# Applied Human Language Technology

## Lecture 3

## Part 1: Sentences, Rewrite Rules, Parsing and Syntax

**Irene Murtagh**

## This week:

1. Phrase Structure and Grammar – The Basics
2. Word Classes
3. Syntactic Categories of English
4. The Representation of Phrase Structure
5. Modelling Language in Prolog
6. Some things to do

Back to Noam Chomsky...


The Chomsky Hierarchy of languages

# Phrase Structure and Grammar – The Basics

## Word Classes

- Word classes in traditional grammar
- Levels of analysis and defining categories
- Category arguments and diagnostics
- Lexical and phrasal categories

## Syntactic Categories of English

- Lexical Categories
- Phrasal Categories
- The relationship between lexical categories and phrasal categories

## The Representation of Phrase Structure

- The NP in English
- Phrase structure rules (PSRs)
- Phrase markers (PMs) and syntactic trees
- Lexical insertion rules

- The noun is a part of speech having case inflections, signifying a person or thing

- The verb is a part of speech without case-inflection, admitting inflections of tense, person and number, signifying an activity or a being acted upon.

- The preposition is a word placed before all other parts of speech in word-formation and syntactic constructions.

## Constituents and Constituent Structure

- Words are organised into **units** which are then organised into larger units.

- These units are called **constituents**, and the hierarchical organisation of the units in a sentence is called its **constituent structure**.

## Phrasal Categories

- A **phrasal construction** can then be defined as any construction which has a **head** and a phrase as any unit which exhibits such a construction

  i.e. any XP not only functions like an X but has an X as its head.

  P stands for phrase

## Phrasal Categories

- In addition to sentences (correct and incorrect), we have intuitions about certain classes of words and groupings of words.

- Today is [$a_{Det}$ $lovely_{Adj}$ $day_{N}$]
- The rebel parties [$must_{Aux}$ $unify_{Verb}$]

- *Today is [$a_{Det}$ $lovely_{Adj}$]
- *The rebel parties [$must_{Aux}$ $unify_{Verb}$]

- We use categories and phrases to characterise these intuitions about word groupings.

**Phrase Structure**: describes the way in which a sentence is organised into successively smaller units.

**Phrasal Categories**

- A **phrasal construction** (e.g. verb phrase / noun phrase) can be defined as any construction which has a **head** and a phrase as any unit which exhibits such a construction

  i.e. any XP not only functions like an X but has an X as its head.

  P stands for phrase

# Syntactic Categories (Parts of Speech) of English

Some English **Lexical Categories**

| | |
|---|---|
| V | verb (arrive, discuss, melt, remain, dislike) |
| N | noun (boy, wheat, policy, moisture, bravery) |
| ADJ | adjective (good, tall, intelligent, fond, big, small) |
| ADV | adverb (silently, slowly, quickly, now) |
| P | preposition (to, in, near, at, by, after, on) |
| DET | determiner (the, a, an) |
| COR | co-ordinator/conjunction (and) |
| SUBORD | subordinator/complementiser (if) |

Further divisions

| | | |
|---|---|---|
| Adverbs | → | adverbs, degree words (so, too, very, etc. |
| Verbs | → | verbs, auxiliary verbs (may, can, will, etc.) |
| Determiners | → | indefinite articles, definite articles, demonstratives, etc. |
| Adjectives | → | adjectives, quantifiers, etc. |

## Some English **Phrasal Categories**

| | |
|---|---|
| NP | noun phrase |
| VP | verb phrase |
| PP | prepositional phrase |
| AP | adjectival phrase |
| ADVP | adverbial phrase |
| QP | quantifier phrase |

## The Representation of Phrase Structure

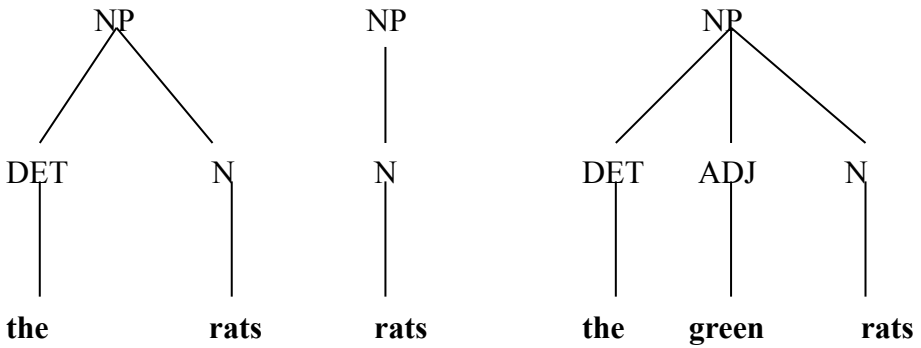
English NP Composition

| | | |
|---|---|---|
| rats: | N | |
| the rats: | DET N | |
| green rats: | ADJ N | |
| the green rats: | DET ADJ N | |
| the large green rats: | DET ADJ ADJ N | |
| the very large green rats: | DET **ADV** ADJ ADJ N | *or* |
| | DET **DEG** ADJ ADJ N | |

# Representing Constituents and Constituent Structure

## Phrase Markers and Tree Diagrams

a. b. c.



## Brackets

a. $[_{NP}\ [_{DET}$ the$]\ [_{N}$ rats$]]$

b. $[_{NP}\ [_{N}$ rats$]]$

c. $[_{NP}\ [_{DET}$ the$]\ [_{ADJ}$ green$][_{N}$ rats$]]$

## Summarising possible structures as ***rewrite*** **or** ***phrase structure rules***

a. NP → N

b. NP → DET N

c. NP → DET ADJ N

d. NP → { (DET) (ADJ) N }

*Abbreviations*:

{ } = *or* list

( ) = optional

We can say that in :

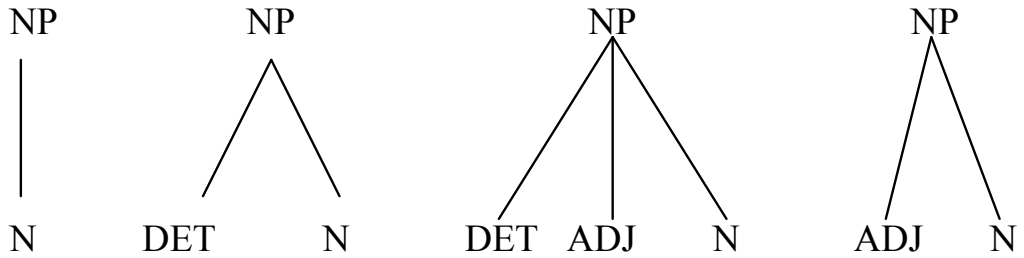
a. the NP **dominates** N,
b. the NP dominates the DET and the N, and
c. the NP dominates the DET, the ADJ and the N.

These phrase structure rules are also called **immediate-dominance rules** or **ID-rules**.

## Phrase Structure (PS) Rules and Phrase Markers (PMs)

- A PS rule can be seen as a prediction about possible structures.
- It reflects a hypothesis or claim based on syntactic arguments.

**NP → (DET) (ADJ) N**



# A Simple **Phrase Structure Grammar (PSG)**

PSRs like **a**. predict structures including **b**.

**a**.

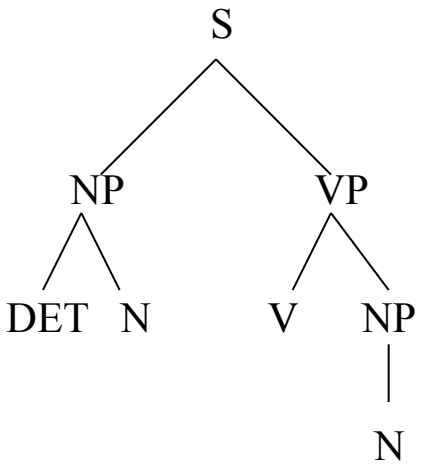
S → NP VP

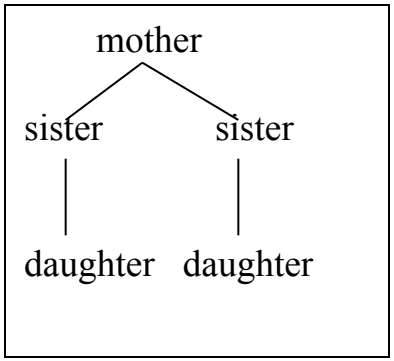
VP → V (NP)

NP → (DET) N

→ Prolog like representation

**b**.



To know what actual sentences a tree like the one above describes, we need to have the

1. **words of the language**
2. listed in their **syntactic category**
3. in a **lexicon**.

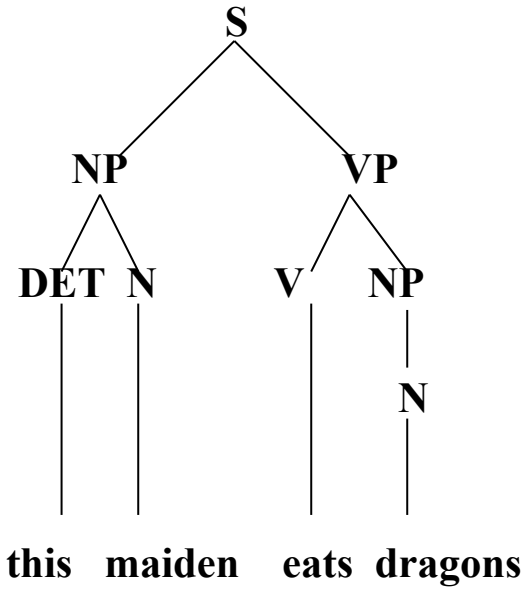
| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| dragon: | PHON: | | eat: | PHON: | | this: | PHON: | |
| | SYN: | **N** | | SYN: | **V** | | SYN: | **DET** |
| | SEM: | | | SEM: | | | SEM: | |
| maiden: | PHON: | | reject: | PHON: | | the: | PHON: | |
| | SYN: | **N** | | SYN: | **V** | | SYN: | **DET** |
| | SEM: | | | SEM: | | | SEM: | |

Now we need to match the prediction in the tree with its lexical information to see what possible sentences it describes. In Chomsky's *Standard Theory* (Chomsky 1965) for example, a **Lexical Insertion Rule** was used. This fills empty PMs with meaningful elements from the lexicon

**Lexical Insertion Rule**

1. Insert from the lexicon an item of the category matching the terminal symbol.
2. Attach the item underneath the relevant symbol (as a daughter).

From our **phrase marker tree** above and the **lexicon**, we can generate the **lexically filled** PM tree. [We ignore morphological details.]

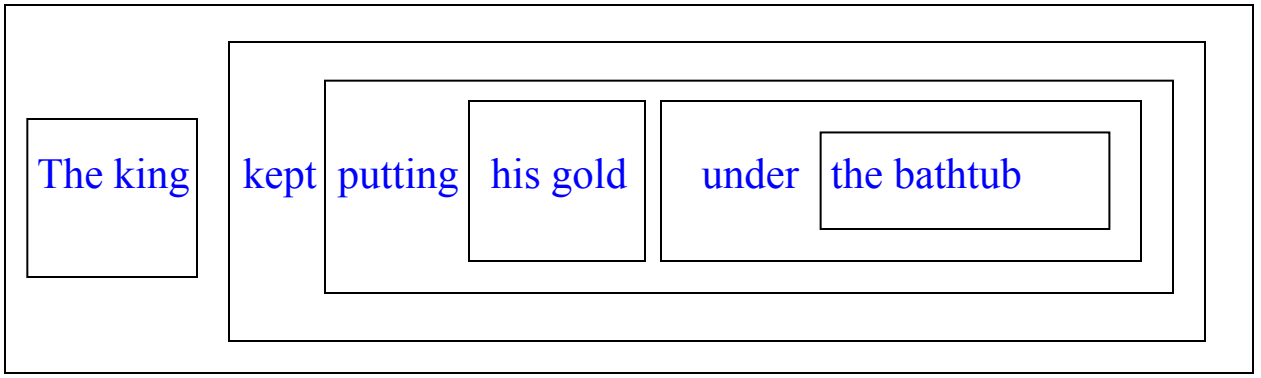


# Some More Remarks about Phrase Structure

**Phrase structure**

The way in which a sentence is organised into successively smaller units.

The king kept putting his gold under the bathtub



## Phrases That Serve As Sentence Predicates

A sentence consists of a subject and a predicate.

Tara fed the dog

The **subject** is "the thing talked about" → Tara

The **predicate** is the "thing said about the subject" → fed the dog

## Phrases that help build subjects

Phrases that can join with the part of speech [the:**DET**]

a. The *king of Spain*. (the + common noun)

b. The *leader of the Labour party*

c. The *members of the committee*.

d. *The ran the club. (*ran* is a verb)

e. *The Joe Smith. (*Joe Smith* is a proper noun)

f. *The angry at the judge. (*angry* is an adjective)

## Phrases Headed by Adjectives

a. The cats seem ***eager*** *to leave the meeting*. (***eager*** is an adjective)

b. *The cats seem ***want*** *to leave the meeting*. (***want*** is a verb)

After *seem*, we can have phrases headed by adjectives but not by verbs.

## Phrases consisting of a single word

a. The dogs ***snore***. (*snore* is a verb)

b. the ***book*** (*book* is a noun)

c. The cats seem ***despondent***. (*despondent* is an adjective)

Our concept of phrase allows for instances consisting of only one word.

# Remember:

**In Phrases That Serve As Sentence Predicates**

**The predicate is the thing said about the subject…..**

A sentence consists of a subject and a predicate.

Tara fed the dog

The **subject** is "the thing talked about" → Tara

The **predicate** is the "thing said about the subject" → fed the dog

## Distinctions among verb forms

| | |
|---|---|
| 1. The cats [*know* everything]. | Present Tense |
| 2. The cats [*knew* everything]. | Past tense |
| 3. *The cats [*knowing* everything]. | Present participle |
| 4. *The cats [*known* everything]. | Past participle |

A phrase headed by a present tense or past tense verb can serve as the predicate of an independent sentence.

a. *The cat [*know* everything].

b. The cat [*know**s*** everything]. Present Tense

c. The cat**s** [*know* everything]. Present Tense

d. *The cat**s*** [*know**s*** everything].

The verb that heads the predicate of a sentence must agree with the subject in number (singular or mass noun, plural) and person (first, second, third).

We can therefore combine these two rules into one consolidated rule as:

A phrase headed by a present tense or past tense verb can serve as the predicate of an independent sentence that agrees with the subject in number and person.

## Other Parts of Speech as Heads

Any phrase whose head belongs to part of speech X can be used in such a way as to form a larger construction.

### Other Parts of Speech as Heads

a. **prepositions** (*in, on, about, to, around*)

b. definite **pronouns** (*I, you, him, them*)

c. indefinite pronouns (*something, anyone*)

d. coordinating **conjunctions** (*and, or, nor, but*)

e. subordinating conjunctions (*that, because, when, what, before, after*)

f. **Determiners** / **articles** (*the, a, an, this, those*)

g. manner **adverbs** (*slowly, wistfully, courageously*)

h. degree adverbs (*very, extremely, rather, quite*)

i. locative adverbs (*here, there, everywhere*)

j. frequency adverbs (*often, occasionally, frequently, never*)

## Locative Phrases

These indicate stationary location

a. Irene stayed *at the café*.

b. The keys are *right up here on the counter*

## Motion Phrases

These indicate some kind of movement in a certain direction or towards a certain location.

a. The bread fell *into the soup*.

b. Martin walked *back up here onto the stage*.

c. The tiger moved *towards the child*.

d. Niamh took the vacuum cleaner *around the house*.

BUT… The interpretation of the phrase depends on its context:

a. Jane **kept** her turtle *over there*. [**locative** phrase]

b. Jane **took** her turtle *over there*. [**motion** phrase]

# Lexical Information about Phrases: Heads and Complements

1. Phrases and their Heads
   - Types of verbs: Auxiliary & lexical (main) verbs; copula (linking) verbs.
   - Verbs and their objects;
   - Intransitive verbs; transitive verbs; ditransitive verbs.
   - Problems identifying transitivity; displaced objects; arguments and adjuncts

2. Verb Phrases – General Issues
   - Auxiliary verb complements
   - Copula verb complements
   - Lexical verb complements

3. Verb Phrases – Types of Complements
   - Adjectival complements in AP
   - Nominal complements in NP
   - Prepositional complements in PP

# Phrases and Their Heads

In linguistics, the head of a phrase is the word that determines the **syntactic type** of that phrase

Traditionally, a complement is a constituent of a clause, such as a **noun phrase** or **adjective phrase**, that is used to predicate a description of the subject or object of the clause

## Phrases and Their Heads

### Head of NP

[NP [NP [DET The] [ADJ giant] [N bat]] [PP [P outside] [NP [DET your] [N window]]]] interests me.

### NPs:

[NP the giant bat]

[NP the giant bat outside your window]

[SUBJ The giant bat outside your window] interests/*interest me

## Verbs

- **Lexical** verbs: *think, walk, eat, grow*, etc.
- **Auxiliary** verbs: *do, have, be, can, will, must, ought to*, etc.

  a. I *did* see him there.

  b. We *must* get back before it's too late.

  c. They *used to* ride the railroad cars.

- **Copula** Verbs: *be, become, seem, sound, smell, taste, look, get, turn*, etc.

  These give an attribute of the subject, or identify the subject.

- **Copula** Verbs: ***be****, become, seem, sound, smell, taste,* ***look****, get, turn*, etc.

  These give an attribute of the subject, or identify the subject.

  a. She ***is*** an architect. vs. She ***shot*** an architect.

  b. [$V_{COP}$ NP] vs [$V_{Lex}$ NP]

  c. Your singing ***sounded*** terrible.

  d. He ***looks*** sick.

  e. He ***looks*** like Bob Hope.

  f. He ***looks*** like a moron.

# Lexical Verbs and Their Objects

We will assume a Phrase Structure rule of: VP → V NP

1. Godzilla ***chased*** the sea monster.
2. *Godzilla ***chased***.
3. Godzilla ***slept***.
4. *Godzilla ***slept*** a sleep.
5. Godzilla ***ate*** Tokyo.
6. Godzilla ***ate***.

## Phrase Structure Rules

1. *chase*: VP → V NP
2. *sleep*: VP → V #
3. *eat*: VP → V (NP)
4. *put*: VP → V NP PP

[where # = the end of the VP, or 'nothing'.]

1. Godzilla ***put*** a train up his nose.
2. *Godzilla ***put*** a train.
3. *Godzilla ***put***.

## Lexicon Entry

1. *chase*:V: [ __ NP] = transitive verb
2. *sleep*:V: [__ #] = intransitive verb
3. *eat*: V: [__ NP] = transitive verb both
4. *eat*: V: [__ #] = intransitive verb both
5. *put*: V: [__ NP PP] = ditransitive verb

We use :

**subcategorisation frames** and complement frames

[**head** | complement]

- Record syntactic information in the lexicon.
- We need information organised lexically.
- There are too many phrase structure rules, therefore put information in the lexicon.
- Used in **parsing** of the sentence.

# TO DO THIS WEEK

1. **Lab session on Finite State Automata using Java**

**2. Read** : Ch.8 & 9: *Speech and Language Processing* by Jurafsky & Martin

3. In J+M book,

...do **exercises in 9.1 and 9.2** on page 355

This counts as CA