02.13.19: Syntax

Language and Mind
Prof. Jon Sprouse
There are rules to building sentences
Time to unlearn some stuff

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

- split infinitive
- end sentence in preposition
- double negative

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.

**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 are subconscious, so we need to use science to figure them out.
Three pieces of evidence that suggest that sentences are governed by syntactic rules
Evidence 1: 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.

One way to explain this fact is to say that there are rules that dictate exactly where each word must go in the sentence. When you break these rules (reversing the order breaks the rules), the sentence becomes ungrammatical.

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.
- * I wonder who ate a cookie.
- What do you think that John ate?
- * What do you wonder who ate?

The importance of precise word orders suggests that existence of precise rules.
Evidence 2: You understand novel sentences

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

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!

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.

For example, what does subjacency mean? You don’t know, because it is a word that linguists made up. You can only know it if we tell you what it means.
Evidence 3: Infinity

**Question**: How many sentences are there in English?

I have one car.
I have two cars.
I have three cars.
I have four cars.

Notice that you can ask this question for morphemes (which are memorized)!

**Question**: What is the longest sentence in English?

I like cookies.
Lisa said that I like cookies.
Clare thinks that Lisa said that I like cookies.
Ben claimed that Clare shanks that Lisa said that I like cookies...

Notice that you can ask this question for morphemes (which are memorized)!

The fact that sentences are infinite in number, and infinite in length, means that they can’t be memorized. We only have a finite amount of memory!
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 frames.
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. It is only about the syntax of the word — the location of the word in a sentence. Therefore, in mad libs, the sentences are syntactically well-formed, but are often semantically odd.
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 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**?

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!
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.
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

The sentence "the boy ate the cookies after the party" is parsed as follows:

```
S
 /   \
V   PP
 /     \
DP    DP
 /  \
D N   D N
the boy ate the cookies after the party
```

Syntactic Category:
- 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 DP VP
VP \rightarrow V DP PP
PP \rightarrow P DP
DP \rightarrow D N
```

```
S
  \rightarrow
    DP VP
      \rightarrow
        V DP PP
          \rightarrow
            P DP
              \rightarrow
                D N
```

- S
- DP
- VP
- PP
- D
- N
- V
- P
- D
- N

**the boy**

**ate the cookies**

**after the party**
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:

- \( S \rightarrow DP \rightarrow VP \)
- \( DP \rightarrow D N \)
- \( PP \rightarrow P DP \)
- \( VP \rightarrow V DP PP \)

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

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.