

Ling 315 Lecture notes. 10 January 2007.

Syntax is that branch of linguistics devoted to the study of sentence structure. Though we use sentences all the time, we don't normally think about how they are structured. However, a little consideration will reveal that the principles by which words are organized into sentences are, in fact, quite complex.

Our aim in this class will be to build a model of *syntactic competence*, that is, to build a system of rules and principles that generates all and only the sentences of English that are judged grammatical by native speakers. We'll call this system a *grammar*.

1. Substitution classes

We'll start by looking at what are called *substitution* (or *distribution*) classes. First we'll examine what underlies the notion of syntactic (or lexical) category (e.g., *noun* or *verb*), and from there we'll move on to phrasal categories.

1.1 Lexical categories. Lexical category in English can be defined in terms of the *distribution* of a word within a sentence. Specifically, *noun* (N), *verb* (V), and so on are *substitution classes* of words. They are substitution classes in the sense that there is a set of positions within a sentence into which any member of that class can be substituted, preserving the grammaticality of the sentence. To illustrate this point, let's assume a simple lexicon with the following words:

squirrels, kittens, love, like

From here, we can state some rules for what may comprise a sentence:

$S \rightarrow \textit{Squirrels love kittens}$	$S \rightarrow \textit{Kittens love squirrels}$
$S \rightarrow \textit{Squirrels love squirrels}$	$S \rightarrow \textit{Kittens love kittens}$
$S \rightarrow \textit{Squirrels like kittens}$	$S \rightarrow \textit{Kittens like squirrels}$
$S \rightarrow \textit{Squirrels like squirrels}$	$S \rightarrow \textit{Kittens like kittens}$

On the one hand, what's good about this grammar is that it succeeds in generating all the grammatical sentences out of the four words -- and *not* generating ungrammatical sentences such as **Love squirrels kittens*.

However, we can immediately observe that what's wrong with these rules is that they miss an important generalization: wherever *squirrels* may occur in a sentence, *kittens* also may occur. Both may occur before and after *love* or *like*. In other words, these words have the same *distribution* within a sentence; *squirrels* and *kittens* can always substitute for one another. *Squirrels* and *kittens* thus comprise a substitution class, the class of nouns (Ns). We can then revise our rules to reflect this generalization:

$N \rightarrow \textit{kittens, squirrels}$	$S \rightarrow N \textit{ love } N$	$S \rightarrow N \textit{ like } N$
---	-------------------------------------	-------------------------------------

The situation is similar for *love* and *like*: these two words have the same distribution within a sentence, and thus also form a substitution class, which we commonly refer to as verb (V). We may then again revise our S rule:

$N \rightarrow \textit{kittens, squirrels}$	$V \rightarrow \textit{love, like}$	$S \rightarrow N V N$
---	-------------------------------------	-----------------------

Consider now the distribution of the words *happy* and *nice*:

- (1) Nice/happy kittens love squirrels.
- (2) Kittens love nice/happy squirrels.
- (3) Nice/happy kittens love nice/happy squirrels.

Since these two words also have the same distribution -- they occur optionally before an N -- we have reason to propose a third substitution class, which we commonly refer to as *adjective* (A). Again revising our rules, we get

$N \rightarrow \textit{kittens, squirrels}$	$V \rightarrow \textit{love, like}$	$A \rightarrow \textit{happy, nice}$
$S \rightarrow (A) N V (A) N$		

Words like *the* and *those* comprise a fourth substitution class:

- (4) The/those kittens love squirrels.
- (5) The/those nice kittens love squirrels.
- (6) Squirrels love those/the kittens. etc.

These words may optionally occur immediately preceding an N or A N, and are commonly referred to in linguistics as *determiners* (Ds). Revising again:

$N \rightarrow \textit{kittens, squirrels}$	$V \rightarrow \textit{love, like}$	$A \rightarrow \textit{happy, nice}$
$D \rightarrow \textit{the, those}$	$S \rightarrow (D) (A) N V (D) (A) N$	

To summarize so far, then, certain sets of words can be substituted for one another within a sentence, and this motivates the classification of words into syntactic categories such as N, V, A, and D.

=====

How to identify a word's syntactic category. The major syntactic categories of English are as follows:

- N(noun): *kittens, squirrels, people, guys, bed...*
- V(erb): *love, like, hug, slept, is...*
- A(djective): *nice, happy, awake...*
- P(reposition): *on, above, near...*
- D(eterminer): *the, those, a...*
- Adv(erb): *quickly, thoughtfully, carefully...*

Nouns

- Can be modified by adjectives
- Can inflect for number (singular/plural)
- Cannot have noun phrase sisters
- Follow determiners

Verbs

- Can inflect for tense (e.g., *-ed*)
- Can immediately follow an auxiliary (e.g., *should, will, has*)
- Can inflect for aspect (e.g., *-ing*)

Adjectives

- Can be modified by *very*
- Can be inflected for the comparative and superlative
- Can appear immediately following *seem, appear, and become*
- Can modify nouns

Adverbs

- Appear at the very beginning or end of a sentence
- Often end in *-ly*

Prepositions

- Can be modified by *right*

Determiners

- Can immediately precede the phrases *other person, other people*

Exercise 1: Given another example of a word for each category.

===

1.2 Phrasal categories. Not only words form substitution classes: *sequences* of words do as well. For example, *squirrels* (N), *nice squirrels* (A N), and *the nice squirrels* (D A N) may all occur in the same position within a sentence, namely, before or after a verb. Thus, we can say that the sequences N, A N, and D A N form a substitution class of their own, which we may refer to as NP. The grammar can then be revised as follows:

N → *kittens, squirrels* V → *love, like* A → *happy, nice*
D → *the, those* S → NP V NP NP → (D) (A) N

Similarly, if we add in intransitive verbs such as *slept* and *yawned*, we see that V and V NP also form a substitution class: both sequences immediately follow the first NP in a sentence. This substitution class is called VP. Revising our rules once again:

N → *kittens, squirrels* V → *love, like, slept, yawned*
A → *happy, nice* D → *the, those*
S → NP V NP NP → (D) (A) N VP → V (NP)

The words *on* and *near* form yet another substitution class. Based on the following examples, where do these words occur within a sentence?

- (7) The kittens slept on/near the squirrels.
- (8) The kittens put the squirrels on/near the bed.
- (9) The kittens are on/near the bed.

These examples show that Ps may immediately follow a V or V NP, and usually immediately precede an NP. We can also see from these examples that P and P NP form another substitution class, which we'll call PP. Revising our rules once again:

Lexicon

N → *kittens, squirrels, bed* V → *love, hugged, slept, yawned*
A → *happy, nice* D → *the, those* P → *on, near*

Phrase Structure Rules

S → NP VP NP → (D) (A) N VP → V (NP) PP → P (NP)

These rules *generate* sentence of English. They group words into phrases, which are themselves grouped into larger phrases, with the largest group forming a sentence.

2. Phrase Structure Trees. The internal structure of a sentence, as generated by the PS Rules, can be represented by a tree diagram. [See class lecture for details.] Tree terminology: *branch, node, root node, terminal node, dominate, immediately dominate, mother, daughter, sister.*

Exercise 2: Draw the phrase structure trees for the following sentences.

- (10) Those kittens slept. (11) The kittens put the squirrels on the bed.

Exercise 3: Revise the phrase structure rules so that they generate the following S:

- (12) The kittens with fleas slept.

Exercise 4: Account for the ambiguity of the following example by drawing two distinct trees for it. Specify which interpretation goes with which tree.

- (13) Sara wrote a letter to the editor.

Exercise 5: Give a sentence that our grammar generates, but is ungrammatical, and give a sentence which is grammatical, but which our grammar does not generate.

3. Constituency tests. Sentences are more than just ordered sequences of words. The individual words in a sentence are organized into natural groupings, which are themselves organized into larger groupings, the largest grouping of all being the sentence. These groupings within a sentence are called *constituents*, or *phrases*.

So far we have justified the groupings of words into constituents based on the fact that they form substitution classes. Further evidence for these groupings comes from a set of tests called *constituent structure tests*. If the group of words you are testing can pass *at least two* of these tests, you have a good argument that they form a constituent.

2.1 *Replacement test.* Can the group of words be replaced by a pro-form?

- (14) a. The kittens put the squirrels on the bed.
 b. They put the squirrels on the bed.
 c. The kittens put them on the bed.
 d. The kittens put the squirrels there.
 e. The kittens did it.
 f. *The kittens put blarg.
 g. *Blarg the squirrels on the bed.

2.2 *Movement test.* Can the group of words be moved to another position in the S?

- (15) a. The squirrels, the kittens put on the bed.
 b. On the bed, the kittens put the squirrels.
 c. ??Put the squirrels on the bed, the kittens (did).
 d. *The squirrels on the bed, the kittens put.
 e. *Put the, the kittens squirrels on the bed.
 f. *The squirrels on, the kittens put the bed.

2.3 *Cleft test.* Can the group of words be clefted?

- (16) a. It was the kittens that put the squirrels on the bed.
 b. It was the squirrels that the kittens put on the bed.
 c. It was on the bed that the kittens put the squirrels.
 d. ??It was put the squirrels on the bed that the kittens did.
 e. *It was the squirrels put that the kittens on the bed.
 f. *It was squirrels on the that the kittens put the bed.

2.4 *Fragment test.* Can the group of words be used as an answer to a question?

- (17) a. Who put the squirrels on the bed? The kittens.
 b. What did the kittens put on the bed? The squirrels.
 c. Where did the squirrels put the kittens? On the bed.
 d. *What did the kittens put? The squirrels on the bed.
 e. What did the kittens do? Put the squirrels on the bed.

2.5 *Coordination test.* Can the group of words be coordinated with a sequence of words of the same kind?

- (18) a. The kittens and that guy put the squirrels on the bed.
 b. The kittens put the squirrels and that guy on the bed.
 c. The kittens put the squirrels on the bed and on the floor.
 d. The kittens put the squirrels on the bed and fell asleep.
 e. *The kittens put the and found some squirrels on the bed.
 f. *The kittens put the and guys found some squirrels on the bed.

Background: Only identical constituents can be coordinated:

- (19) a. I wrote to you and to him. PP Conj PP
 b. I wrote a letter and a postcard. NP Conj NP
 c. *I wrote to you and a postcard. PP Conj NP