# Syntax: the structure of sentences

Salvador Mascarenhas

Introduction to Linguistics, ENS, Fall 2023

Lectures #4 and #5

## 1 Syntactic competence

- Syntactic theory aims to account for speakers' syntactic competence, and for the patterns of cross-linguistic variation.
- Syntactic competence includes (a)–(d) below, **among other things**.

**(1)** The knowledge of which sequences of words are sentences of a language — which strings are grammatical and which are not

(1) a. I watched his steps.
b. I watched his every step.
c. *I watched his many steps.

(2) a. I watched his five steps.
b. *I watched his all steps.
c. *I watched his most steps.

(3) a. It is likely that John has left.
b. It is probable that John has left.

(4) a. John is likely to have left.
b. *John is probable to have left.

(5) a. I should have bought it.
b. I should have more money.
c. I never got to school on time.
d. I have got to leave.

(6) a. I shoulda bought it.
b. *I shoulda more money.
c. *I never gotta school on time.
d. I've gotta leave.

**(b)** An ability to recognize a sentence as well-formed even if it does not "make sense" or consists of highly unlikely sequential transitions

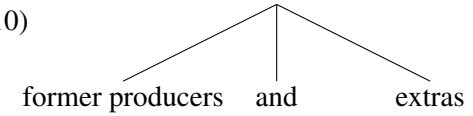
(7) Colorless green ideas sleep furiously.

**(c)** An ability to determine what the sentence means, whether it is ambiguous or not

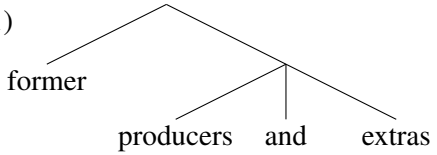
(8) I went to the bank — lexical ambiguity
"I went to the financial institution"
"I went to the river bank"

(9) I met former producers and extras. — structural ambiguity
"I met former producers and I met extras"
"I met former producers and I met former extras"

(10)




(11)




- More structural ambiguity:

(12) a. They attract people with money.
b. Visiting relatives can be boring.
c. When did you say they got married?

**(d)** A crucial feature of human languages is that every language has an infinite set of sentences. Native speakers, although they are finite beings with finite minds, can handle this set.

- How do we prove this? We show that there is no longest sentence in any language (cf. proof of the infinity of natural numbers).
- Does this entail that humans are capable of processing sentences of arbitrary length?

## 2 Lexical categories (parts of speech)

- N(oun), V(erb), A(djective), P(reposition), etc. Traditionally, parts of speech are defined with reference to **meaning**. This suggests that you can only tell what party of speech a word is if you know its meaning. But witness:

(13) 'Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogroves,
And the mome raths outgrabe.

- A better way to define lexical categories: look at the sentential environments that it can occur in and what affixes it can take — **grammatical distribution**
- A verb in English satisfies the conjunction of the following conditions:
  1. it can occur right after "I will," and
  2. in some environments it can take the endings -s or
  3. -ing

## 3 Constituency

- **Heuristic:** A sequence of words forms a **constituent** in a sentence if
  1. it can be replaced with a minimal unit, ideally one word, preserving grammaticality, and conversely,
  2. occurrences of that minimal unit in other sentences can typically be replaced by our sequence of words, preserving grammaticality
- If 1 and 2 hold, then the sequence and the minimal unit are systematically interchangeable and, in the sentence under investigation, the sequence has the same category as the minimal unit.
- Example: "fall asleep" is systematically interchangeable with "hiccup"

(14) a. She may fall asleep $\Longrightarrow$ She may hiccup
b. Hiccup, she never will $\Longrightarrow$ Fall asleep, she never will

- but "John and" is not systematically interchangeable with "yesterday"

(15) a. John and Mary won $\Longrightarrow$ Yesterday Mary won
b. I overslept yesterday $\Longrightarrow$ * I overslept John and

- Fronting and Pronominalization

(16) a. Marie a mangé [la pomme]
b. [La pomme], Marie **l**'a mangée $\Longrightarrow$ [La pomme] is a constituent

(17) a. Paul n'ira jamais [au cinéma]
b. [Au cinéma], Paul n'ira jamais
c. Il n'**y** est pas allé $\Longrightarrow$ [Au cinéma] is a constituent

(18) *Jamais au cinéma, Paul n'ira

(19) Clefts
a. Mary went [to the theater] [with her friend].
b. It was [to the theater] that Mary went with her friend.
c. It was [with her friend] that Mary went to the theater.
d. *It was [to the theater with her friend] that Mary went.

Only constituents can be fronted, pronominalized or clefted

- Elided answers

(20) a. Qui est-ce que Paul a vu
b. (Paul a vu) [le garçon dont je t'ai parlé]

(21) a. Où Paul est-il allé?
b. (Paul est allé) [au cinéma]

(22) a. Qu'a fait Paul?
b. (Il a) [lu un livre]

• Coordination. This is more complex

(23) Marie a lu [un livre] et [un journal]

(24) Marie a écrit à Pierre et à Paul

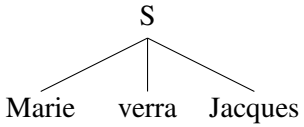
(25) *Marie a écrit à Pierre et quand elle était à Paris.

## 4 A small French fragment

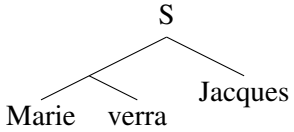
(26) Marie verra Jacques

Three possible structures in principle.

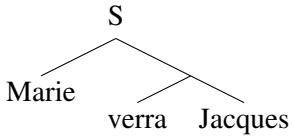
(27) a.




b.




c.




• (27c) is preferred. . . but there is little *direct* evidence.

• Coordination:

(28) a. Marie [ [verra Jacques] et [entendra Paul] ]
b. ?Marie verra et Jacques entendra Paul

• Let's build a rewriting grammar for a fragment of French . . .

(29) a. Marie verra/entendra/. . . Jacques
b. Un type verra Jacques
c. Marie verra un type
d. Le type connu a vu Marie
e. Marie dormira
f. etc.

S $\longrightarrow$ NP VP
NP $\longrightarrow$ Proper Name
NP $\longrightarrow$ Det N (Adj)
VP $\longrightarrow$ $V_{intr}$
VP $\longrightarrow$ $V_{tr}$ NP

(30) Marie pense que Paul verra Jacques

CP $\longrightarrow$ que S
VP $\longrightarrow$ $V_{cl}$ CP

$V_{tr} \longrightarrow$ voir, entendre, . . .
$V_{intr} \longrightarrow$ dormir, briller, . . .
$V_{cl} \longrightarrow$ penser, croire, . . .

• Introducing adjunction

(31) a. Marie verra Jacques quand Paul dormira
b. Marie verra Jacques quand Paul dormira avant que Pierre n'arrive

**1.** 'quand Paul dormira' and 'avant que Pierre n'arrive' are constituents

$C_{temp} \longrightarrow$ avant que, quand
$CP_{temp} \longrightarrow C_{temp}$ S

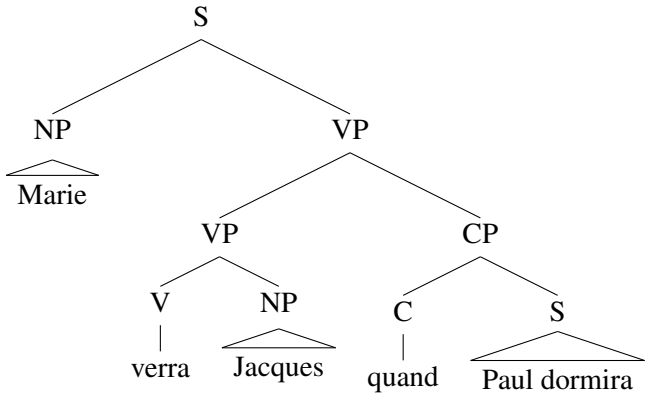
**2.** Where are they attached? Several possibilities in principle

(32) a. S $\longrightarrow$ NP VP ($CP_{temp}$)
b. (i) VP $\longrightarrow$ $V_{intr}$ ($CP_{temp}$)
(ii) VP $\longrightarrow$ $V_{tr}$ NP ($CP_{temp}$)
c. S $\longrightarrow$ S $CP_{temp}$
d. VP $\longrightarrow$ VP $CP_{temp}$

- (32a) predicts that 'VP CP' is not a constituent.

(33) Marie [verra Jacques quand Paul dormira] et [mangera une pomme]

- (32b) is not economical and does not make 'V NP' a constituent.
- (32c), just like (32a), makes it impossible that 'verra jacques quand Paul dormira' be a constituent
- So we pick (32d).




- Rule (32d) is an *adjunction rule*. It says that a constituent of a certain category can combine with something and yield the same category.

## 5 Complements and Adjuncts

(34) Paul compte sur Marie

(35) Paul dort sur le lit

(36) Paul dort

(37) *Paul compte [under the intended meaning]

(38) Paul dormira sous le lit

(39) Paul dormira à côté de la chaise

(40) Paul pensera à côté de la chaise

(41) *Marie compte sous le lit

(42) *Paul compte sous Marie

(43) Paul dormira sous le lit à côté de la chaise près du lavabo

- No shared coordination:

(44) *Paul compte et dort sur le lit

VP $\longrightarrow$ $V_{compte}$ $PP_{sur}$

- In Paul compte sur Marie', 'sur Marie' is a *complement*. In 'Paul crie sur le pont', 'sur le pont' is an *adjunct*

VP $\longrightarrow$ VP PP
PP $\longrightarrow$ P NP
P $\longrightarrow$sur, sous, à côté, ...

(45) Marie comptera sur Pierre sur le pont

(46) Marie verra Pierre à côté du pont

- Distinguishing complements and adjuncts

**1.** Complements cannot be iterated.

(47) *Marie compte sur Paul sur Pierre

(48) Marie travaille près de l'église sur le pont

**2.** Complements are selected: different verbs select different prepositions. One says that a verb like *compter* subcategorizes for a prepositional phrase (PP) whose preposition is *sur*.

**3.** Adjuncts are optional, complements may be obligatory (cf. *Marie compte).

**4.** The above phrase stucture rules predict that complements precede adjuncts — in fact complements are expected to follow the verb.

(49) a. Marie a lu le livre sur le lit.
b. ?Marie a lu sur le lit le livre.

**5.** Finer predictions

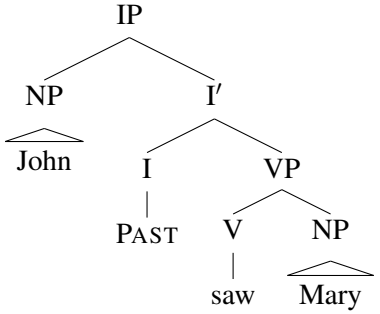
(50) a. Marie dormait sur le lit, et Pierre en faisait autant sur le canapé.
b. Marie dormait sur le lit, et Pierre en faisait autant.

(51) a. *Marie comptait sur Paul et Pierre en faisait autant sur Jacques.
b. Marie comptait sur Paul et Pierre en faisait autant.

## 6 The Inflection Phrase

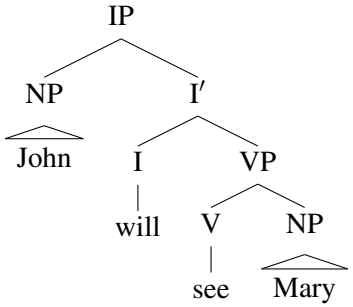
- Modern syntactic theories postulate an Inflection Phrase (IP, or sometimes InflP) instead of an S phrase.
- This addresses the embarrassment of having a phrasal constituent without a head. The inflection (tense, mood, person/number), seen on the verb, is the head category of a sentence.

(52) John saw Mary.




- In sentences with modals or auxiliary verbs, we can see phonologically overt material occupying the I node.

(53) John will see Mary.




- Importantly, certain facts about the placement of *adverbs* help us figure out where precisely in the tree the main verb is to be found.
- These facts can be different in different languages. English and French illustrate such a contrast:

(54) a. John often saw Mary.
b. *John saw often Mary.

(55) a. *John often will see Mary.
b. John will often see Mary.
c. *John will see often Mary.

(56) a. *Jean souvent parle avec Marie.
b. Jean parle souvent avec Marie.

(57) a. *Jean souvent va parler avec Marie.
b. Jean va souvent parler avec Marie.
c. ?Jean va parler souvent avec Marie.

# 7 Displacement

- Many pairs of sentences in English are intuitively "variations" on each other. They are related by displacement, in that one appears derived from the other by moving certain constituents around.

(58) a. John didn't read that book.
b. That book, John didn't read. (That other book, he did.)

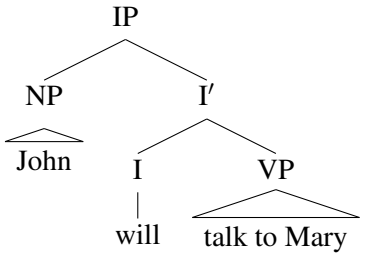
(59) a. Mary is very smart.
b. Is Mary very smart?

(60) a. John likes to talk to Mary.
b. Who does John like to talk to?

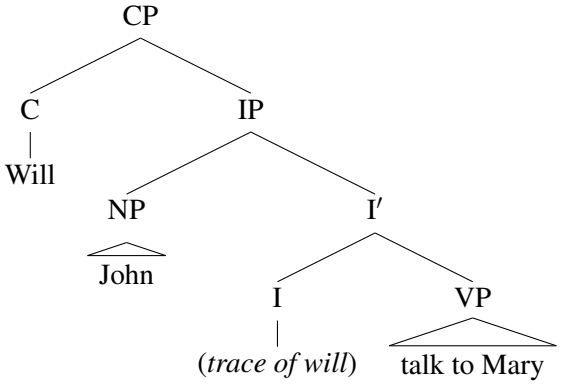
- In the first formal syntactic theories, the pairs in (58)–(60) were taken to have a common *D-structure* (for deep structure) but different *S-structures* (for surface structure).
- S-structures are the results of applying certain *transformations* to D-structures.
- (Contemporary syntactic theories in the Chomskyan tradition instantiate this intuition differently, taking a derivational approach.)

**A transformational rule — I-to-C**
To make a yes-no question from a D-structure with phonologically overt material in I:
1. Project a CP above IP and
2. Move I to C.

(61) a. John will talk to Mary.




b. Will John talk to Mary?




- What about D-structures *without* phonologically overt material in I?

(62) a. John likes Bill.
b. *Likes John Bill?
c. Does John like Bill?

(63) a. Mary talked to Phil about Sue.
b. *Talked Mary to Phil about Sue?
c. Did Mary talk to Phil about Sue?

- What think you of the following sentences? What tell they us about *be* and *have*?

(64) a. Mary is very clever.
b. Is Mary very clever?
c. *Does Mary be very clever?

(65) a. Mary has a brother.
b. *Has Mary a brother?
c. Does Mary have a brother?

(66) a. You have a lighter.
b. ??Have you a lighter?
(perfect in British English)
c. Do you have a lighter?

(67) a. Mary has seen a lot in her lifetime.
b. Has Mary seen a lot in her lifetime?
c. *Does Mary have seen a lot in her lifetime?

# 8 Syntactic processing

- Recall that, alongside the *theoretical linguistics* we've been concentrating on so far there is also *psycholinguistics*, which cares about **how** we make use of the knowledge of language that theoretical linguistics aims to describe and explain (see slides from the first lecture).

**Grammar**
Formal specification of structures allowed in a language. The object of study of theoretical linguistics.

**Parsing**
The process of determining the syntactic structure of an input string. A parser takes as input a string of words and produces as output a parse tree: a labeled bracketing of the input sentence. The object of study of psycholinguistics.

- We need a grammar in order to have a parser!
- **Bottom-up parsers**: Exclusively guided by properties of the input; they make no predictions about future elements of the parse tree.
- **Top-down parsers**: Start with strong assumptions about what typical structures look like and try to fit the input into the presupposed structures. They make a lot of predictions, but those predictions will fail many times and force the parser to backtrack, incurring important memory and time costs.
- Modern parsing theories use **hybrid parsers**, with bottom-up and top-down characteristics.

### Ambiguity

- There is ambiguity at different representational levels:
  - Lexical — a word may be rewritten by more than one category (e.g. "saw")
  - Lexical semantic ambiguity — a word may have more than one sense (e.g. "bank")
  - Structural ambiguity — a string may have more than one legal syntactic representation (more than one legal parse).
- Also important to distinguish:
  - Global ambiguity: no information in the sentence resolves the ambiguity.
  - Local or Temporary ambiguity: at one or more points during parsing of a string, there isn't sufficient information to determine the correct structure. Local ambiguity disappears when the entire string has been parsed.

(68) a. John saw a saw.
b. Mary loves banks.
c. Phil writes beautiful shopping lists and poems.
d. The athlete realized his goals were unattainable.
e. The man examined yesterday came back today.

**Garden Path phenomena**

- Temporary syntactic ambiguity can be resolved in an inappropriate manner before disambiguating information is reached.
- We say that we were led down the (wrong!) *garden path*.

(69) The horse raced past the barn fell.

(70) The man left the money
a. in the will invested it wisely.
b. on the table for the waiter.

(71) Since Jay often jogs a mile
a. seems like nothing to him.
b. two miles seems like a lot to him.

- The garden path theory of Frazier and Rayner explains this effect. The theory is
  - Serial — only one syntactic representation is built at any given time. If that fails, you have to backtrack or start from scratch.
  - Heuristic-based — it relies on two heuristics (problem-solving strategies / rules of thumb):
    1. Late Closure — Attach new items to the phrase or clause currently being processed.
    2. Minimal Attachment — Attach new items to the existing structure in a manner that requires establishing as few new nodes as possible.
- What do the heuristics have to say about the following two sentences:

(72) a. Since Jay often [$_{VP}$ jogs a mile] [$_{IP}$ two miles seems like a lot to him]
b. Since Jay often [$_{VP}$ jogs] [$_{IP}$ a mile seems like nothing to him]

- What about these two:

(73) a. [$_{NP}$ The man] [$_{VP}$ left the money on the table for the waiter]
b. [$_{NP}$ The man [$_{CP}$ left the money in the will]] [$_{VP}$ invested it wisely]