

Introduction to Linguistics

Syntax 1.

1. The theory of grammar

In the first week:

- **prescriptive** grammar: sets down guidelines for how people **should** speak or write
- **descriptive** grammar: attempts to describe how people actually speak

Rules of descriptive grammar: a list of certain properties of the language in question

e.g., English: subject usually comes before the verb and the object usually comes after.

most present tense verbs have an -s ending with 3rd person singular subjects.

John eats scrapple. Loretta walks to school.

Descriptive grammars of unfamiliar languages - indispensable for serious linguistic work.

- **explanatory** grammar: attempts to describe **the system of knowledge** that underlies how people speak.
- E.g., to explain why the object sometimes comes before or the subject after the verb in English:

That guy I don't like .

Down goes Frazier! (Howard Cosell's call of George Foreman knocking out Joe Frazier in 1973)

What is a theory of grammar?

A grammar is the **complete** system that underlies a language, not just a few facts about it

- grammar = (subconscious) knowledge that a speaker has of their own language a grammar
- grammar = our theory of what that knowledge might be like

Goal: get our theory of grammar to approximate the grammar in the speaker's mind.

Analogy: car specs for customers vs diagrams for mechanics.

Last several weeks: some fragments of a theory of English grammar (*sound and word structure*).

E.g., **Phonological rules** for flapping and raising, and the interaction of these rules

the kind of description that tells us how language-sound is put together out of a series of smaller parts.

Morphological rules for affixation tell us how words are put together out of morphemes

The syntax: how to generate an infinite number of new sentences from a finite set of words (lexical items)

2. Background: why syntax?

- Language is not just a jumble of words thrown together, but involves rigid structures and rules of combination. Why?
- Several words together provide more information than one word alone, e.g. top 10 words:

"College: the Undergraduate Experience:" undergraduate faculty campus student college academic curriculum freshman classroom professor.

"Earth and other Ethics:" moral considerateness bison whale governance utilitarianism ethic entity preference utilitarian.

"When Your Parents Grow Old:" diabetes elderly appendix geriatric directory hospice arthritis parent dental rehabilitation

"Madhur Jaffrey's Cookbook:" peel teaspoon tablespoon fry finely salt pepper cumin freshly ginger

To understand a recipe, we need more exact information about how the words (and ingredients!) combine.

Syntax rule? - Semantically-related words will be closer together than unrelated ones.

- **Francine, pizza, eat** ok
- **Pat, Chris, glass, break, saw** – two interpretations possible:
"Pat saw Chris break the glass." "Chris saw Pat break the glass."
- We need **principles for the combination** of these words to distinguish these meanings.

Similar: **word-for-word gloss of a passage in a language we don't know:**

- **Kashaya**, an American Indian language of northern California:

tiiqa mito taqhma c'ishkan shaqac'qash

I wish you dress pretty might wear

"I wish you might wear a pretty dress"

- But in other cases, **knowledge of the syntax is essential.**

muukín' tito 'ama dút'a' dihqa'khe' dúucic'iphi t'o daqaac'i'ba

he him job will give if know would like

A speaker of English might be **inclined to interpret this** as: "He will give him a job if he knows that he'd like it."

- In fact it is: "He would like it if he knew someone was going to give him a job"
- The only way to be sure about this is to know Kashaya syntax.

Syntax:

- a set of procedures and rules for putting words together in consistent ways to represent ideas in predictable ways.
- Both the speaker and the hearer have access to the same rules, so understanding is assured.

Rules can vary from language to language, e.g. a language could put a verb in front of its arguments (English) or behind them (Kashaya), as long as it systematic.

- English examples of what syntax does for us: grammatical nonsense

Colorless green ideas sleep furiously. (Chomsky 1965)

- Also: ungrammatical sense:

This child seems sleeping.

Harris dined the steak.

Neurological agrammatism

Broca's aphasia:

- **cannot speak fluently**, tend to omit e.g. articles and verbal auxiliaries, and sometimes can hardly speak at all. Comprehension, by comparison, *seems* relatively intact.
- **But:** their ability to understand sentences turns out to be deficient in systematic ways.
 - They always do well when only one sense is plausible:

It was the mice that the cat chased

The postman was bitten by the dog

- However, if more than one semantic arrangement is **equally plausible**:
It was the baker that the butcher insulted
 or if a syntactically wrong arrangement is more plausible:
The dog was bitten by the policemen
 then they do not do so well.

Clearly Broca's aphasia has a **negative impact on the processing of syntactic structure**.

So how does the syntax of human languages actually work?

Main characteristics of syntax: **A discrete combinatory system**

3. Parts of syntax.

Very roughly speaking, (Chomsky 1967) the syntax of a language consists of

- A lexicon – set of words;
- a set of phrase structure rules that tell us how to put lexical items together;
- a set transformation operations – movement rules – that tell us, e.g., how to arrive from a declarative sentence to the corresponding question

(john arrive → who arrived?; john is here → is john here?); movement rules also tells us to get a passive sentence from an active one – (john hit bill → bill was hit by john)

- much of the above has changed as syntactic theory progresses. The three broad parts of syntax – the lexicon, rules for putting words together, and movement – have remained

4. The lexicon

- simply put, the lexicon is a bunch of words which are the building blocks of sentences
- the lexicon is divided into two broad categories:
 1. **lexical words** and
 2. **function words**.
- Lexical words are an **open class** – nouns, verbs, adjectives, and adverbs. We can always add new ones to this set of words simply by making them up.
John blicked the ball.
- Function words are a **closed set** – articles and other determiners, pronouns, negation, numerals, – no new ones can be created, no matter how imaginative or poetic you may be.
John took the book frunder the bed meaning “from under”

Possible exception – prepositions – they often have a lot of semantic content, so we will treat them as a **lexical** category in syntax. But they are definitely a **closed** class – no new additions.

Lexical vs. functional categories

What is a lexical category (lexical word/ item) and how do we know what category a word belongs to? What is a noun, verb, adjective?

Otto Jespersen (end 19th century): **The definitions [of the parts of speech] are very far from having attained the degree of exactitude found in Euclidean geometry. [Philosophy of Grammar, 1924]**

Operationalism: any concept involved in "experimental reasoning concerning matters of fact and existence" must be defined in terms of **clearly defined operations**, to be performed in any given situation in order to decide whether an instance of that concept has been observed.

- **Non-operational:** "a noun is a person, place or thing"
- Operational definition must be precise, exact, deterministic and without exceptions, and must be in terms of how the given entity is related to processes of the system.

Zellig Harris (a key figure in American structuralist linguistics, Noam Chomsky's teacher) undertook to provide operational definitions of core linguistic concepts such as "phoneme," "morpheme," "noun," "verb".

- The basis for his definitions - observations of utterances
 - a mass of sentences that people had actually written or said
 - tried to define things like "noun" based on how they appeared in that mass of data.

Consider: "*the frump gromed the dron from out of the croled krune*"

Now consider "*the grome frumped*"

"the krune grome frumped the dron out of the crole"

... you get the idea...

- What are the categories of the novel words?
- How do you know them if you don't know the meaning of each individual word?
- How do you know that "frump" is a noun if you don't know whether it is a thing or a person or a concept?

Does the category of 'grome' and 'frump' change in the above sentences? Why is this happening?

- there is no unanimous agreement in the field as to what defines a category of a word,
- BUT: the functional words/elements surrounding a word (endings, articles/determiners) play a crucial role in letting us know what category a word is.
- consider a more radical example:

The blah blah blahes and blahas

OR:

Buffalo buffalo buffalo buffalo buffalo. - this is a grammatical sentence of English

Noam Chomsky: took up the task of providing operational definitions for the concepts of syntax (sentence structure) -- *noun, verb, sentence* and the like.

- He realized that he was working on the problem of **language learning**.
 - Figuring out grammatical concepts from a set of utterances: an idealized or abstracted version of what children accomplish in learning their native language.
 - Children must figure out
 - what types of words there are and
 - what the rules are for combining them into sentences. (later in the course)
- Chomsky's second innovation: the basic mathematics of syntax
 - along with collaborators like the French mathematician Marcel-Paul Schuetzenberger.

- We won't go into the details of this here, but the important thing is:
 - Chomsky adopted a very specific definition of a language.
 - For him, it was just a **typically infinite set of strings of symbols**.
- E.g., in a human language the strings=sentences, set of symbols=words.
 - English = all grammatical sentences that can be made from English words.

Below, we'll put the word "language" in scare quotes when we mean it in this abstract mathematical sense.

NOTE: This perspective ignores the question of meaning, and focuses exclusively on form.

Chomsky's revolution: **generative grammar** - our theory of grammar should be a formal "recipe" that can "generate" all the possible sentences of a given language, and nothing else.

- Syntacticians aim to answer the question:
 1. what do we know when we know that a sentence is grammatical or ill-formed in our language?
 2. How are we able to understand and generate an infinite number of new sentences

So a **generative grammar** of English should be a **set of rules** (or processes or whatever) that can produce sentences like *John likes pizza* but not *John pizza likes*

- **Syntax** is concerned with discovering and explaining **the hierarchical structure of sentences** in the human language;
- What are the rules that underlie sentence formation and structure; which rules are **universal** and which are **language specific**?
- In this regard, we can say that the grammar we aim to achieve is **generative and explanatory**. - the mechanism by which we can make new good sentences in a language.
- It is distinct from being a *descriptive* grammar concerned with descriptions of surface appearance of sentences: the rules we want **explain why** a sentence is grammatical.
- It is also distinct in an even more crucial way from being a *prescriptive* grammar

Other things we know when we know a word

In addition to knowing the category of a word we also know:

- Its **subcategorization** frame, i.e what other words it can combine with syntactically. This is especially relevant to verbs: **John devoured vs. John ate*.
- The word 'devour' subcategorizes for a noun object, while the word 'eat' does so optionally.
- We also know the word's semantic selection: we know which words can combine with other words given their meaning. For example we can say "John sleeps" but not "#This idea sleeps" even though the structure is grammatical syntactically
- Function words and morphemes tell us the category of the lexical words
 - A **noun** is a word that can follow a **determiner** (article, quantifier, numeral)
- Function words and inflectional morphemes can change subcategorization of a word
 - A **singular count** noun must have some determiner. But a **plural count** noun does not have to have a determiner (though it can have one)

5. The rules for building sentences out of words

5.1 Sentences are not word strings

So the syntax puts words together in specific orders, but how does it do so?

- Stringing words together, like beads on a string?

- OK for simple tasks (e.g. reading of recorded phone numbers, even parts of grammar)
- but such grammars are fundamentally unable to do the whole job.
 - The reason: such “generators” have **no memory**.
 - Choice of words from the first column cannot affect choice from the third column.

If Andrea hadn't eaten the fish, then she wouldn't have gotten sick.

By the time that we get to *fish*, there are at least three dependencies we need to keep track of in order to finish off the sentence properly:

1. The word *if* requires a **following clause** where the consequence (that will follow if the condition in the *if*-clause is fulfilled) is laid out. This clause can only be introduced by *then* or nothing, not *that*, *when*, *because* or the like.
2. Because *Andrea* is a woman's name (at least it usually is in the U.S.), **the pronoun** in the *then* clause that refers back must also be feminine, in this instance *she*.
3. The **past tense and irrealis mood** of the verb *had* in the *if* clause restricts the tense and mood of the *then* clause to a subjunctive with *would*.

5.2 Constituent structure

Like every level of linguistic structure, the syntax is **hierarchically organized**.

- A sentence is not composed directly of words, but of word groups, called phrases.

Creating a question in English: take the **auxiliary verb** and move it to the front of the sentence.

the dog is in the yard → is the dog __ in the yard?

the dog that is in the yard is named Rex → *is the dog that __ in the yard is named Rex?

the dog that is in the yard is named Rex → is the dog that is in the yard __ named Rex?

- Which “is” moves is determined by **how the pieces of the sentence fit together**, i.e. The **constituency** of the sentence.
 - **the subject of the sentence might be complex, potentially containing one or more verbs of its own** (in relative clauses that modify a noun).

[my dog] is named Rex

[the dog you just saw] is named Rex

[the dog that is in the yard] is named Rex

[the dog whose owner was arrested yesterday by the police for using him in a drug-running scheme] is named Rex

- All these sentences have the same structure except for the contents of the subject.
 - question formation only concerns with the level of the sentence,
 - not with what is inside the subject.
 - Once we know that *the dog that is in the yard* is the subject, it is treated as a single unit,
 - and the question **rule looks past it** to find the next verb.
 - If sentences were just strings of words, this would be impossible.
- The second part of syntax, phrase structure rules combine words into these groupings / units called **constituents**.
 - Words are not strung together in a linear fashion like beads on a string. They are hierarchically grouped with some words belonging closer together than others.

- **Hierarchical structure pervades language:** at the phonological (syllable), morphological (word), and sentence level.

5.3 From words to phrases

- Constituents generally are phrases built around **lexical categories** or, more traditionally, parts of speech.
 - The phrases built around different lexical categories have some common properties.
 - E.g, a **noun** is a single word (or a compound).

N
|
dog

- A **noun phrase** is a constituent built around a noun
 - (which is what we call the **head** of the phrase),
 - with other elements such as determiners and adjectives, or a relative clause or other modifiers.
 - Importantly, **every NP has a head** – the noun.
 - It may or may not also have a complement or a modifier.
 - E.g. *The man with the telescope.*
 - The noun “man” is the head (determines the category of the phrase).
 - “with the telescope” is a modifier. It is optional.

- The same can be done with the verb:

John runs **John runs and screams** **John sees a fat white cat**
John put the fat white cat on the floor

- Verb Phrase (VP) constituent can consist of only a single verb,
- it may also have one or more complements (an object, an indirect object).

- **All phrases have heads.**
 - English is a **left-headed language**,
 - other languages may have the head on the right (e.g. in Kashaya, verb is right of object).
- A noun phrase as well as a verb phrase are constituents.

5.4 Constituency tests

- an NP is a constituent:
 - we can replace a very long NP with a single pronoun “it”/”she/they”.
 - So, for syntax, even though the NP is very long, it occupies only one – a single – syntactic position.

The small cute cat with a striped tail and a funny pink nose runs = He runs

- Verb phrases (VPs) are constituents as well.
 - Verbs that take direct objects (transitive verbs) form a constituent with the object, but not with the subject:

Ex. John sees the big fat white cat that runs around and Bill does too. / so does Bill

Does too/ so does = sees the big fat white cat that runs around

- we can replace the entire VP with a VP-“pronoun” such as “Does too”
 - ◆ it replaces not just the verb but V+Object.
 - ◆ There is no such pronoun that would replace V+Subject
 - even though it is conceivable in theory:

* *John sees the fat cat and so does the dog* =/= *John sees the dog as well*

There are several ways to determine whether a string of words is a **constituent**, i.e. a coherent grouping of words functioning as a syntactic unit.

Test 1: Stand-alone in answers All the subjects in the sentences about Rex can **stand alone as answers** to the question, *Who is Rex?*

Who's Rex? My dog! That dog! The dog you just saw! The dog that's in the yard!
The dog whose owner was arrested yesterday by the police for using him in a drug-running scheme!

In contrast, many strings of words are **not able to stand alone** when we **question the original sentence**, and they are therefore probably **not constituents**.

***The dog you just!** ***The dog that's in the!** ***The dog whose owner!**

Test 2: Pro-form/Substitution A good test for constituency that we just saw is whether a **pro-form** -- that is, a **pronoun** such as *it* or *them*, or the **pro-verb** *do*, *do so*, *do it* -- can replace the string of words.

[the dog you just saw]_{NP} is named Rex [He] is named Rex
I gave [the book]_{NP} to Pat I gave [it] to Pat
Did you [give the book to Pat?]_{VP} Yes, I [did] Yes, I [did it] already Yes, I [did so]

The pro-form **replaces an entire constituent**.

For some types of phrases, there are no pro-forms, but we can sometimes substitute a single word instead:

The [very awfully big]_{AP} dog runs The [scary] dog runs

The test fails some non-constituents, and you should be careful when the result of substitution seems to work: do all the remaining words still play the same role?

I saw [the very awfully big dog in my neighbour's house]_{NP} run
*** I saw it in my neighbour's house run**

I saw [the very awfully big dog in my neighbour's house]_{NP}
I saw it in my neighbour's house
- *WRONG ROLE FOR "in my neighbour's house"*

Test 3: Movement/clefting Another test for a constituent is whether it can **move as a unit**. An example is the construction called a **cleft sentence**.

It's [my dog] that's named Rex It's [the dog that is in the yard] that's named Rex

Examples for **I gave the book to Pat** (the underline indicates where the words are from).

It's [the book] that I gave to Pat It's [to Pat] that I gave the book
It's [Pat] that I gave the book to *It's [the book to Pat] that I gave
***It's [gave the book] that I to Pat *It's [gave to Pat] that I the book**
It's [give the book to Pat] that I did

In the last example, a **pro-form** substitutes for the verb phrase, which of course is another test.

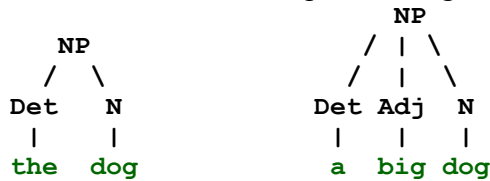
5.5 Phrase-structure rules

Recap:

- Words in a sentence are grouped into units called constituents.
That is, some words belong closer together than others.

Tests to determine if words form a constituent (stand-alone, substitution by a word or pro-form, move-together)

- Parts of a sentence = phrases; e.g. subject of a sentence is NP, [noun phrase], not just a noun.



- Tree diagrams like these **encode** constituency
 1. They are created (in the speaker's mind, or in a computer model of the speaker's knowledge -- by means of **general principles** called phrase structure rules.
 2. These determine **what kinds of sentence structures are possible** in a language.

The **phrase structure rules** that permit these configurations are as follows:

NP ---> **Det N** **NP** ---> **Det Adj N**

Need better rule: **the dog** **the big dog** **a big brown dog** **big angry brown dogs**

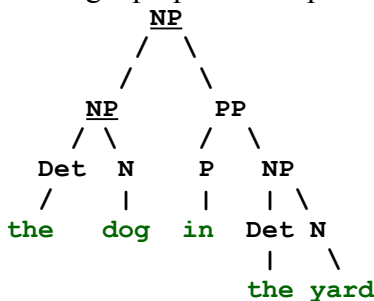
To **unify these different types of noun phrase**: **NP** ---> **Det Adj* N**

The asterisk means that an NP consists of a determiner, **any number** of adjectives, and a noun.

One more tweak: need an **Adjective Phrase** rather than just an adjective:

[**a**]_{Det} [**very big**]_{AdjP} [**slightly brownish**]_{AdjP} [**dog**] So, correcting the rule again: **NP** ---> **Det AdjP* N**

Adding a prepositional phrase as in *the dog in the yard* :



The phrase structure rule for this structure is rather simple: **NP** ---> **NP PP**

Key property of language: a single category can appear on both sides of the arrow of a phrase structure rule.

- This is called **recursion**
- It is the main way that a finite grammar can derive an infinite number of structures.

"Spelling out" the **prepositional phrase**: **PP** ---> **P NP**

Because we already have a rule for NP, we can **apply the rules in succession**

NP ---> **NP PP** **NP PP** ---> **NP P NP** **NP P NP** ---> **Det AdjP* N P Det AdjP* N**

Together these rules will produce the structures for complex NPs such as:

[**dogs**] [**on tables**] [**the young brown dog**] [**under the big green tree**]

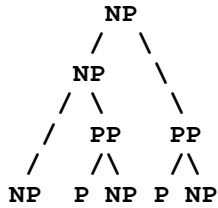
We can also apply the rules **successively** to produce more than one PP.

NP ---> **NP PP** **NP PP** ---> **NP PP PP** **NP PP PP** ---> **NP P NP PP**

NP P NP PP ---> **NP P NP P NP**

- This yields the following tree.

- Follow the top node (NP) down to its dependents (NP & PP); you see the effect of the first rule;
- keep going down to see the effect of the remaining rules.
- The tree structure is **built** by the rules.



After we get to the **individual words that make up the NPs** at the bottom of this tree, e.g.
[the dog] [under the tree] [with brown fur]

But suppose we **apply the rules in a different way**:

NP ---> NP PP

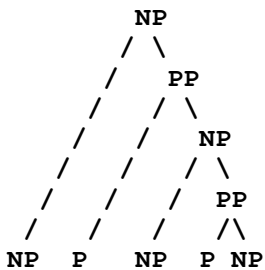
NP PP ---> NP P NP

NP P NP ---> NP P NP PP

NP P NP PP ---> NP P NP P NP

Now we get a **different tree**:

A **phrase** matching this structure is the following:
[the dog] [under [the tree with dead leaves]]



These two sentences: **same linear order** of elements but **different constituency**.

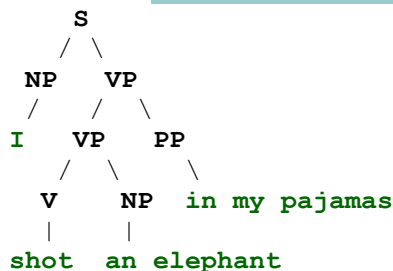
Structural ambiguities.

Groucho Marx: **One morning I shot an elephant in my pajamas. How he got into my pajamas I dunno.**

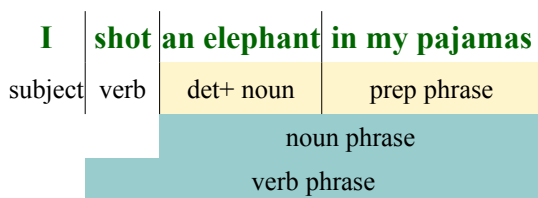
I	shot	an elephant	in my pajamas
subject	verb	noun phrase	prep phrase

"I shot an elephant **while (I was)** in my pajamas." || I [bought a book [with my credit card]].

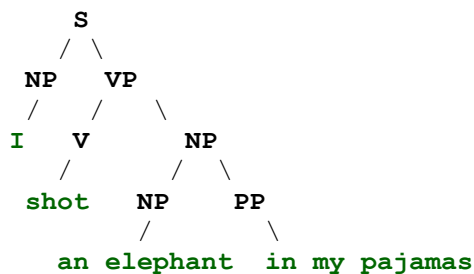
I	shot	an elephant	in my pajamas
subject	verb	noun phrase	prep phrase
verb phrase		verb phrase	



"I shot an elephant **that was in** my pajamas." || I bought [a book [with a red cover]].



This **closer structural connection** encodes the idea that the PP modifies **elephant** & not the act of shooting.



- The ambiguity comes from possibility of 2 orders in which the phrase structure rules apply.
- Again, a grammar that consisted of simple word strings would be incapable of capturing this. Why does the existence of sentential ambiguity indicate that there is hierarchy?

Formal universals:

- recursion
- hierarchical structure, and
- the **generality** of the phrase structure rules
 - the same spelling out of NP occurs whether the NP is part of a PP or not.

A **verb phrase** is another type of constituent which includes the verb and its complements, such as a direct object, indirect object, and even a sentence.

they [saw me] she [gave the book to me] you [said that you would arrive on time]

5.6 From phrases to sentences

- **Fundamental division in a sentence** between the subject (an NP) & the verb phrase (VP).
 - these are often referred to as **subject** and **predicate**
 - the thing discussed, and what's said about it

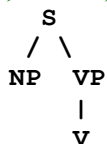
Because of this, the first rule in our phrase structure grammar will look something like this:



The **internal structure of the verb phrase** depends on the nature of the verb.

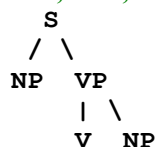
intransitive: verb does not have an object

laugh, frown, die, wait, fall VP → V



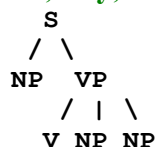
transitive: verb has an object (= an NP complement)

see, want, like, find, make VP → V NP



ditransitive: verb has two objects

give, tell, buy, sell, send VP → V NP NP



OR: VP → V NP PP

I gave the book to you

I gave you the book.

Often the same verb can **belong to more than one class**.

I already [ate]

I already [ate the apple]

She [told] as in, "I'm gonna tell!"

She [told a story]

She [told me a story]

It's also possible for a verb to take **complements that aren't NPs**, at least not in any simple sense.

We [told him the truth]

We [told him that we were leaving]

They [want the book]

They [want to leave]

They [want you to leave]

There are a number of other complications and details with phrase structure that you can learn about if you take a course in syntax, like **Ling 120** taught by Dr. Lotus Goldberg.