Syntactic Parsing: Phrase structures

COMP61332: Text Mining
Week 2
Riza Batista-Navarro

Phrase structure

Sentence structure based on phrases

Phrase structure trees show:

- groupings of words into phrases
- hierarchical structure of phrases

Phrase structure

Types of phrases

NP: noun phrase

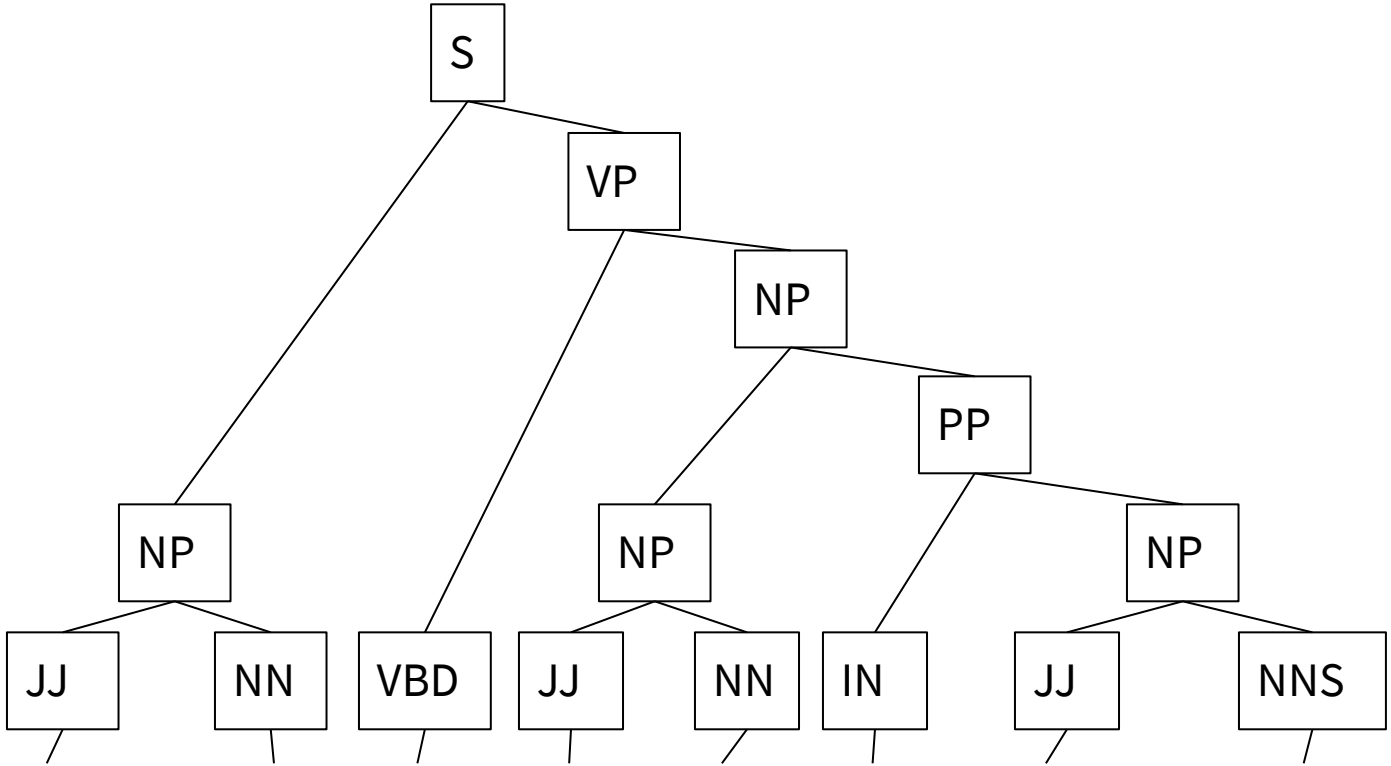
VP: verb phrase

PP: prepositional phrase

AdjP: adjectival phrase

AdvP: adverbial phrase

Phrase structure



Economic news had little effect on financial markets

Context Free Grammars

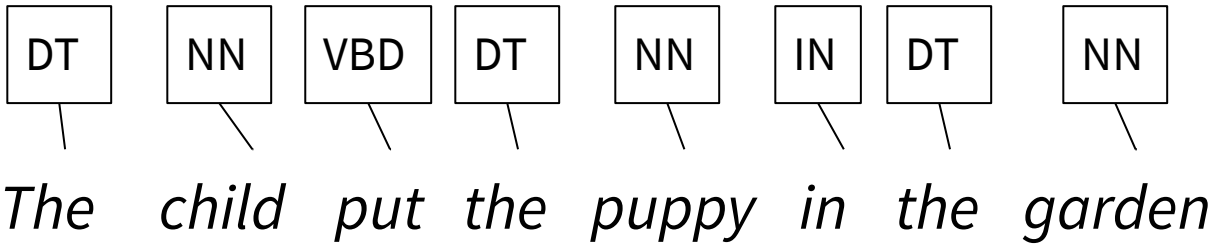
Production rules

Toy grammar for English:

```
S  → NP VP
NP → NN
   → PRP
   → DT NN
   → JJ NN
   → NP PP
VP → VBD
   → VP NP
   → VP PP
PP → IN NP
```

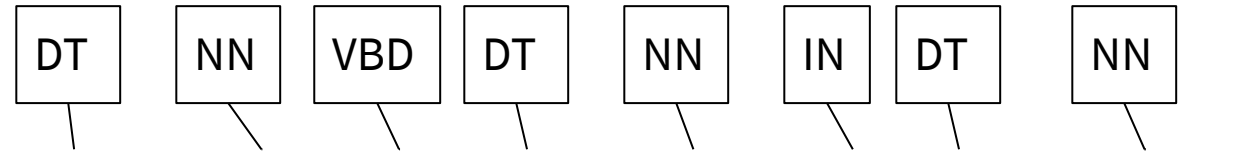
Analysis of Phrase Structure

Step 1: Label the syntactic category (POS) of each of the tokens



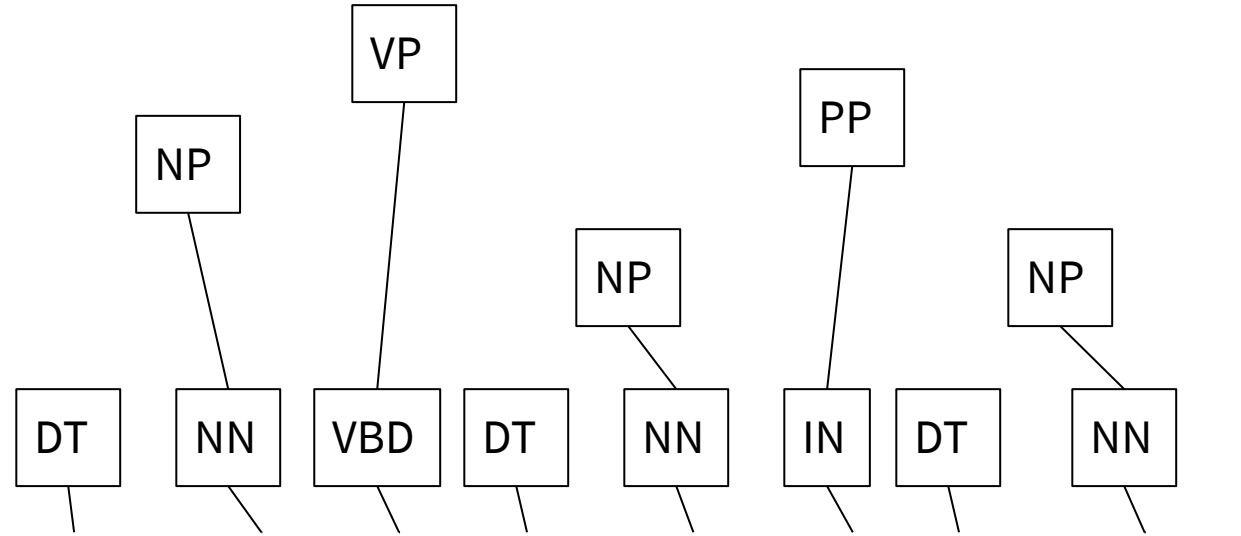
Analysis of Phrase Structure

Step 2: There are two principal constituents--a noun phrase (NP) and a verb phrase (VP). Locate the boundary between the two.



Analysis of Phrase Structure

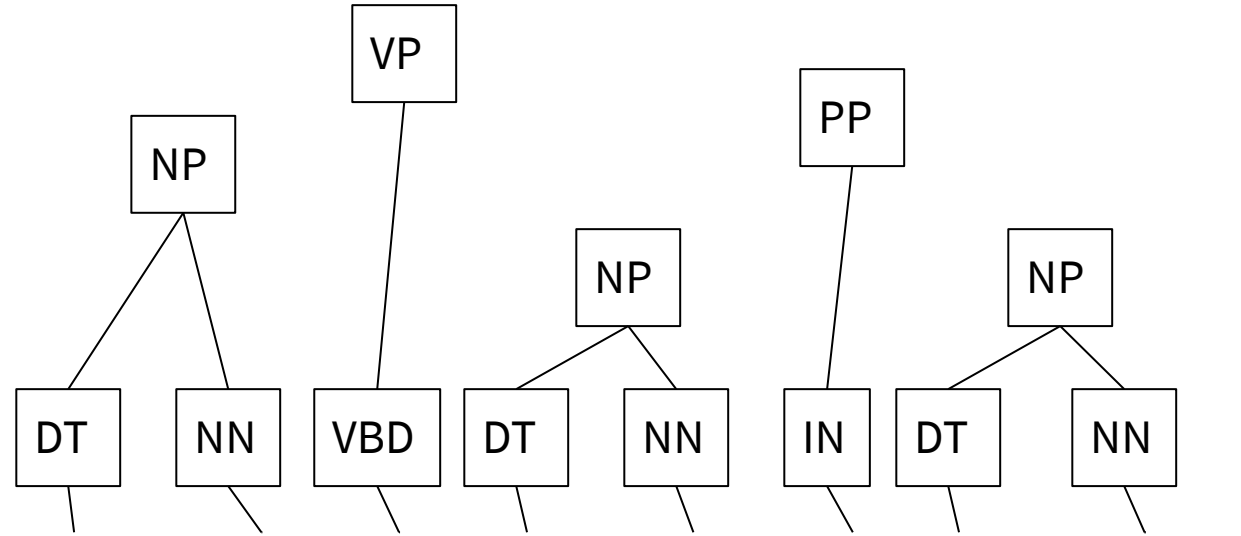
Step 3: For each head noun/pronoun, verb, adjective, adverb and preposition, project a phrasal node: NP, VP, AdjP, AdvP, PP.



[The child] [put the puppy in the garden]

Analysis of Phrase Structure

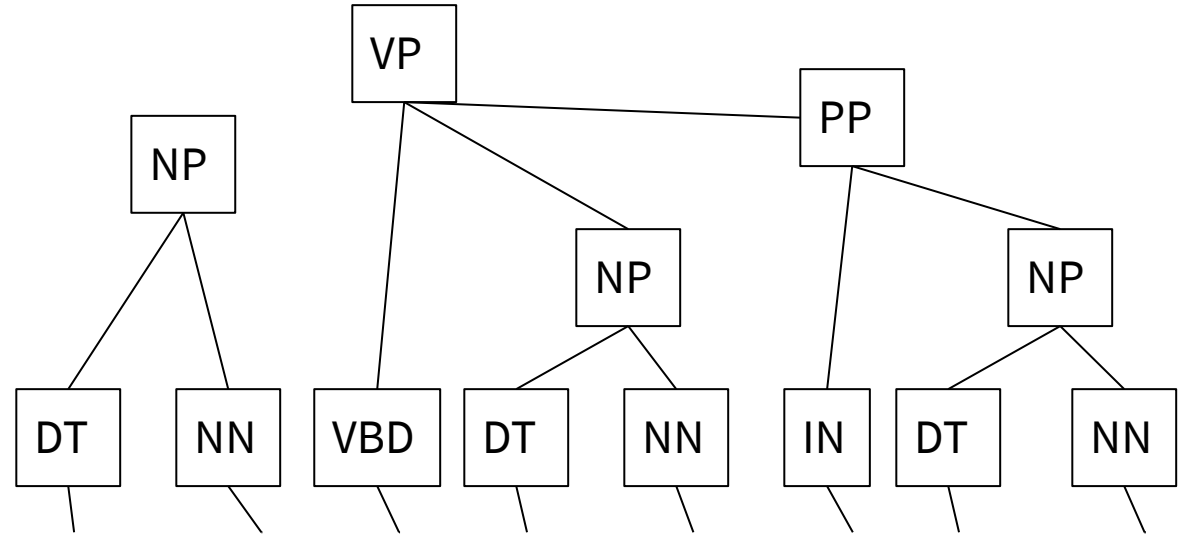
Step 4: Connect the remaining tokens to the nodes they belong to



[The child] [put the puppy in the garden]

Analysis of Phrase Structure

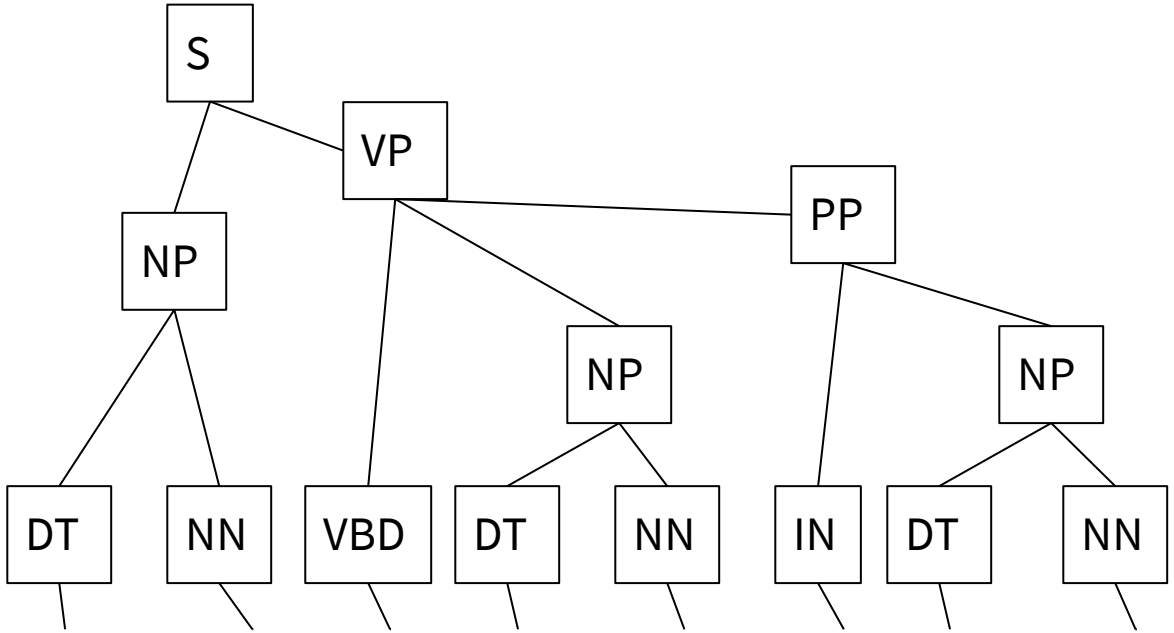
Step 4: Connect the remaining tokens to the nodes they belong to



[The child] [put the puppy in the garden]

Analysis of Phrase Structure

Final tree



[The child] [put the puppy in the garden]

Comparison: What can be represented?

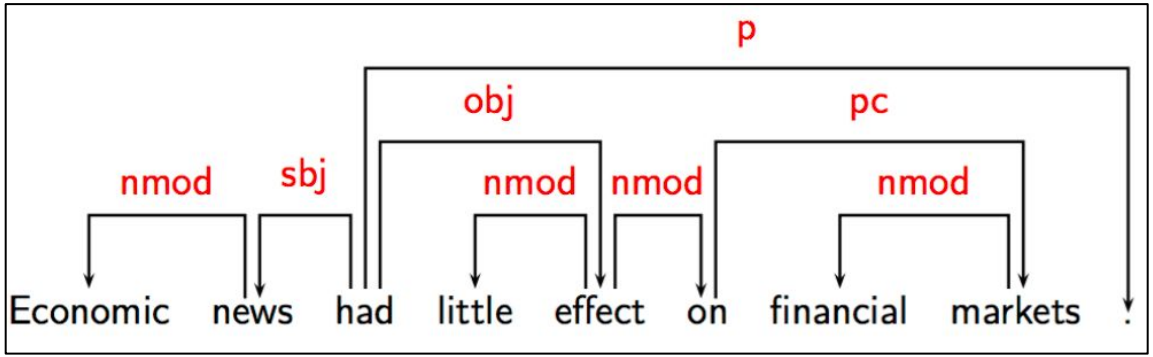
Dependency structure

- directed edges: head-dependent relations
- edge labels: grammatical functions
- POS tags: syntactic categories

Phrase structure

- non-terminal nodes: phrases (and their hierarchical structure)
- POS tags: syntactic categories

Dependency structures: Endocentric vs Exocentric Constructions



Construction	Head	Dependent	Required?
Exocentric	verb	subject (sbj)	yes
	verb	object(obj)	yes
Endocentric	verb	adverb (vmod)	no
	noun	adjective (nmod)	no

Verb valency

Certain verbs require certain types of dependents

Valency: the number of *arguments* controlled by the *predicate* (verb)

Verb valency

Verb class	Arguments	Example verb	Example sentence	Example
Univalent	1 subject	dance(subject)	She danced.	danced(She)
Divalent	1 subject, 1 object	eat(subject, object)	He ate the chocolates.	ate(He, chocolates)
Trivalent	1 subject, 2 objects	give (subject, object, indirect object)	He gave her the document.	gave(He, document, her)
Avalent	none/dummy	rain()	It rains.	rains()

Predicate-argument Structures (PAS)

Generalisation of the concept of verb valency

Predicates can be any part of speech (not just verbs)

Arguments are specified using an argument number

Predicate-argument Structures (PAS)

The executive order by Trump caused confusion at airports and embassies.

Example PAS given by the Enju Parser:

$\text{caused}_{\text{verb\_arg12}}$(order, confusion)

$\text{executive}_{\text{adj\_arg1}}$(order)

$\text{The}_{\text{det\_arg1}}$(order)

$\text{by}_{\text{prep\_arg12}}$(order, Trump)

$\text{at}_{\text{prep\_arg12}}$(caused, airports)

$\text{and}_{\text{coord\_arg12}}$(airports, embassies)

What can we use the results of parsing for?

Relation extraction

Information extraction