

LING 1010



Language and Mind

Prof. Jon Sprouse

02.15.21:

Syntax

Unlike morphemes, sentences
are not memorized

Let's start by falsifying the simplest theory

We know that we need to memorize morphemes - we've already seen that.

And now we want to build a theory of how to combine morphemes (words) into sentences.

One simple theory would be to say that we use the same mechanism: **memory**. Maybe we just **memorize sentences**.

This is **not true**. So what I want to do first is show you three pieces of evidence that suggest that sentences are not memorized, but instead are constructed from **specific rules**.

Three pieces of evidence that suggest that sentences are, not memorized, but are constructed using syntactic rules

Evidence 1: Infinity

Words/morphemes

How many words are there in English?

We've seen this is complicated because words can be changed by adding affixes, but it seems possible to do.

If you google this, you get answers like 171,000 from entries in the OED. Or 1 million and up from people who count all forms of words.

Sentences

How many sentences are there in English?

This feels unanswerable, right? It feels like a trick question.

And we can demonstrate that there an infinite number of sentences:

I have **one** car.

I have **two** cars.

I have **three** cars.

I have **four** cars.

The fact that sentences are **infinite in number**, means that they **can't be memorized**. We only have a finite amount of memory!

Evidence 1: Infinity (continued)

Words/morphemes

What is the longest word in English?

Pneumonoultramicroscopicsilicovolcanoconiosis

Antidisestablishmentarianism

Sentences

What is the longest sentence in English?

This feels feels like a
trick question again.

I like cookies.

Lisa said that I like cookies.

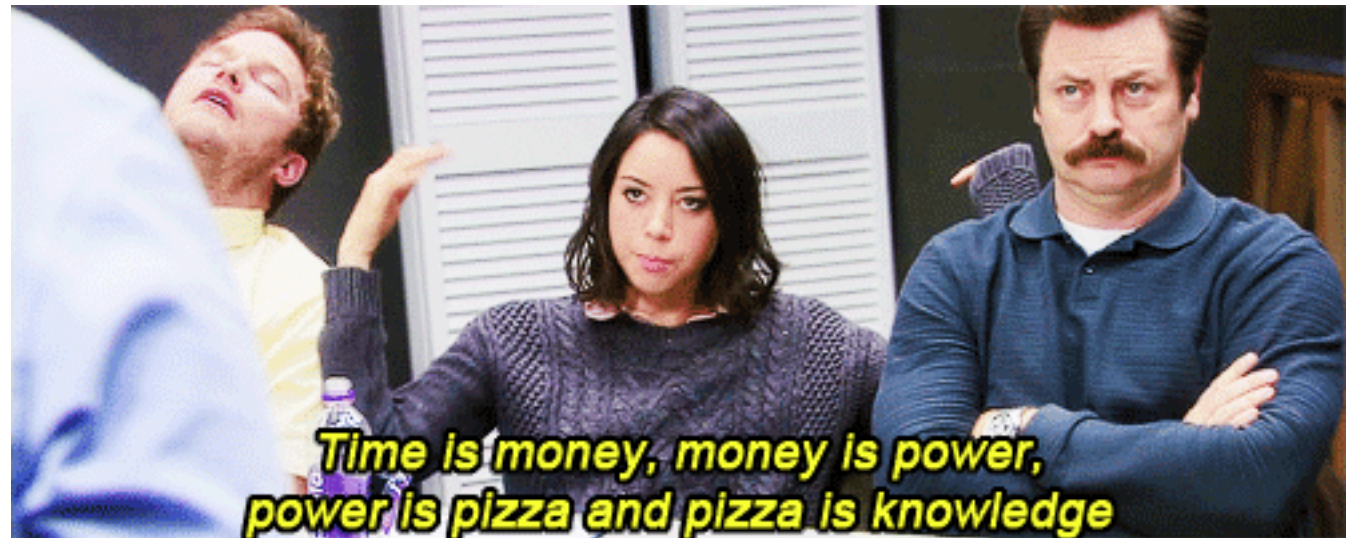
Clare thinks that Lisa said that I like cookies.

Ben claimed that Clare thinks that Lisa said that...

The fact that sentences are **infinite in length**, means that they **can't be memorized**. We only have a finite amount of memory!

Evidence 2: You understand novel sentences

Here is a sentence that you have probably never heard before (or before you heard it on Parks and Rec). In other words, it is a novel sentence:



(Almost all jokes that you find funny are novel sentences — because jokes tend to be less funny the second time you hear them.)

The fact that you can say and understand **sentences that you've never heard before** means that **sentences are not memorized**. They must be constructed by rules!

Compare this to things that we know are memorized, like **morphemes**:

Can you understand words that you've never heard before? **No, you have to ask what it means (or look it up in a dictionary).**

Evidence 3: Word order matters

You know that words in sentences need to go in a certain order. If you reverse the order of the words in a sentence, it is no longer grammatical:

- ✓ Revolutionary new ideas occur infrequently.
- * Infrequently occur ideas new revolutionary.

OK, perhaps you want to say that we just memorized the grammatical sentence, and didn't memorize the ungrammatical sentence.

But the ungrammatical sentence is systematically bad. Except for a few special cases, the reverse of a grammatical sentence is (almost always) an ungrammatical sentence.

- * flavor that like didn't I
- * extended was homework the for deadline the

It seems like quite a coincidence that every single reversed sentence is bad. Science tries to minimize explanations by coincidence. So this strongly suggests that there are patterns to grammatical sentences - and patterns suggest rules, not memorization.

Evidence 3: Word order (continued)

I just want to add that the patterns that we see in the grammaticality of sentences are very subtle, and very interesting.

The rules seem to be even more subtle than this. Sometimes just changing one or two words will lead to ungrammaticality:

✓ I **think that** John ate a cookie.

✓ **What** do you **think that** John ate?

✓ I **wonder who** ate a cookie.

* **What** do you **wonder who** ate?

The left two sentences show us that we can replace **think** with **wonder**, and we can replace **that** with **who**, and the sentence is still grammatical.

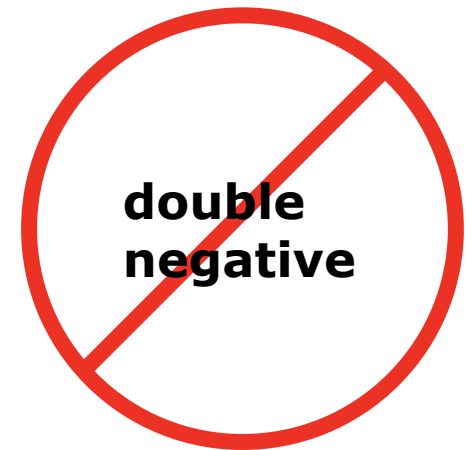
The top left shows us that we can turn the **think** sentences into a question, and it is still grammatical.

But the bottom right shows us that we can't turn the **wonder** sentences into a question. It becomes ungrammatical. That is a puzzle — it shows us that the rules for constructing these sentences are complicated!

An important distinction:
Prescriptive vs Descriptive
Rules

Time to unlearn some stuff

When you were in school, your teachers probably told you some of the “rules” to constructing sentences in English:

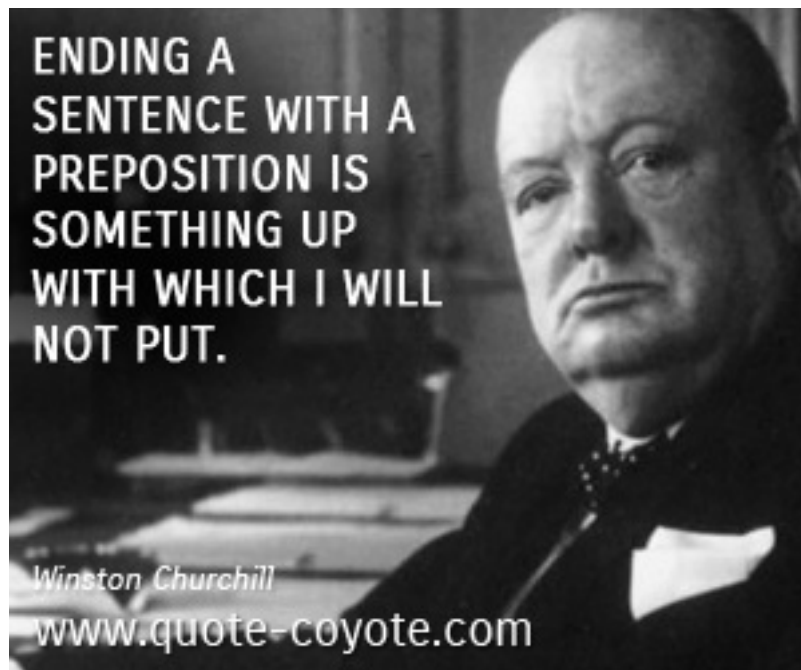


These aren't the rules for constructing sentences. Notice that they don't tell you where to put all of the words in the sentence. They just tell you about **one word** (or so), and what **NOT** to do with it. Someone who didn't know English already couldn't use them. They'd be completely lost.

These are rules for showing off your education. It is a way to show that you've been to school, and paid attention.

We don't even follow these rules

Listen to the way we actually speak. Not only do we not follow these rules, but when we do, it often sounds ridiculous.



Linguists call these types of rules **prescriptive rules**. These are rules that are **prescribed** by people who really care about style. They are **NOT** the rules for constructing sentences that are in your mind. So we are not interested in them. (In fact, they are just a way to create two groups of people, those who follow them and those who don't, so they are kind of nefarious!)

We want to study the rules in your mind!

The rules we are going to study are **the actual rules for building sentences**.

These rules are in your mind - you use them every time you speak/hear a sentence.

These rules are complex - they tell you where to put every single word in a sentence.

These rules were never taught to you explicitly - you learned them as a small child when you were learning your language.

Linguists call these rules **syntactic rules**. The word “syntax” means “word order”, so you can see why this would be a good term to use. The field that studies the structure of sentences is called the field of **syntax** (hence the name of today’s lecture).

Linguists call a set of syntactic rules a **grammar**, so they will also sometimes call these rules **grammatical rules**.

Warning: Confusing Terminology

The word “grammar” is used both by people who are interested in prescriptive rules, and by linguists. I am sorry about this. It is confusing.



=

Prescriptive Grammar / Prescriptive Rules

These are rules intended to tell people how to use their language in order to appear educated. They are about style. They **prescribe** - they tell you what to do.



=

Descriptive Grammar / Descriptive Rules

These are the rules that are actually in your mind, the ones you learned as a child when you learned your native language. They **describe** what you actually know. They are subconscious, so we need to use science to figure them out.

Let's build a theory of
syntactic rules

Component 1: Syntactic Category

We don't want a different rule for each word

Estimates for the number of words in English range from 100,000 to 1,000,000. It turns out it is really complicated to count words (remember the Inuit). But it doesn't really matter what the exact number is. It is very large!

Syntactic rules tell us where to put each word in the sentence. If we had a rule for each word, that would be a lot of rules for children to learn, and a lot of rules to store in our minds.

Fortunately, it looks like the human mind takes advantage of the idea of **categories** to reduce the number of rules. The idea is that two objects in the same category share some relevant property. For example, two movies in the "horror" category will share the property of being scary.



The same idea seems to apply to words. Two words that share the same **syntactic category** can **appear in the same position in a sentence**.

Syntactic Categories

Two words that share the same **syntactic category** can **appear in the same position in a sentence**.

We can begin to build an inventory of syntactic categories by taking an ordinary sentence, and deleting one of the words. Then we can ask, which words can replace the missing one?

The ____ existed.

dog

homework

idea

* eat

* of

* quickly

All of the words that can fit in this position are the same **syntactic category**, which in this case we call **nouns**.

The words that can't fit in this position are **not nouns**. We need more tests to see which category each of these words are.

More syntactic categories

Nouns:

The ____ existed.

Verbs:

The cat will ____.

Prepositions:

It died right ____ here.

Adjectives:

They are very ____.

Adverbs (manner):

She coughed ____.

Adverbs (sentential):

____, you are a liar.

Determiners:

He wrote ____ other work(s).

Complementizers:

I know ____ John is a liar.

We can define a set of syntactic categories by defining a set of sentence frames, and asking which words fit in which frame.

Mad Libs!

At some level, you already knew that some words could show up in the same position in a sentence as another word... because you probably played Mad Libs as a kid.

The game Mad Libs takes advantage of the fact that words of the same syntactic category can replace each other in a sentence.

The meaning will sometimes be strange (or funny), because syntactic category **is not about meaning**. Syntactic category is just about where in the sentence the word can go according to the grammar. It is not about semantics!



MAD LIBS
TALK LIKE A PIRATE

Ye can always pretend to be a bloodthirsty _____,
NOUN
threatening everyone by waving yer _____ sword
ADJECTIVE
in the air, but until ye learn to _____ like a pirate,
VERB
ye'll never be _____ accepted as an authentic
ADVERB
_____. So here's what ye do: Cleverly work into yer
NOUN
daily conversations _____ pirate phrases such as
ADJECTIVE
"Ahoy there, _____," "Avast, ye _____,"
PLURAL NOUN PLURAL NOUN
and "Shiver me _____." Remember to drop all yer gs
PLURAL NOUN
when ye say such words as *sailin'*, *spittin'*, and *fightin'*. This will
give ye a/an _____ start to being recognized as a
PART OF THE BODY
swashbucklin' _____. Once ye have the lingo down
NOUN
pat, it helps to wear a three-cornered _____ on yer
NOUN
head, stash a/an _____ in yer pants, and keep a/an
NOUN
_____ perched atop yer _____. Aye,
NOUN PART OF THE BODY
now ye be a real pirate!

Syntactic category is not the same as the “parts of speech” that you learned in school

Though we are using labels for syntactic categories that you have seen before (noun, verb, etc), they are not the same.

Think about the definitions that your teachers gave you for parts of speech:

“A noun is a person, place, thing, or idea.”

“A verb is an action or state of being.”

These are definitions based on **semantics/meaning**. They are not based on syntax. **Syntactic category** is just about syntax - the places in the sentence the word can appear.

Also, science abhors disjunctions - if you have to say “something is either x or y”, you don’t have a very good theory. Notice that we don’t have that issue with syntactic category. You either are, or are not, each category.

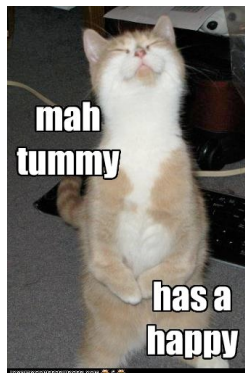
Syntactic Categories must be part of lexical entries!

Up until now, items stored in the lexicon (morphemes) have consisted of a pair of representations: a sound (phonetic representation) and a meaning (semantic representation):

Phonetic representation:

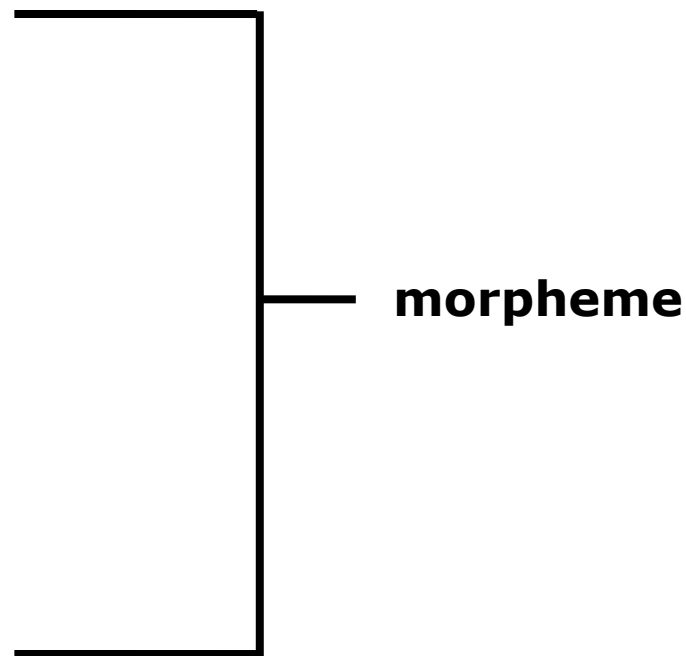
[k æ t]

Semantic (meaning)
representation:



Syntactic Category:

noun

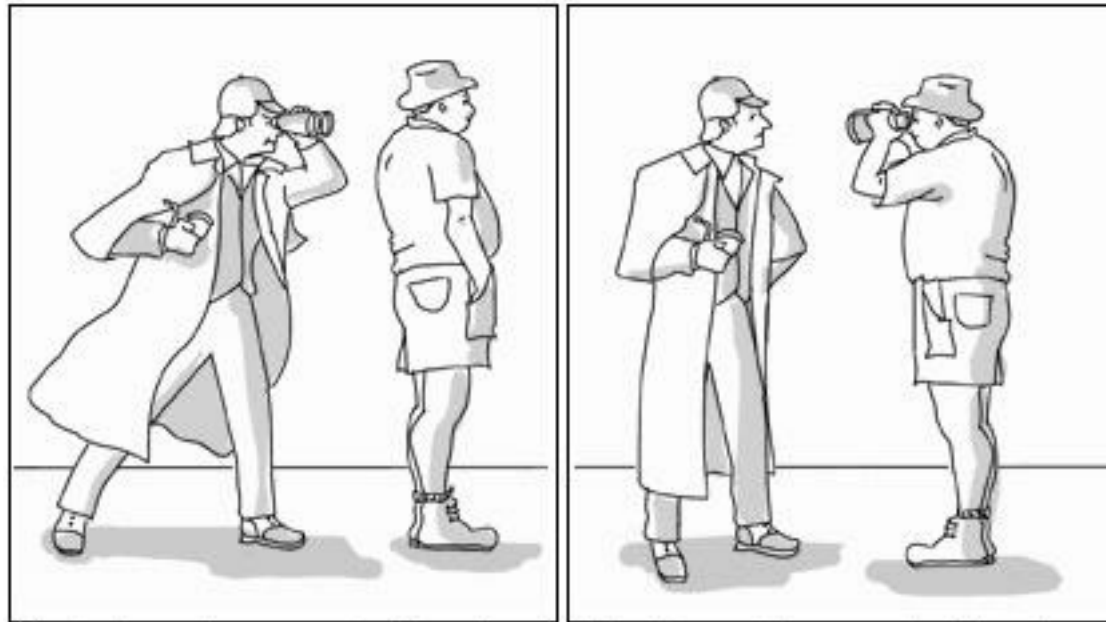


Now we need to add a third piece of information to the entry: **syntactic category**.

Component 2: Phrase Structure Rules

Sentences can be ambiguous!

Sentences can be ambiguous just like we saw with words. For example, this string of words can have two different meanings. Assuming that the meaning of sentences is **compositional** (based on the words in the sentence), how can we explain the same set of words leading to **two distinct meanings**?

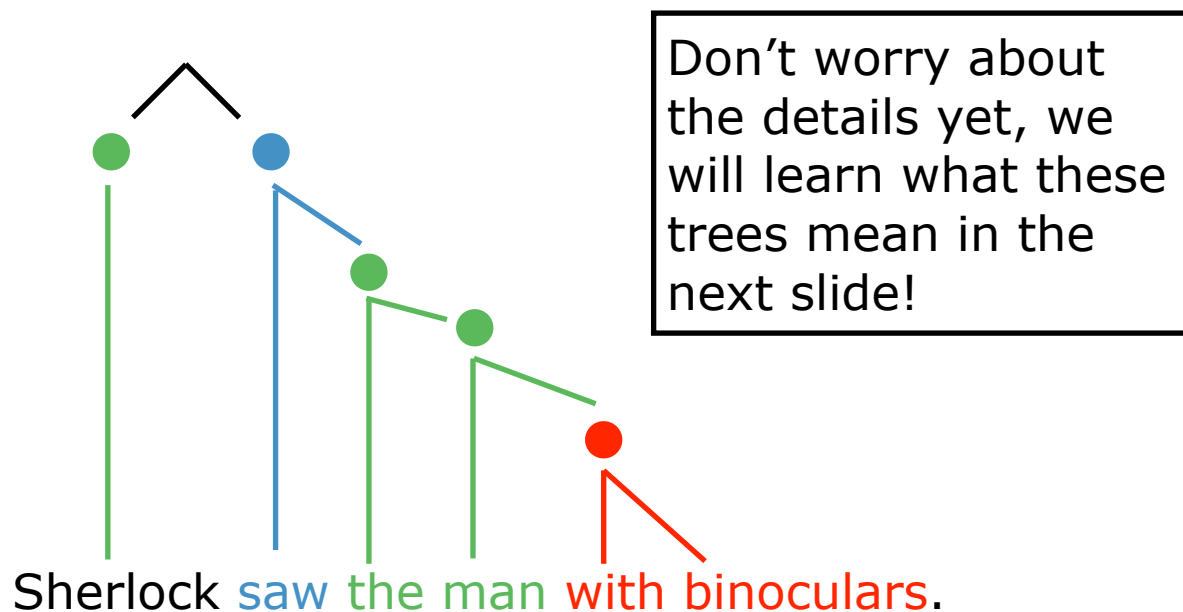
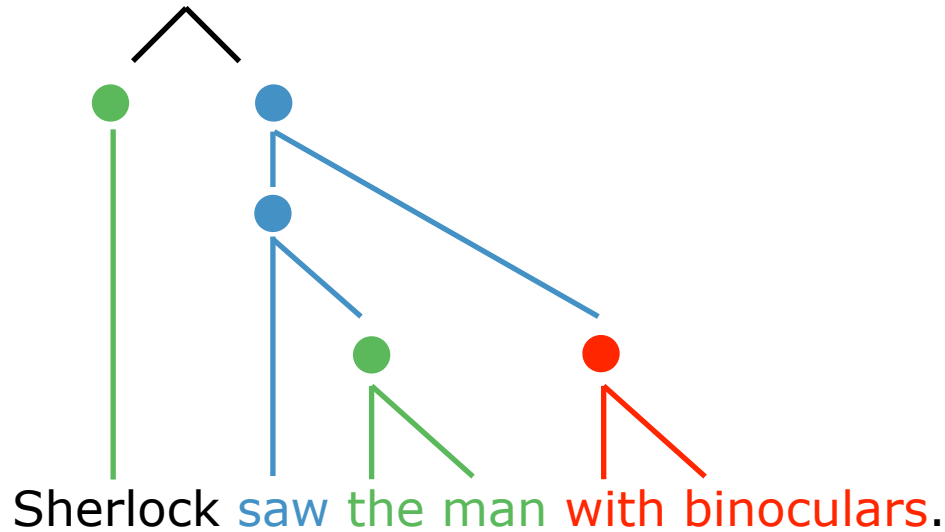


Sherlock saw the man with binoculars.

We can use the same logic that we used for morphology, and conclude that sentences must have **hierarchical structure**.

Ambiguity = Hierarchical Structure

Just like we saw with morphology, we can explain the ambiguity if we assume that **sentences have hierarchical structure**:



Don't worry about the details yet, we will learn what these trees mean in the next slide!



Structure-building rules for syntax

We've already seen structure-building rules for creating complex words. Now let's try to come up with some structure building rules for constructing sentences.

The first step is to label the syntactic categories of words. Syntactic categories are the units that will go into our rules.

D = determiner

N = noun

V = verb

P = preposition

To save space, we can use the first letter of each syntactic category instead of the full name. No big deal.

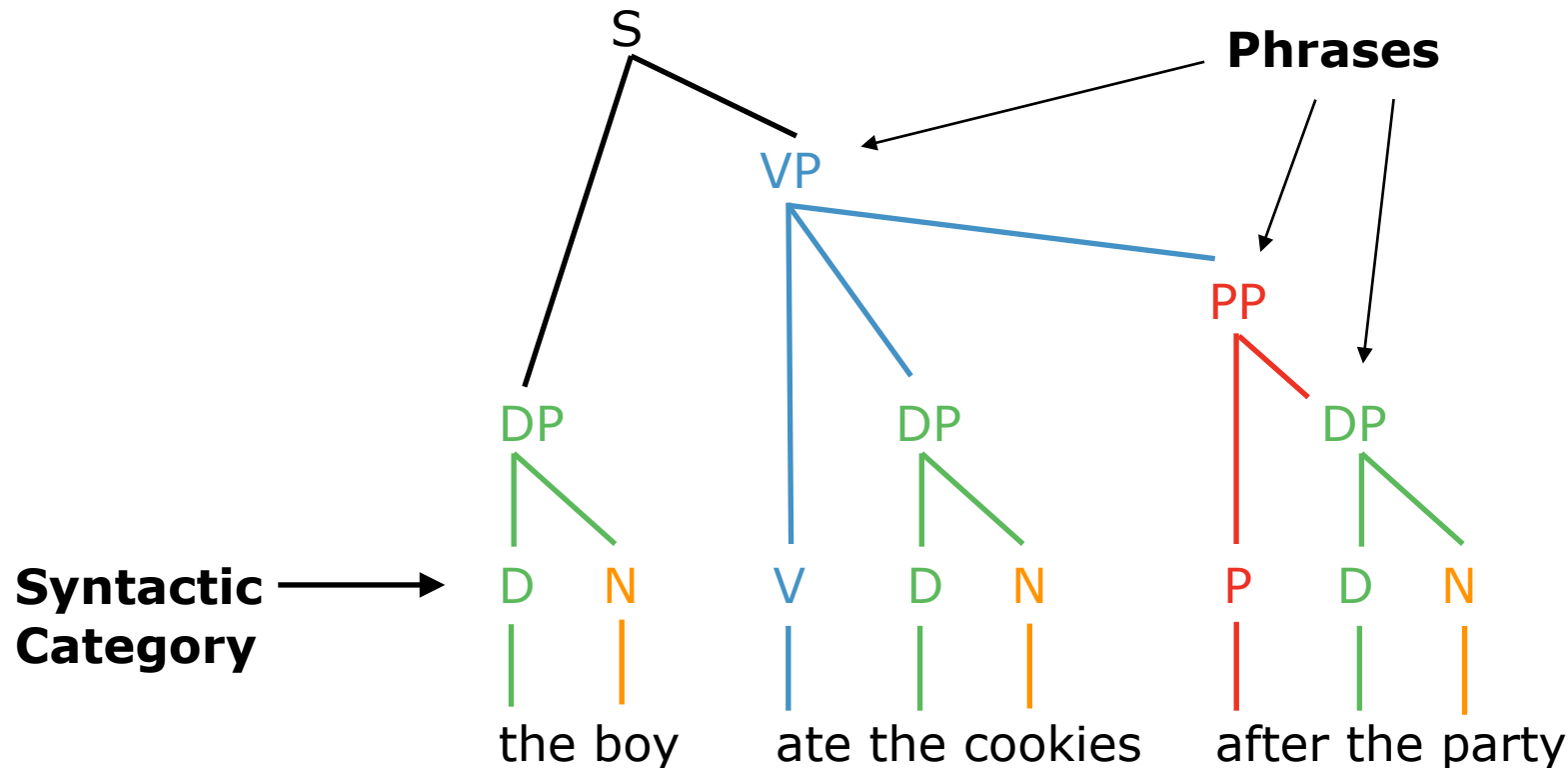


Structure-building rules for syntax

The next step is to figure out which words combine together.

When two or more words (or phrases) are combined together, we call it a **phrase**.

Linguists have various tests that they use to figure out the phrases in a sentence, called constituency tests. But I won't cover them here. You can take a full course on syntax to learn those sorts of details.



Structure-building rules for syntax

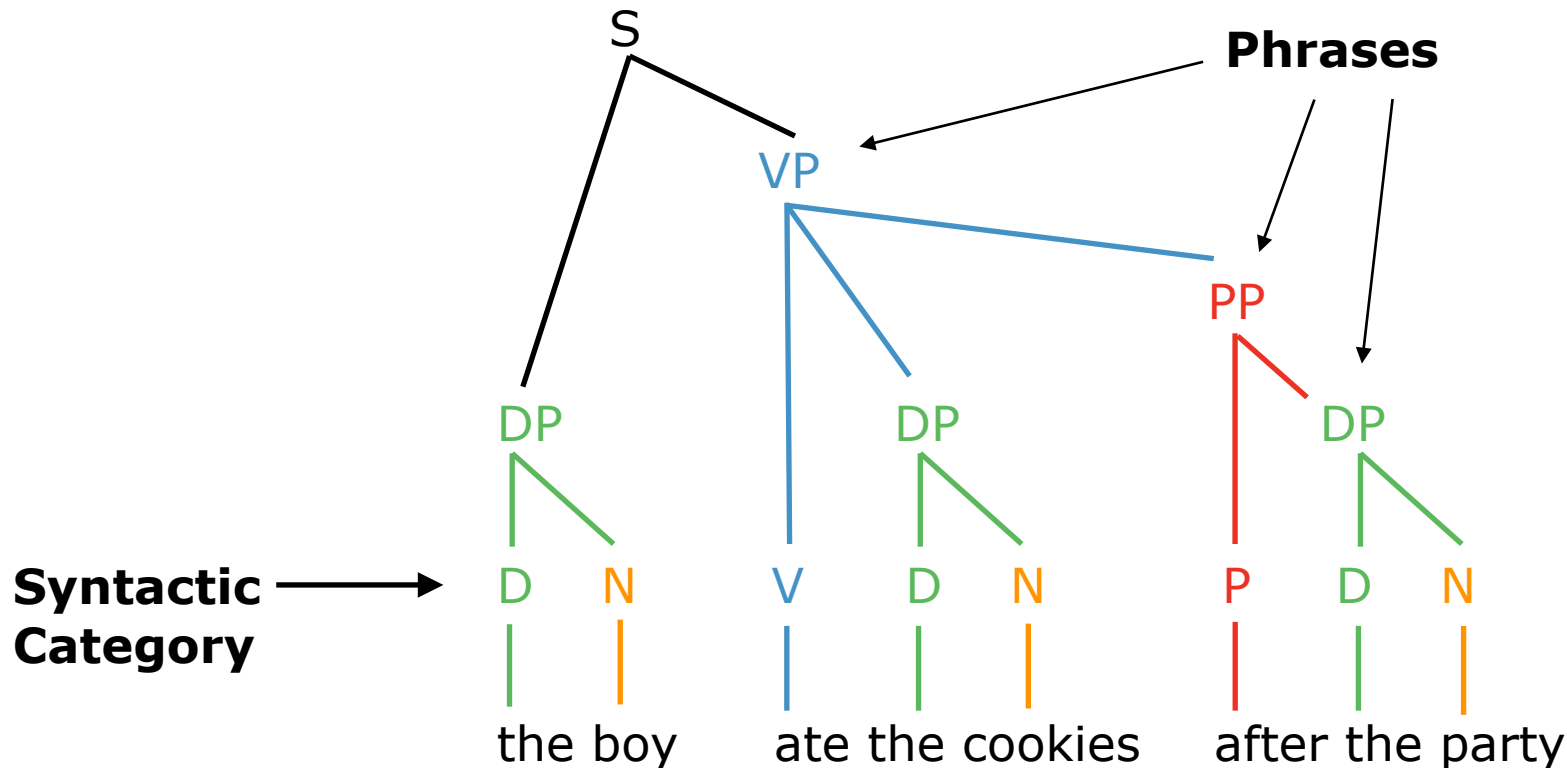
Notice that we have labeled each phrase with a two letter acronym. The first letter stands for one of the syntactic categories in the phrase. And the “p” stands for phrase.

DP = determiner phrase

NP = noun phrase

VP = verb phrase

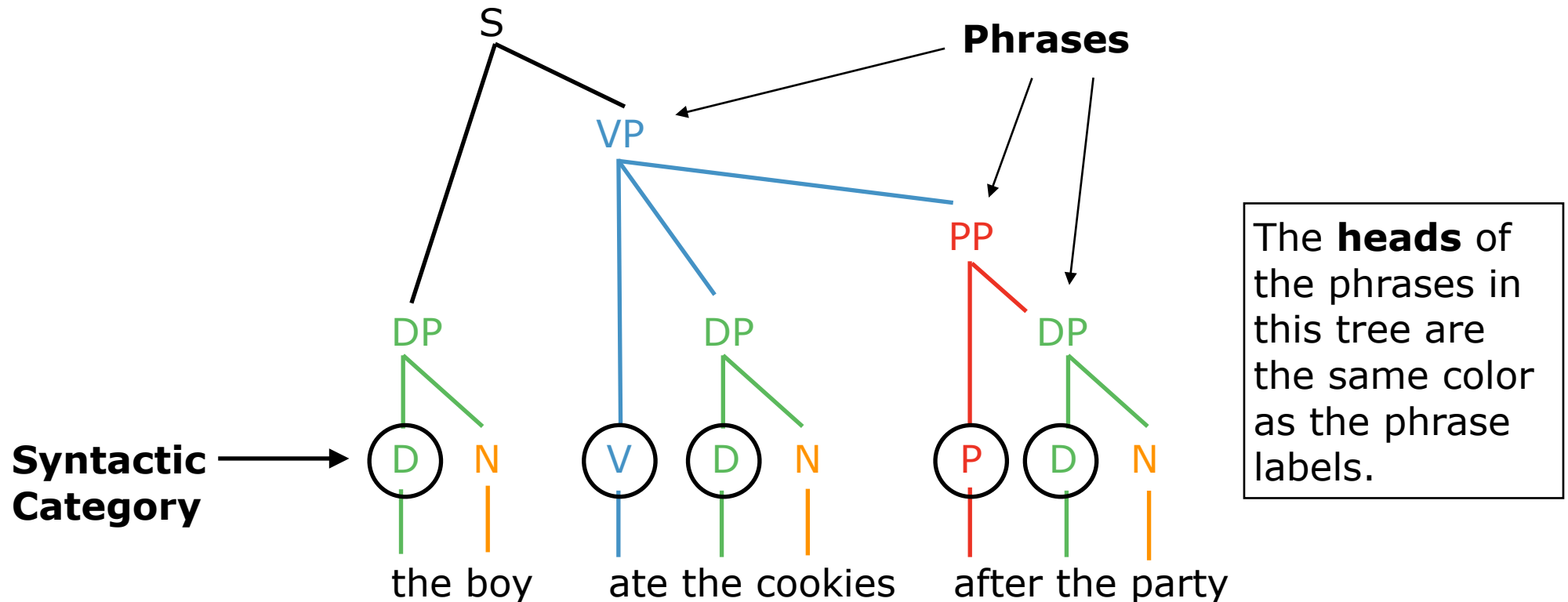
PP = prepositional phrase



Structure-building rules for syntax

We call the syntactic category that lends its name to the phrase the **head** of the phrase. Every phrase has a head.

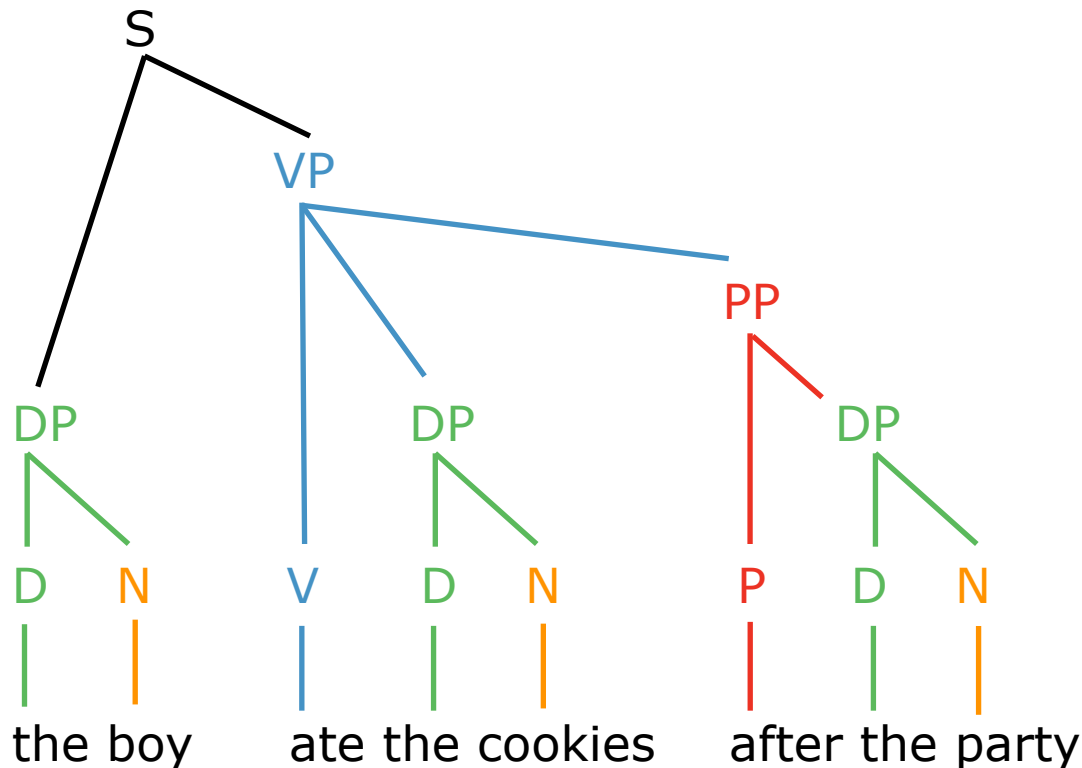
The head of the phrase is important, because it determines the properties of the phrase. Just as syntactic categories can only appear in certain location in the sentence, so too can specific phrases only appear in certain locations.



Structure-building rules for syntax are called **Phrase Structure Rules**

Because the structure-building rules in syntax are used to construct phrases from two or more words or phrases, we call them **phrase structure rules**.

We can read the phrase structure rules that built this tree right from the tree itself. Basically, there is one phrase structure rule every time two lines connect:



$S \rightarrow \text{DP VP}$

$VP \rightarrow \text{V DP PP}$

$PP \rightarrow \text{P DP}$

$DP \rightarrow \text{D N}$

Structure-building rules for syntax are called **Phrase Structure Rules**

Because the structure-building rules in syntax are used to construct phrases from two or more words or phrases, we call them **phrase structure rules**.

We can read the phrase structure rules that built this tree right from the tree itself. Basically, there is one phrase structure rule every time two lines connect:

You do not need to memorize these phrase structure rules. I haven't taught you how to figure out the phrase structure rules of a language, I have just given you examples. You can learn how to find them for a language in a course like LING 2010Q. For this course, I just want you to know that phrase structure rules exist, and what they do for languages! The fact that human languages use phrase structure rules for sentences is structure in the mind!



Putting it all together

Linguists want to study the **grammatical rules in your mind** (descriptive rules), not the prescriptive rules that are taught in school.

There are three pieces of evidence that sentences are constructed from syntactic rules, and not, for example, memorized:

1. Sentences are infinite in number and length (infinity).
2. You can understand sentences you've never heard before.
3. Word order matters.

And we have seen two components to the theory of syntax:

1. Syntactic category (which reduces the number of rules we need)
2. Phrase-structure rules (which explain ambiguity through hierarchical structure)

In full disclosure, there is another type of syntactic rule called a "transformation" which I haven't told you about yet. We will talk about transformations a bit when we discuss the acquisition of syntax.