

# LING 101 : Introduction to Linguistics

Salvador Mascarenhas

Michael Goodale (TA)

21 Septembre 2023

## Table des matières

<b>I</b>	<b>Cours 1 21/09</b>	<b>3</b>
1	Remarks and Observations about the Nature of Language	4
2	The Goal of Linguistics	4
3	Different levels of Study	4
4	Language and Societies	5
4.1	Language and Classes . . . . .	5
4.2	Language and Dialect . . . . .	5
4.3	Infinity of Language . . . . .	5
4.4	Description . . . . .	5
4.5	Phonological Differences . . . . .	5
4.6	Internal Structure of Sentences . . . . .	5
<b>II</b>	<b>TA 1 27/09</b>	<b>5</b>
5	Animal Communication	5
<b>III</b>	<b>Cours 2 28/09</b>	<b>6</b>
6	Different yet Similar	6
7	Similarities between distant unrelated languages	6
7.1	Reciprocal pronouns . . . . .	6
7.2	Sensitivity to negative elements . . . . .	7
7.3	Relations between Sentences . . . . .	7
7.4	The Puzzle . . . . .	7
8	Universal Grammar	7
8.1	Chomsky's Hypothesis . . . . .	7
8.2	Universal Grammar . . . . .	8
8.3	Language Acquisition . . . . .	8
8.3.1	Critical Period . . . . .	8
8.3.2	Problem of Induction . . . . .	8
8.3.3	Absence of Negative Evidence . . . . .	8
8.3.4	Conclusion . . . . .	8
8.3.5	Chomsky's Argument . . . . .	8

<b>IV</b>	<b>TD 2 : 04/10</b>	<b>9</b>
<b>9</b>	<b>Homework 1</b>	<b>9</b>
9.1	Question 1 . . . . .	9
9.2	Question 2 . . . . .	9
<b>10</b>	<b>Exercise</b>	<b>9</b>
10.1	Swedish Extrapolation . . . . .	9
10.2	Rule Extrapolation . . . . .	9
<b>V</b>	<b>Cours 3 : 05/10</b>	<b>9</b>
<b>11</b>	<b>Use of Morphology</b>	<b>9</b>
<b>12</b>	<b>The Study of Morphemes</b>	<b>9</b>
12.1	Words . . . . .	9
12.2	Morpheme Types . . . . .	10
<b>13</b>	<b>Forming Words</b>	<b>10</b>
13.1	Derivational Processes . . . . .	10
13.2	Constituent Structure . . . . .	10
13.3	Compounding . . . . .	11
<b>VI</b>	<b>TA 3 : 11/10</b>	<b>11</b>
<b>14</b>	<b>Homework 2</b>	<b>11</b>
<b>15</b>	<b>Cross-Linguistic Variation</b>	<b>11</b>
<b>16</b>	<b>Morphological Derivation</b>	<b>11</b>
16.1	Deriving <i>Antidesestablishmentarianism</i> . . . . .	11
16.2	Deriving <i>unfriendliness</i> . . . . .	12
16.3	Deriving <i>reinstatement</i> . . . . .	12
<b>VII</b>	<b>Cours 4 12/10</b>	<b>12</b>
<b>17</b>	<b>Syntactic Competence</b>	<b>12</b>
17.1	Lexical Categories . . . . .	13
17.2	Constituency . . . . .	13
17.3	Rewriting Grammar . . . . .	13
17.4	Complements and Adjuncts . . . . .	15
<b>VIII</b>	<b>TA 4 : 18/10</b>	<b>15</b>
<b>18</b>	<b>Homework 3</b>	<b>15</b>
<b>19</b>	<b>Study of an abstract language</b>	<b>15</b>
<b>20</b>	<b>Structure of sentences</b>	<b>15</b>
20.1	John sees Mary . . . . .	15
20.2	John Sleeps . . . . .	15
<b>21</b>	<b>Heads, Complements, Adjuncts</b>	<b>16</b>
<b>22</b>	<b>Structure of Noun Phrases</b>	<b>16</b>

23 Teaser for next class	17
IX TA 5 : 25/10	17
24 Kinda like Homework	17
25 Grammar Generating a Language	18
26 On Disambiguation	19

## Première partie

# Language : Psychological and Social Entity

## Introduction

DEC is called like that because it would have bothered literatures Prof : Master at NYU, Junior Research at Oxford, here since 2016. Using language as a window into human mind

Linguistics is a broad term for serious principled study of language. Many perspectives, from the cognitive point or study language by an external perspective (structuralism : Ferdinand de Saussure, or Leonard Bloomfield) looking out on structures. Also quite general here. Language looked at as a social entity. History of languages, typology of languages. Not only about cognitive studies, pretty broad look out.

Teaching assistant : Michael Goodale PhD student, LRSCP, computational models of Language Acquisitions. Statistical inference and model language after formal tools. Practical Skills really.

Assessments : Homework, graded on a qualitative schedule, due on lecture days. Final : Last Lecture, 30% of grade 10% of grade in TA participation. Website : Moodle hosted by Université de Paris, Syllabus Syllabus :

## Schedule

1. Language as a psychological and social entity
2. Language (non-)variation : Universals, variation within parameters
3. **Morphology**, language typology
4. **Syntax I** constituent structure ; selection and subcategorization
5. **Syntax II** subcategorization ; X-bar theory
6. **Semantics I** first look at meaning *Studied by Salvador*
7. **Semantics II** philosophy of language and the case for methodological solipsism.
8. **Phonology** phonetic macro classes ;
9. Language and Reasoning
10. Sign languages (guest lecture)
11. Language in the brain ; deficit-lesion method ; functional brain imaging ; psycholinguistics ; parsing, reading, lexical access.
12. Language and thoughts in minds vs. machines

# 1 Remarks and Observations about the Nature of Language

R.Descartes "*Discourse on the Method*" : Humans, everyone of them, can speak. Animals, though they have what is needed, can't express their thoughts. Insights :

1. It doesn't matter on general intelligence, social intelligence or any measure of your intellectual abilities. Happens despite any other difficulties.
2. *To Our Knowledge* Not any other animal can do what we can do. Article from the *NYT*, saying animals **can** speak, though it's embarrassing. A Chasm *appears* between humans and animals. Yet, it is continuous of what happens in the animal kingdom.
3. Animals are not incapable of language because they can reproduce human language, or use signs to communicate. Studies on Non-Human primate vocalizations, 3-4 words, all alarm calls → Language is independent on organs and communication systems. Yet is it a panic reaction or a real communication. Cannot conclude on A.Is then...

## 2 The Goal of Linguistics

A complete understanding of how sound (/sign/etc) relates to meaning :

- in terms of the speaker's knowledge : the state their mind is in by virtue of having acquired a natural language (competence). Distinguished from mastery of language/way it is produced. Describing skill, not usage.
- in terms of using that knowledge in linguistic tasks, like uncovering meaning from sound in real-time comprehension ; executing motor commands necessary to externalize meaning in language production (performance)

Competence/performance is not really a sort of Chomsky. Chaz Firestone (Yale) published on competence/performance saying machines have been tested on performance and not competence. Tight connection between thinking and speaking. Behaviourism = school of thought that tried to figure out a way of studying humans that postulated and said absolutely about our mind. Not only linguistic behaviour : we shouldn't describe what happens in people's heads, just study what outputs comes from what inputs, without looking at the stimuli. Not what we will postulate. Freud sucks. Cannot look at functions : functions = algorithms, studied based on input/output pairs. Cannot do deduction nor induction but only abduction. Yet, we have no less reasons to believe it is right than to believe black holes exist.

## 3 Different levels of Study

Example : *Mushrooms are an edible fungus*

1. Sound Categories : Studying the sound signal based on the phonemes, represented in the mind.
2. Morphemes : first chunks of phonemes that has a meaning : Morphology. Here : *Mushroom* and *z* or *edi* and *able*. Sometimes they are not pronounced : need for a rigorous description. Sometimes, they're redundant, and appear with the same meaning in different places : compare theories. What is the probability of that happening ? And how about irregularities : Past = laugh-*ed* or gave ? Past in a concept that can manifest itself in different places : simple theory. FMRI theory can improve this theory. Morphemes don't always come in the same orthograph nor sound.
3. Words : Not much to do here.
4. Semantics : Organizing words into phrases. Here : *edible fungus* is a phrase, but *edible* is not. Must be done formally.

Three way of looking : Us, looking from a native's judgement - introspection **will** answer some questions. Exaggeration, yet : no written language, only looking at spoken language.

## 4 Language and Societies

### 4.1 Language and Classes

Language display depends only on human factors, political relationships, genetic factors : distinction between the animal and the meat (names coming from French, spoke by the upper class) in english. Happened in other languages. Different from hyperonyms like *poultry* for *chicken*. Context can explain linguistic aberrations.

### 4.2 Language and Dialect

Language and Dialect are political constructs and arbitrary decisions :

A language is a dialect with an army and a navy. (M.Weinreich)

### 4.3 Infinity of Language

Sentences are of arbitrary length, and can always be augmented. Yet infinite-ish sentences are impossible to comprehend because Performance is finite, i.e. cognitive resources are finite. There is a *finite* system generating infinitely many linguistic representations : recursions are of the order.

### 4.4 Description

Not looking for rules prescribing language (fuck l'académie française), but only for rules describing them. No better way to speak, the way you ought to speak has nothing to do with linguistics and only with politics. Yet, language are principled, even those which are *proscribed* : adding  *fucking* in the middle of a word : *Phila-fucking-delphia*. Rule here :  *fucking* comes before the stressed syllable and the material right before needs to be heavy. Heavy comes from phonology, rigorous theory of the weight of syllables.

### 4.5 Phonological Differences

Languages have different constraints on the syllables composing their words : *\*pnick* works in French but not in Engl\*sh.

### 4.6 Internal Structure of Sentences

Sentences are made of constituents that don't act up the same in every language : *des* is not used in Engl\*sh (*~ of the*). They cannot be broken : *des burgers et des frites*. It is mysterious tho ? Maybe language has something else to do for us than communicate. . .

They cannot be considered alone : *Fat cats eat* and *Fat cats eat accumulates*. Supposition that two words next to each other are related in written language. Also, prosody is a big help in understanding.

## Deuxième partie

## TA 1 27/09

## 5 Animal Communication

Language is also communication, not only hearing (trees ?). Many (if not most) Animals Communicate, and almost all react to sound. (cf. NYT Article) : dolphins communicating by signs, bees dancing, monkeys having muscles/organs to 'talk', bird songs, ant pheromones, great apes...

Differences between human and animal communication? Human language has : composition (recursion : meaning of sentence can be derived from its constituents), abstraction, no hypothetical/long term/prevention discussion, intentionality, arbitrary length of sentences, systematic neologisms/nonce words (when learning a word, it is usable immediately), non-instrumental.

Many experiments about teaching great apes language suck and were not really concluent. There is a poor, noisy, contradictive and unrepresenting stimulus that child have to make do with. Deaf child make their own language if they need one. For example of the poverty of the stimulus : 2 Layer Embedding of possessive happened 67 times in 120k sentences, while kids at 6 can do 4 level possessive embedding.

The words *stop*, *mat*, *tap*, *butter* all have 't' yet have different sounds : there is a sense where this is the same sound, but another one where they have different sounds.

## Troisième partie

# (Non-)Variation and Languages

## 6 Different yet Similar

Languges are made of signs, composed of a form and a meaning. They only look like they have multiple forms/meanings, on another level of analysis they only have one, e.g. past tense in english and the morpheme *ed*. Variants depend only and purely on properties of the root, and are entirely predictable. With *bat*, there is an ambiguity phenomenon, but there is also a phenomenon of polysemy, e.g. *book*, also, similar meanings often derive from a central point.

« A sign presents itself to the senses, and something distinct from itself to the mind »-  
St-Augustine

There are two types of signs based on the link between form and meaning :

1. Those with a causal link : *smoke* implies there is a fire
2. Those with a arbitrary, conventional link : *black attire* implies mourning (*in some cultures but...*)

After Ferdinand de Saussure,

« Language is a systemn of conventional (arbitrary) signs. »

Example : The word *man* in different languages : German Mann, Spanish hombre, Français homme, Hungarian ember, Turkish adam. With the sound produced by the rooster, the words differ in languages, but there is still a partial causal link, because you cannot mimic perfectly the sound of the animal, given the differences between vocal organs. Arbitrary doesn't mean random : it just doesn't matter what choice is made. It is no accident there is a resemblance between German and English words for *man*

## 7 Similarities between distant unrelated languages

### 7.1 Reciprocal pronouns

Pronouns marking reciprocity always have a mysterious constraint where they must be in the same, finite clause as their antecedents : *They thought I talked about each other* seems weird. Generally, it seems that reciprocal pronouns must refer to a thing that lies in the same proposition, but why?

First question is : Do we have a reason to say why they seem weird, just because they are longer? No, *I thought that they talked about each other*, is equally long as *They thought that I talked about each other* and doesn't sound half as bad.

Then, the sentence *They thought we talked about each other* shows it's not about third person. Coming up with a precise answer implies looking for phrases where each bit of the sentence has been replaced, one at a time, to isolate the *issue*.

In *They thought that I talked about each of them, each of them* is not necessary reciprocal, as it only include (*each other*)'s meaning so it is compatible.

## 7.2 Sentivity to negative elements

Words occur sometimes with negative elements. Every natural language seems to contain at least one lexical items that is sensitive to whether the context in which they occur exhibits negative or positive polarity.

Exemple : *Jean a fait le moindre effort* doesn't seem natural, but *Jean n' a pas fait le moindre effort* does.

Facts are subtle : the mere presence of a negative item is enough to license a negative polarity, and sometimes negative polarity is inferred without the obvious presence of a negative item. A sentence with an empty slot in place of the negative polarity item is the context that needs to be negative in order to license an NPI.

Positive polarity also exists : *John didn't see someone* is really weird, and it requires a really particular prosody and/or context to work. You have a meta interpretation of this. The word anyone, more than a negative item is also a free choice item.

## 7.3 Relations between Sentences

In all languages, declaratives and interrogatives are linked by *transformations* ; i.e. a reorganisation of the terms of the declarative to build the interrogative. It creates pairs of assignations, not necessarily questions and answers. Also some declaratives link to multiple questions.

There is a finite palette of strategies for these transformations : no known human language builds questions by mirroring completely the order of the words. Grammars do not count, e.g. there is no language with transformation swap words 1 and 2.

## 7.4 The Puzzle

Languages are systems of arbitrary signs. Any sign will do for internal calculation, and any convention widely known within a given community will do for communication. This observation does not predict the existence of strong systematic, pervasive similarities whose speech communities have had no contact.

How can the conventional character of language be reconciled with pervasive cross-linguistic similarities.

« We would learn so much if we could do horrible things to babies. We can do horrible things to animals though, not saying we should, but we can. »- Salvador Mascarenhas

# 8 Universal Grammar

## 8.1 Chomsky's Hypothesis

Human faculty for language :

- enables humans to acquire and use languages ;
- delimits what linguistic structures humans are capable of acquiring and using
- delimits the kind of linguistic conventions that a community of humans can develop and successfully hand down to new generations

Thus we need to know all about gene reproduction, biological evolution and so on, to understand it fully

Universal Grammar in this sense is part of all humans' biological endowment and it is reflected in all natural languages.

## 8.2 Universal Grammar

This is not a language in itself and it doesn't imply the idea all languages come from a common source. This does not imply there is no other factors of relevance shaping actual natural languages than *a priori* constraints, there being no rhyme or reason to those constraints. Chomsky added it might be to optimize computation. This is *not* a collections of generalizations about trends in linguistic diversity, other projects (Dryer, 2005) have been doing it, especially on word order. It has for purpose to classify the languages the human mind can learn and the *impossible* ones. It is a two steps projects : Finding facts, photographing the human languages, then deriving generalizations.

## 8.3 Language Acquisition

### 8.3.1 Critical Period

There is a critical period for human (and animal aswell) acquisition of language : post-puberty, acquisition is severely impaired, e.g. Genie, kid discovered in L.A. in 1970 at 13.5y. Any child can learn any language, it depends on features of the environment.

Acquisition doesn't seem to be full related to mathematical/intellectual/reasoning skills and so on... Learning a second/third language is totally different tho, it is correlated to musical and computational skills it seems, there even seems to be a purely linguistic talent. Almost nothing about adult phenological is due to purely biology and genes.

### 8.3.2 Problem of Induction

Grammars make predictions for infinitely many word sequences, yet the input if finite. Therefore, there is something not in the input is playing a key role in learning.

### 8.3.3 Absence of Negative Evidence

Experiments, can give negative information : they can show that under certain conditions, an outcome is *not* observed, but a child has extremely limited access to such negative data.

### 8.3.4 Conclusion

A child is far better at figuring their native language than a linguist. Linguistics is an empirical science not a fundamental one.

Linguistic data consist of judgements on utterances, grammatically judging strings of characters, truth-value judging sentences/scenarios. It can come from introspection, but introspection is limited : there is a limit to what one can infer, e.g. *I have more pictures of my kid on my phone than my dad ever saw me* isn't grammatical, but it is understandable. The unboundness of human mind might mean we can find sense in any sentence.

The better way is to speak with other linguists or conduct experiments on a large amount of naive participants. There is a risk of falling into delusion from thinking to much about the same sentence.

### 8.3.5 Chomsky's Argument

There is such a gap between what a child is exposed to and the sophistication of the acquired grammar the child must have expectations as to what a language can be, analog to the concept of triangle : You've never seen a triangle but you know what it is. Mathis Hademeyer

Universal Grammar is a set of principles with a number of free parameters.

Principles and Parameters is a method now limited to small studies on variations, mostly in Italy (just you and me). By this we mean the fact that language, unrelated, share a structure. In fact most languages only occupy a minuscule part of the mathematical space of possibilities for a language. Principles and Parameters is a way to cash out on this idea : the problem is figuring what values has a child gotten for its parameters, and how he can deduce from some principles, the structure of language.

The null-subject parameter : In Catalan, *he speaks* is said *Øparla*, there is no need for a pronoun,



contrary to French or English. How do children handle this situation? It seems that Italian children go through a phase where they omit subject when learning French.

## Quatrième partie

# TD 2 : 04/10

## 9 Homework 1

### 9.1 Question 1

The word *nous* is used in formal context, so it is expected that people use formal grammar in the whole sentence, hence the *ne* of negation is expected.

### 9.2 Question 2

You can drop the copula in AAVE when you can contract the verb.

## 10 Exercise

### 10.1 Swedish Extrapolation

Separate the sentences in groups of morphemes, then extrapolate the meanings based on the other sentences.

### 10.2 Rule Extrapolation

Language is Chikisaw or something.

## Cinquième partie

# Morphology : The Structure of Words

## 11 Use of Morphology

Do you spell kick-ass or kickass? Is it one or two words? This is not very well understood.

## 12 The Study of Morphemes

Morphology is the study of words, and morphemes. Morphemes are the smallest linguistic unit that makes sense, e.g. *a* or *I*. The sound has no meaning, but the morpheme has. It is the smallest entity with both a form and a meaning.

Words are formed by processes, so they have a constituent structure, they can be constituted from multiple morphemes.

They can be derived from rules of morphology. Languages have a huge range of variations in what they can do.

### 12.1 Words

A Word is an identifiable unit of phonology with a prosody (cf. phonology class).

TABLE 1 – (Simplified) Derivational Rules in English

Affix	Rule	Output
-able	Verb + -able	= Adjective
re-	re- + Verb	= Verb
un <sub>1</sub> -	un- + Adjective	= Adjective
un <sub>2</sub> -	un- + Verb	= Verb

Words have a structure : The *noun* (part of speech) *reuseables* is made of four morphemes : *re*, *use*, *able*, *s*. Here, the morphemes are, in order, a *prefix*, a *root*, a *suffix*, a *suffix*.

Words can be derived from other words also : *to invite* - *an invite*. They can then have an ambiguity in sense : is *reusable* a noun or an adjective.

## 12.2 Morpheme Types

Prefixes, suffixes, infixes (e.g. *abso-fucking-lutely*), circumfixes (e.g. *em-bold-en* - not a good example as historically it might have never been a circumfix) together are called affixes. Those morphemes are called bound morphemes, meaning they cannot appear by themselves (e.g. *mang* root of *manger* in French)

When looking at the derivation, the thing that is not an affix is called the stem. The root is the smallest stem.

Things called clitics can also come into words, e.g. *l'*, it has more information than another morpheme : *l'aime* has two concepts in it.

Roots and Stems are called open class morphemes, new instances can easily emerge or be invented, e.g. *blink-ing* would be easy to understand. Affixes are closed class morphemes, meaning new instances develop slowly.

Inflectional morphemes only add grammatical information. They can have the same sound shape as some derivational morpheme. Derivational morphemes on the other hand, create new concepts out of existing ones.

There is a universal attested word-structure : [[[root] [derivational affixes]] [inflectional affixes]]

## 13 Forming Words

Particular grammars make certain derivational processes available, which we can describe by means of rules.

### 13.1 Derivational Processes

A rule specifies the category of the input and the category of the output (cf 1). Affixes typically add further restrictions on stems.

### 13.2 Constituent Structure

From a derivational analysis of a word in terms of the processes that generated it, we can extract its constituent structure. We can represent it in a sort of phylogenetic tree.

For *un/enjoy/able* : *unenjoy* is not a word, as you cannot reverse enjoyment. Even though *un-see* exists, it is more of a creative product of language as you might not understand its meaning at first glance; it is understood because it is frequent. So we get a structure like : Node(un, Node(enjoy,able)). We might go deeper in analysis, but there is no real rule behind *en-joy*.

For *re/read/able* : We cannot base an argument on a stem that doesn't exist, because both hypothetical stems do. Yet, as *re-* does not work with an adjective, we can derive the structure to be :

Node(Node(re,read), able)

For *un/wrap/able* : Here, both Node(Node(un, wrap), able) and Node(un, Node(wrap, able)) are valid processes of derivation involving different rules. The second is the negation of the adjective *wrapable*, which means something that cannot be wrapped. The first on the other hand, is something that can be *unwrap(ped)*, since we add *able* to a verb. This is an example of structural ambiguity, that is often disambiguated by prosody e.g. *fat cats eat...* Yet, this not only allows to express ambiguity, but also to explain it.

### 13.3 Compounding

This is a quite mysterious area of morphology.

Compounding is a process to build words from two or more stems. It constitutes in combining those stems with a novel meaning, with lower predictability than derivational processes, e.g. *bitter/sweet*. There is compounding between, adjectives and adjectives, adjectives and nouns, nouns and nouns, nouns and verbs, verbs and verbs, verbs and nouns... There is a very blurry line between what people think and historical reasons, are people thinking of *sleep* and *walk* in *sleepwalk*?

It can be an ambiguous process : If a *houseboat* is a boat that is a house, what is a *housecat*? English compounds are typically headed by their rightmost element, meaning the precedents are qualifying it, but sometimes compounds are headless : *kick-ass*, *ceasefire*...

Are there more rules? Is an N-N compound necessarily a *for* relation, a *from* relation? Is there simply just a hidden preposition in it?

## Sixième partie

# TA 3 : 11/10

## 14 Homework 2

Nothing to add, see file.

## 15 Cross-Linguistic Variation<sup>1</sup>

Languages can be put on a spectrum from analytic to polysynthetic, based on how much they use morphemes and/or syntax. Analytic use the most syntax, polysynthetic don't care about it and only add morphemes.

- Analytic (Isolating) languages where each morpheme is a word on its own (e.g. Chinese)
- Agglutinating languages where each grammatical bit of meaning is an affix to a stem (e.g. Japanese, Korean, Hungarian)
- Fusional languages where an affix or change in the stem can lead to multiple grammatical meaning variations (e.g. Latin, German, Arabic)
- Polysynthetic Languages where affixes have extremely rich content (e.g. many native North American languages)

Both agglutinating and polysynthetic languages pack a lot of information into a single word, the difference being that agglutinating use affixes for grammatical meaning only.

## 16 Morphological Derivation

### 16.1 Deriving *Antidesestablishmentarianism*

We can see this word from a sequence of trees :

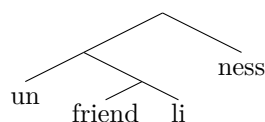
---

1. This is part of Lecture 3, see the handout

	Tree :	Explanation
Step 1 :	<pre> graph TD     A[ ] --- B[dis]     A --- C[establish]           </pre>	$dis : verb \mapsto verb$
Step 2 :	<pre> graph TD     A[ ] --- B[dis]     A --- C[ ]     C --- D[establish]     C --- E[ment]           </pre>	$ment : verb \mapsto noun$
Step 3 :	<pre> graph TD     A[ ] --- B[dis]     A --- C[ ]     C --- D[establish]     C --- E[ment]     B --- F[anti]           </pre>	People are against <i>disestablishment</i> , not the people
Step 4 :	<pre> graph TD     A[ ] --- B[dis]     A --- C[ ]     C --- D[establish]     C --- E[arian]     B --- F[anti]     D --- G[dis]     D --- H[ ]     H --- I[establish]     H --- J[ment]           </pre>	$arians : action \mapsto peopleinit$
Step 5 :	<pre> graph TD     A[ ] --- B[dis]     A --- C[ ]     C --- D[establish]     C --- E[ism]     B --- F[anti]     D --- G[dis]     D --- H[ ]     H --- I[establish]     H --- J[ment]     I --- K[ ]     K --- L[arian]           </pre>	$ism : people/concept/action \mapsto idea$

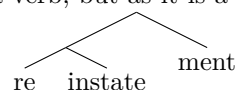
## 16.2 Deriving *unfriendliness*

We only give the tree :



## 16.3 Deriving *reinstatement*

You cannot attach *in* to a noun, it needs to be attached to a verb, but as it is a rare example of borrowing from latin, we cannot separate *instate*



## Septième partie

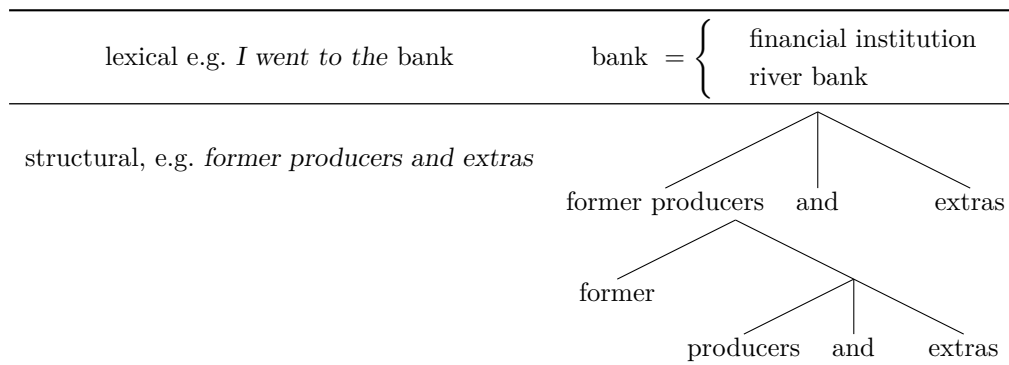
# Syntax 1

## 17 Syntactic Competence

Syntactic competence is a wide concept, it includes :

- The knowledge of what sentence is grammatical and what isn't. This comes from the fact there is no syntax without semantics in linguistics, unlike in arithmetics. The idea of an autonomous syntax has disappeared since the 70s. It is however useful to talk about syntax in a somewhat disconnected way from semantics. For example, you cannot contract auxiliaries when they carry meaning : *I shoulda bought it* and *\*I shoulda more money*.

- An ability to recognize a sentence as well-formed even if it does not make sense or consists of highly sequential transitions, meaning, even though some transitions have really low probability, e.g. *Colorless green ideas sleep furiously* and *Colorless Green* or *sleep furiously*...
- An ability to determine what the sentence means, whether it is ambiguous or not either structurally or lexically :



- Every language has a countably infinite set of sentences (proven by showing there is no longest sentence in a language). Native speakers can handle this set despite having finite minds. We can construct arbitrary long sentences based on the center embedding idea, i.e. *le jeu du Johnny Depp*.

## 17.1 Lexical Categories

The lexical categories (or parts of speech) such as *N(oun)*, *V(erb)*... This does not mean you can only tell what part of speech a word is if you know its meaning.

« *The thing about thing is that everything and anything is a thing* »- Salvador

**Définition 17.1.1.** We define lexical categories given the sentential environments in can occur in and the affixes it can take, i.e. the grammatical distribution of the word.

« *You shall know a word by the company it keeps* »- Ferguses

For example, a verb in English satisfies all of the following :

1. It can occur right after the auxiliary *will*
2. It can take the endings *-s* and *-ing*

## 17.2 Constituency

**Définition 17.2.1** (A Heuristic). A sequence of words forms a constituent in a sentence if :

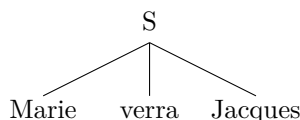
1. It can be replaced with a minimal unit, ideally one word, preserving grammaticality and conversely
2. Occurrences of that minimal unit can be replaced by the sequence of words, preserving grammaticality

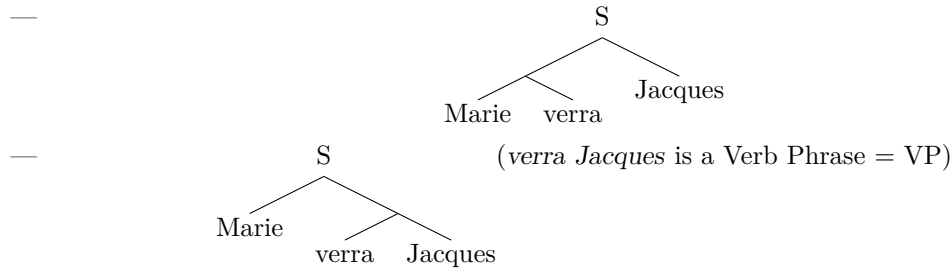
If those conditions hold, then the sequence has the same category as the minimal unit.

A constituent can also be found by Fronting and Pronominalization, Clefting, Eliding or Coordination

## 17.3 Rewriting Grammar

*Marie verra Jacques* leads to three possible structures in principle :





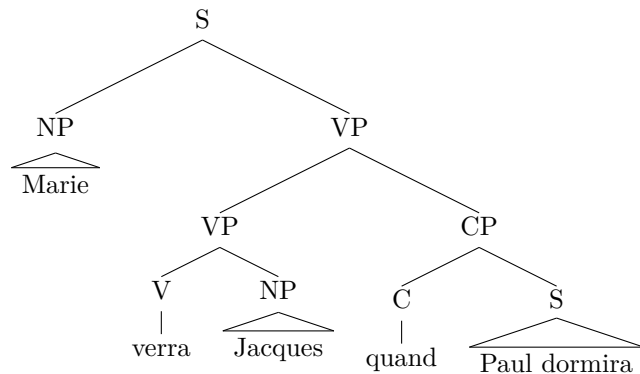
We might build a rewriting grammar for a fragment of French :

1. *Marie verra/entendra/...Jacques*
2. *Un type verra Jacques*
3. *Marie verra un type*
4. *Le type connu a vu Marie*
5. *Marie dormira*
6. *Marie pense que Paul verra Jacques*
7. ...

We then can get an idea on French Grammar :

Id	Node	→	Rephrased as
0	S(entence)	→	N(oun) P(hrase) + VP
1	NP	→	Proper Name
2	NP	→	Det + N + (Adj)
3	VP	→	$V_{intr}$
4	VP	→	$V_{tr}$ + NP
Up to this point, we cannot generate an infinite number of sentences			
5	C(omplementizer) P(hrase)	→	<i>que</i> + S
6	VP	→	$V_{cl(ausal)}$ + CP
7	$V_{tr}$	→	voir, entendre,...
8	$V_{intr}$	→	dormir, briller,...
9	$V_{cl}$	→	penser, croire,...
There is now a loop in rules, so there is recursion and we can generate an infinity of sentence			
10	$C_{temp}$	→	avant que, quand
11	$CP_{temp}$	→	$C_{temp}$ + S
12	S	→	NP + VP + ( $CP_{temp}$ )
	VP		a) $V_{intr}$ + ( $CP_{temp}$ ) b) $V_{tr}$ + NP + ( $CP_{temp}$ )
14	S	→	S + $CP_{temp}$
15	VP	→	VP + $CP_{temp}$

We can then infer a tree from a sentence with the lexical categories :



## 17.4 Complements and Adjuncts

Complements and adjuncts are ways to add things to a constituent. We get new grammar rules :

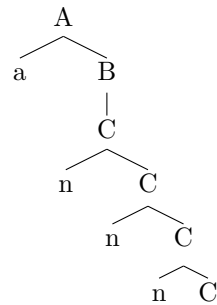
## Huitième partie

TA 4 : 18/10

## 18 Homework 3

Nothing to say.

## 19 Study of an abstract language

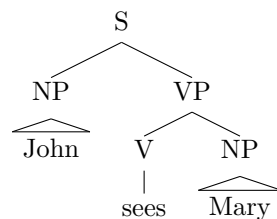
$$\begin{array}{l} A \rightarrow aB \\ B \rightarrow c \\ B \rightarrow z \\ B \rightarrow C \\ C \rightarrow nC \\ C \rightarrow d \end{array}$$


This gives us many strings possible :  $ac, az, an^*d...$

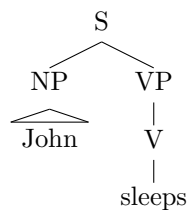
We can draw trees from those strings, e.g. *annnd*, see above.

## 20 Structure of sentences

## 20.1 John sees Mary

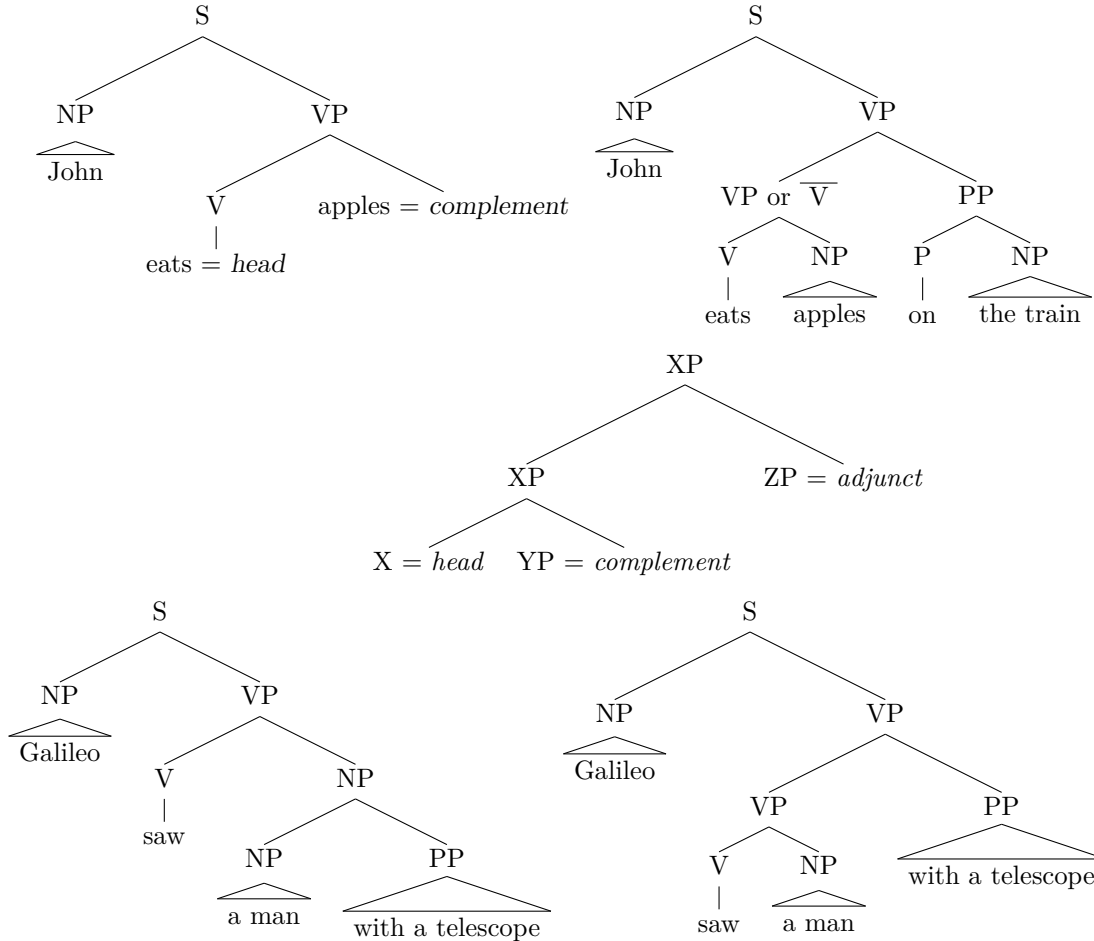


## 20.2 John Sleeps



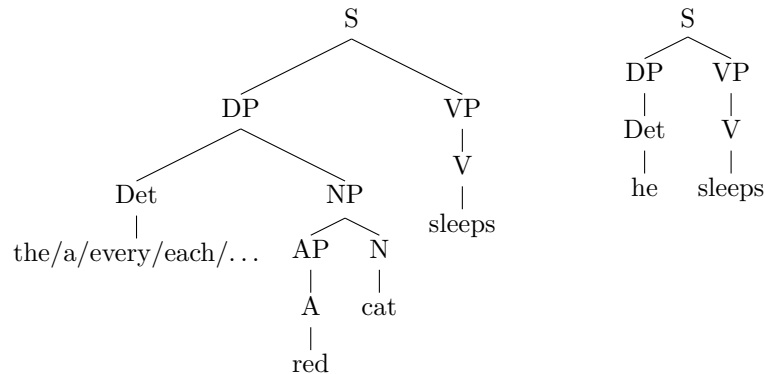
## 21 Heads, Complements, Adjuncts

**Définition 21.0.1.** A head is a word. A head is connected to a phrase. Anything that is the sister of a head is a complement. When a complement is unnecessary, it is called an adjunct.



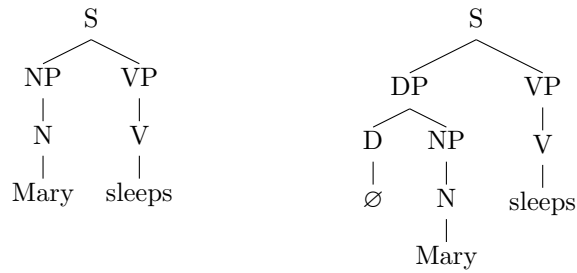
## 22 Structure of Noun Phrases

To analyze Noun Phrases with adjectives, we can base on the model from the first example below, introducing Determiner Phrases. We can do the same with verb phrases, seeing pronouns as determiners (this is called the DP hypothesis) :



We may guess that in the sentence *Mary sleeps*, *Mary* is a determiner, yet we could say *every Mary sleeps*. There are two ways of seeing the sentence *Mary sleeps* : either *mary* is a noun phrase or *Mary* is a determiner phrase with an empty determiner.

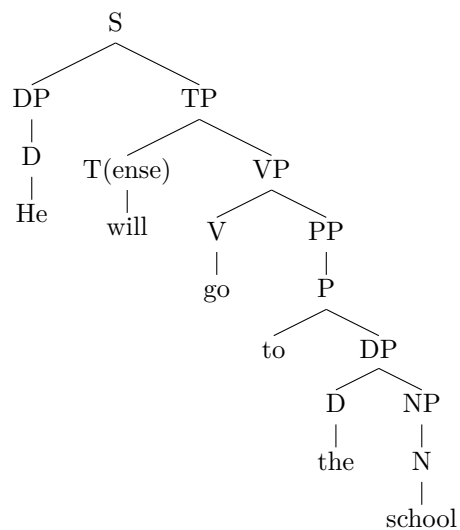




The latter is better from the study of other languages such as Portuguese where the determiner is not empty : *a Maria dorm.*

## 23 Teaser for next class

Consider *He will go to the school* :

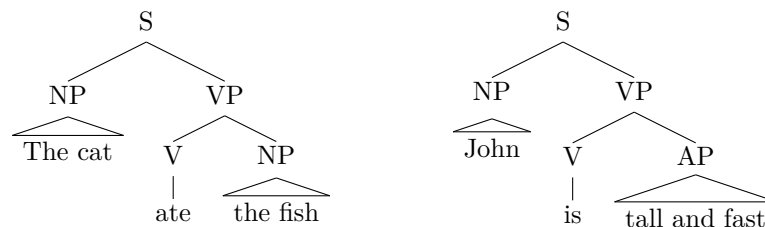


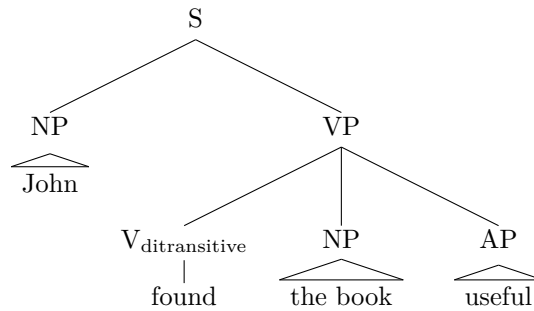
We here only did binary branching, but there might be ternary branching sometimes.

Neuvième partie

**TA 5 : 25/10**

## 24 Kinda like Homework





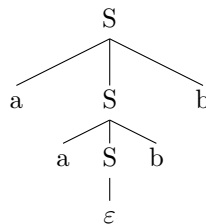
We could leave the VP in the last example with only two children by leaving the topological order. We would need to implement inside the heads a rule to tell the order of pronunciation. Another way to achieve this is by saying *the book useful* has secretly a more complex structure such as *the book to be useful*.

## 25 Grammar Generating a Language

Language :  $L = \{a^n b^n \mid n \in \mathbb{N}\}$   
 Grammar :

$$S \rightarrow aSb + \varepsilon$$

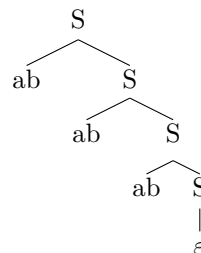
We get, for *aabb* :



Language :  $L = (ab)^*$ .  
 Grammar :

$$S \rightarrow abS + \varepsilon$$

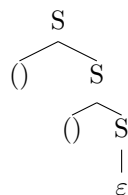
For *ababab* :



Language : Well-Parentthesized Expressions.  
 Grammar :

$$\begin{aligned} S &\rightarrow \varepsilon \\ S &\rightarrow ()S \\ S &\rightarrow (S) \end{aligned}$$

For *()()* :



## 26 On Disambiguation

*John looks like Mary* might mean *John looks in the same manner as Mary*. Syntactician do not care about priority rules, they don't want to disambue their grammar. They just add probability to their production rules, to generate parses probabilistically.