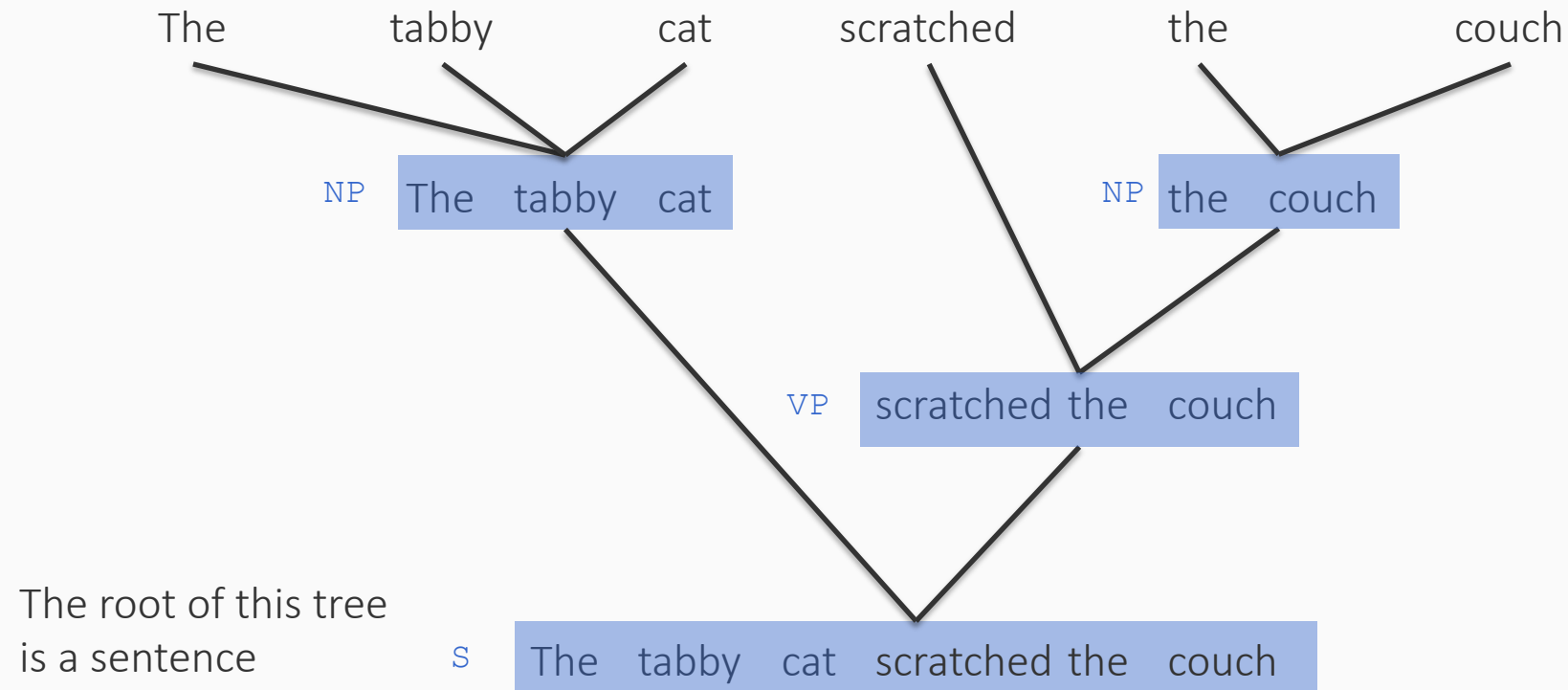
Words in a sentence are grouped into nested constituents. Together, these form a tree



Constituents group with with other constituents to form a hierarchical structure.  
Here we have a *verb phrase*

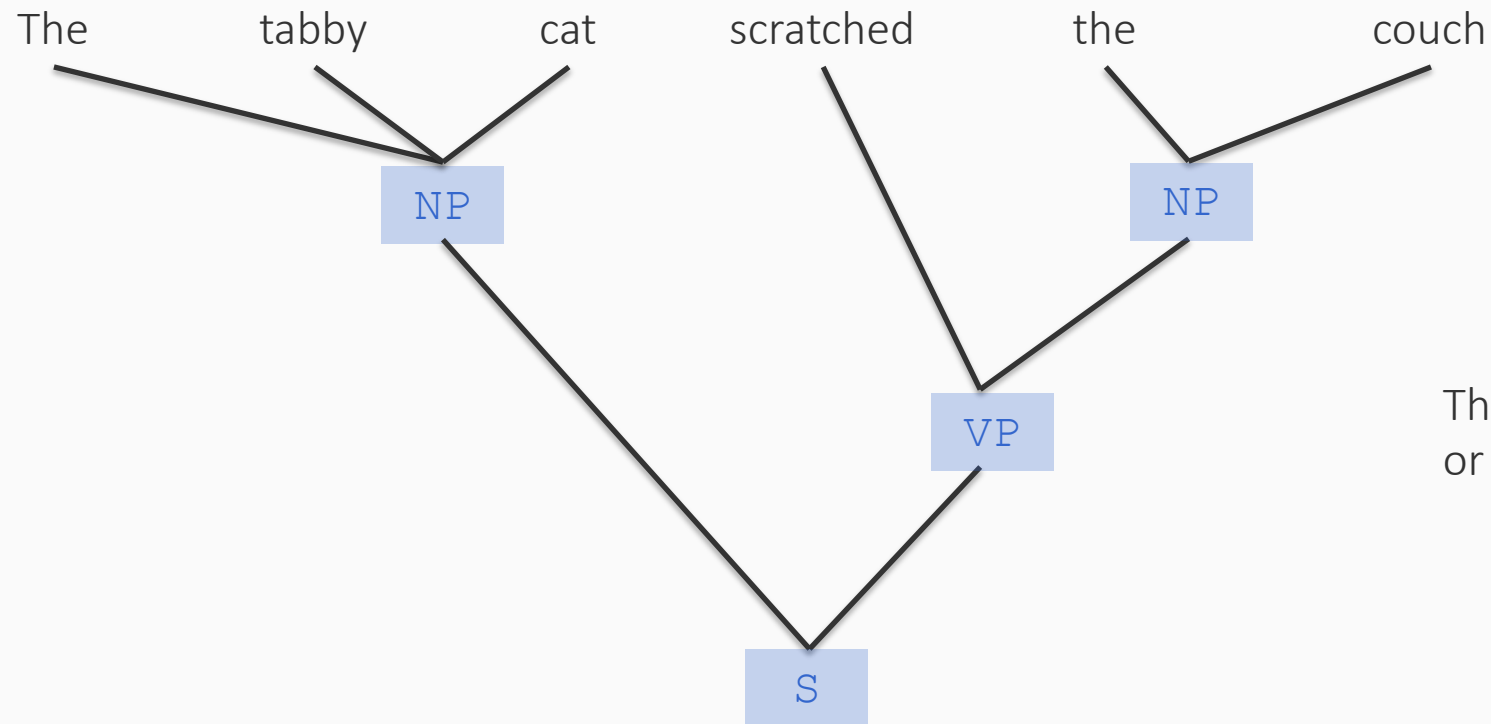
# Phrase structure

Words in a sentence are grouped into nested constituents. Together, these form a tree



# Phrase structure

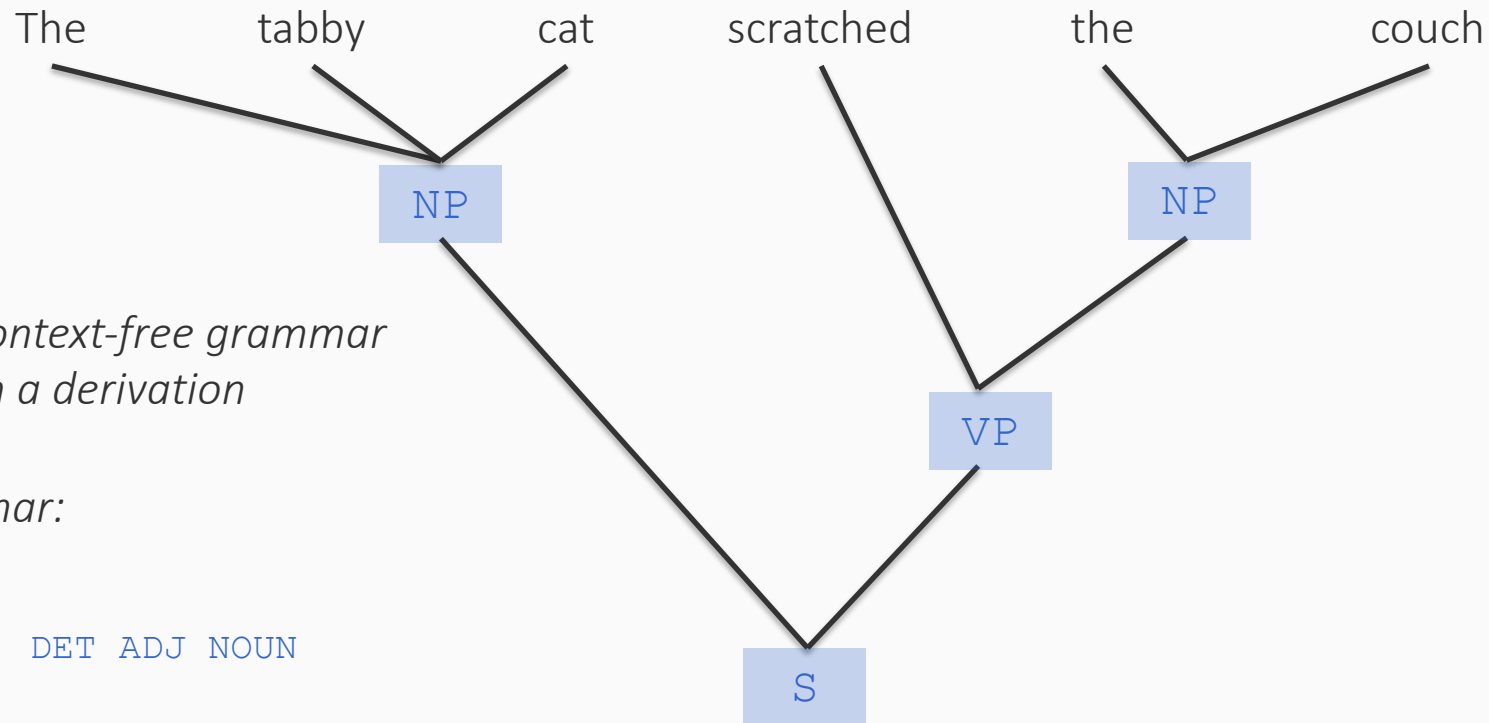
Words in a sentence are grouped into nested constituents. Together, these form a tree



This is a constituency tree or a phrase structure tree

# Phrase structure

Words in a sentence are grouped into nested constituents. Together, these form a tree



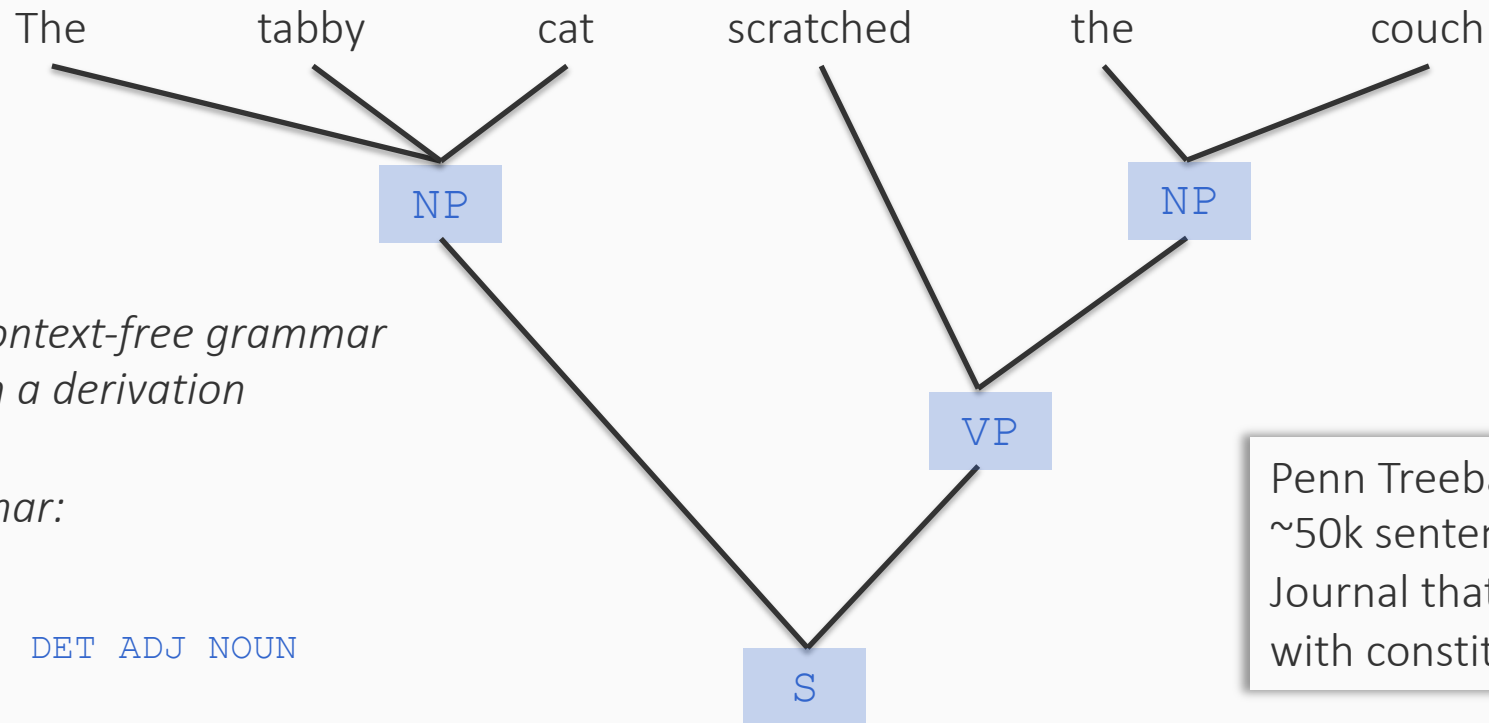
*Usually there is a context-free grammar that generates such a derivation*

*Here's a tiny grammar:*

```
S → NP VP  
NP → DET NOUN | DET ADJ NOUN  
PP → ADP NP  
VP → VERB NP | VERB PP
```

# Phrase structure

Words in a sentence are grouped into nested constituents. Together, these form a tree



*Usually there is a context-free grammar that generates such a derivation*

*Here's a tiny grammar:*

```
S → NP VP  
NP → DET NOUN | DET ADJ NOUN  
PP → ADP NP  
VP → VERB NP | VERB PP
```

Penn Treebank (1993): A dataset with ~50k sentences from the Wall Street Journal that were manually parsed with constituency trees



# Linguistic syntax

How do words combine with each other to construct valid phrases, clauses and sentences?

Governs word order and grammatical relations between words such as subject verb agreement and preposition attachment

One of the more commonly seen structures in the NLP and linguistics literature

Other kinds of linguistic structure: phonological, morphological, semantic, discourse

Two types of syntactic structure are commonly seen in the literature:

1. Constituency structure or phrase structure
2. Dependency structure

# Dependency structure

Syntactic structure of a sentence = directed binary grammatical relations between the words called *dependencies*

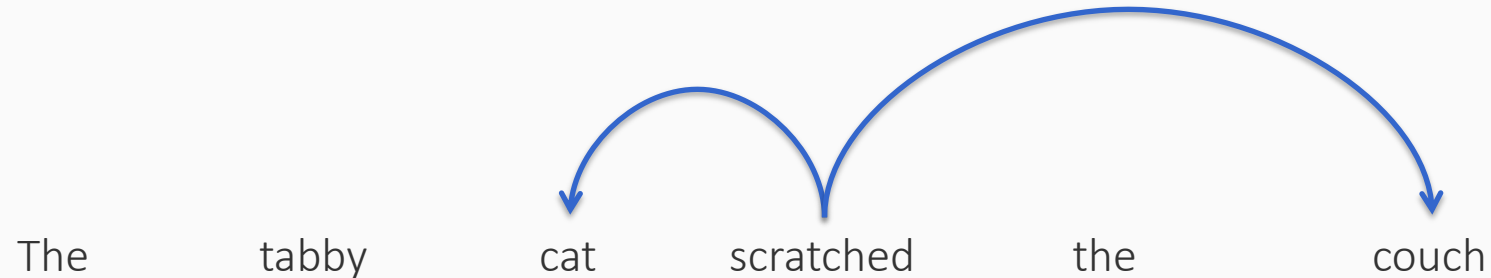
# Dependency structure

Syntactic structure of a sentence = directed binary grammatical relations between the words called dependencies

The tabby cat scratched the couch

# Dependency structure

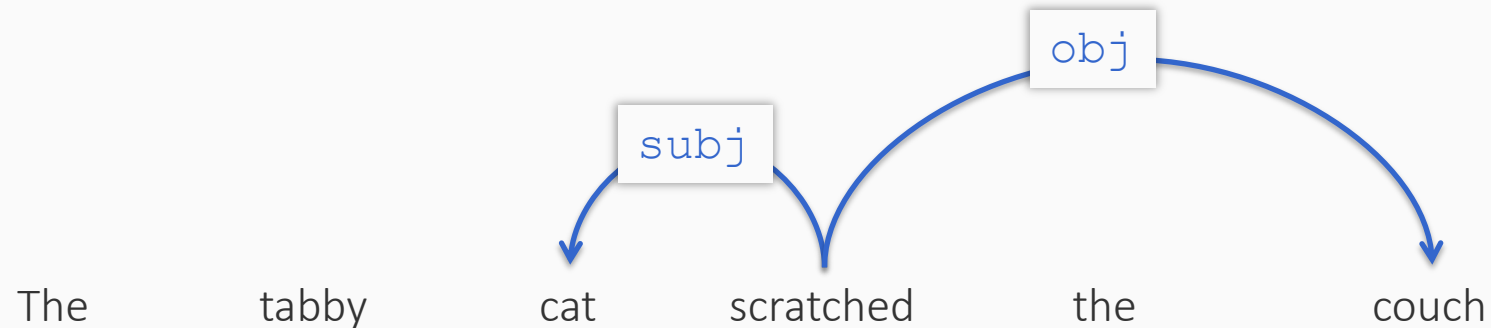
Syntactic structure of a sentence = directed binary grammatical relations between the words called dependencies



Dependency arcs go from the **head** word to its **dependents**

# Dependency structure

Syntactic structure of a sentence = directed binary grammatical relations between the words called dependencies

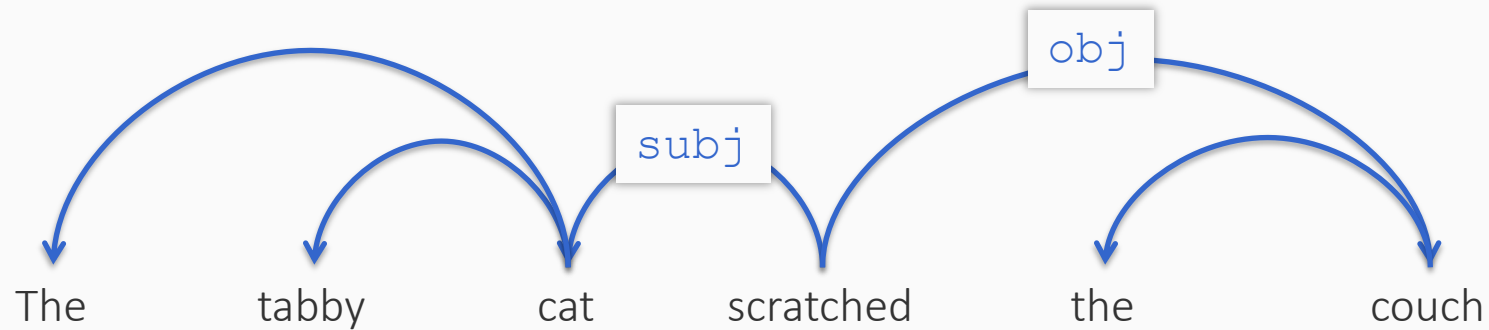


Dependency arcs are labeled with grammatical relations such as subject, object, noun modifier, determiner, etc.

The label describes what grammatical function the dependent plays with respect to the head

# Dependency structure

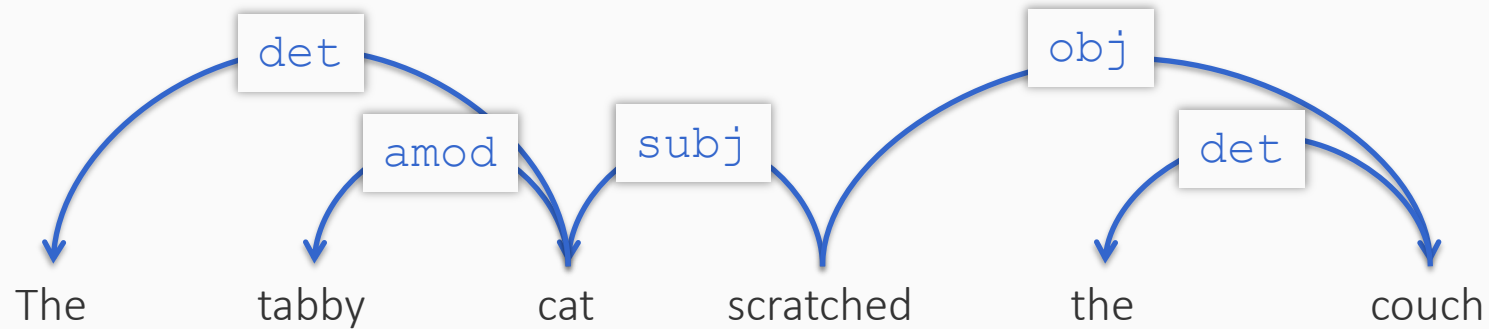
Syntactic structure of a sentence = directed binary grammatical relations between the words called dependencies



Every word in a sentence should be covered by the tree

# Dependency structure

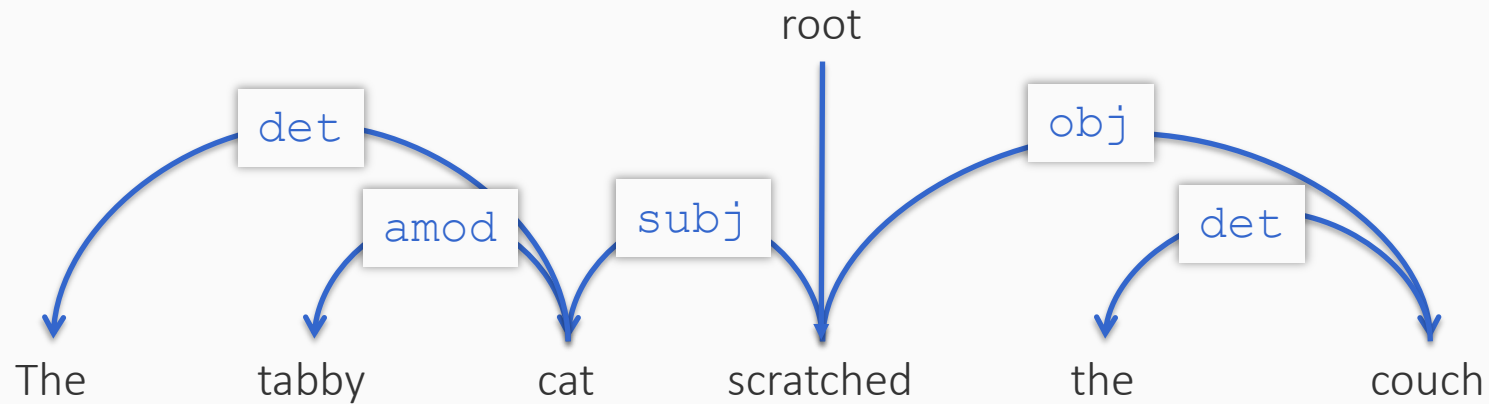
Syntactic structure of a sentence = directed binary grammatical relations between the words called dependencies



Every edge is labeled with a grammatical relation from a fixed set of labels

# Dependency structure

Syntactic structure of a sentence = directed binary grammatical relations between the words called dependencies



A *root* node explicitly marks the root of the tree



# Dependency datasets

The Universal Dependencies project: cross-linguistically consistent treebank annotation for many languages

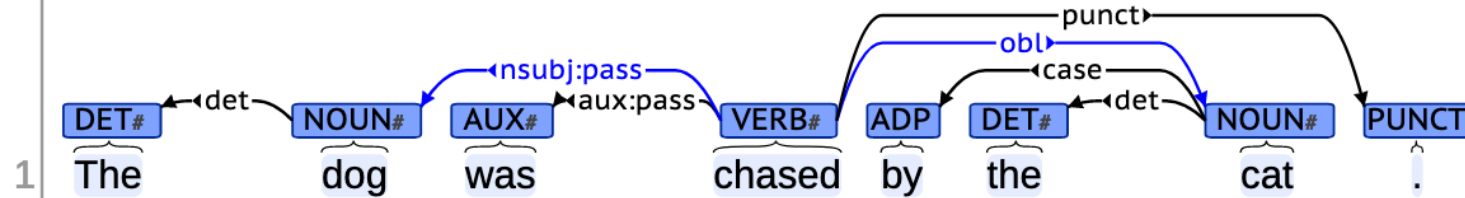
- Evolved from previous projects like Stanford dependencies, Google universal part-of-speech tags, etc

A common annotation scheme that has been used to annotate >200 treebanks in >100 languages (as of Sep 2023)

- And growing

# An example of universal dependencies

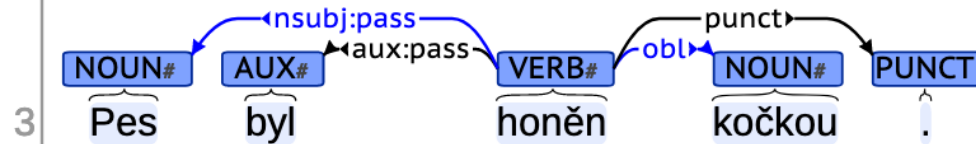
English



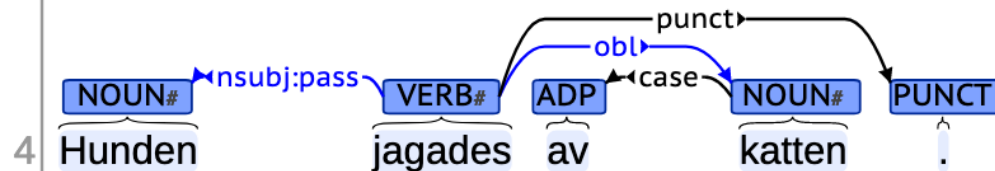
Bulgarian



Czech



Swedish



# History of dependency representations

- The idea of dependency grammar goes back to the Indian grammarian Panini (6<sup>th</sup>-4<sup>th</sup> century BCE)
  - Also slightly later in ancient Greek and Arabic linguistics
  - Phrase structure grammar is a relatively recent idea (mid-20<sup>th</sup> century)
- Modern dependency grammars draw from the work of Lucien Tesnière's 1959 work
- David Hays built one of the first automatic dependency parsers at the RAND corporation in 1962
- Resurgence after the 1990s with the algorithmic work of Eisner (1996), Nivre (2003), McDonald et al. (2005), etc, new modeling techniques and datasets like the Prague Dependency Treebank (Hajič, 1998) and the Universal Dependencies project (de Marneffe et al. 2021).