CAS LX 321 / GRS LX 621 *Syntax: Introduction to Sentential Structure*

September 11, 2018

1 Doing science

There are several levels of thing we're trying to do here, but the primary one for now is that we want to characterize what English speakers know about English syntax.

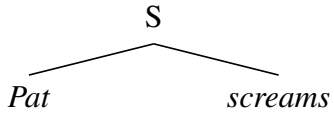
An English speaker, confronted with a string of words, can determine (through some procedure/knowledge) whether that string of words is an English sentence or not. English speakers agree on the divisions.

We've got a formalism that we're going to work with, it is a **generative grammar**—so called because it generates sentences. It is an algorithm that expands an abstract S (sentence) into a string of words. This is tied to reality by the hypothesis that **English speakers will accept as English all and only the sentences generated by the grammar**. And so now we want to work out the grammar.

Starting small.

Everybody agrees that *Pat screams* is a sentence of English.

(1) a. S → *Pat screams*



This grammar has several problems that keep us from accepting it as the thing English speakers know that makes them English speakers.

The first is that there are essentially infinitely more sentences that English speakers would accept as English other than *Pat screams*.

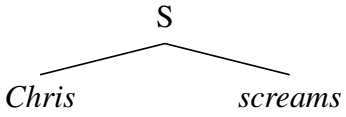
This grammar is a theory of English. It is a fairly poor theory of English. It predicts that the only sentence of English is *Pat screams*—{ *Pat screams* } is an exhaustive set of the sentences English speakers accept as English.

To demonstrate that this is a bad theory of English, we look at what it predicts (no sentence other than *Pat screams* is English), and check to see if the world bears this out. It does not. One reason, among others, is that *Chris screams* is also English, but the grammar does not generate it.

- Propose a theory
- Identify the predictions
- Test the predictions against the observed data
- If they fail, revise and return to the first step

So, given that *Chris screams* is also English, we can revise our theory of English like so:

(2) a. S → *Pat screams*
b. S → *Chris screams*



We have now patched the theory to account for the failure we identified in what it predicts. It now correctly predicts both that *Pat screams* and that *Chris screams* are English.

However, it incorrectly predicts that no other sentences are English.

But, turns out, *Tracy screams* is also English.

There are patterns here that are already obvious. We can always come up with a new name, and whatever name we put in there as the first word is going to wind up resulting in an English sentence. (For example, suppose that there's a science fiction story in which a group of people wake up in a spaceship having lost their memory and they adopt names based on the order they wake up. One woke up first, followed by Two. *One screams* and *Two screams* are both English. Extend this as far as you like. Whatever grammar you provide for a scenario where anyone up to *N* screams, I can invent a new scenario where there's one additional person, who wakes up last and has a correspondingly higher number, *M*, and *[M] screams* will be English too.)

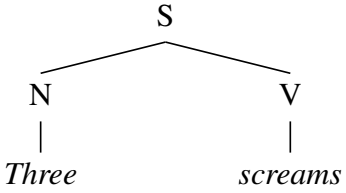
Similarly, *Pat sleeps* is also English, and any sentence that ends with *screams* has a correspondng English sentence that ends with *sleeps*.

But *Pat Tracy* is not a sentence. Nor is *Screams sleeps*. So we can't just put anything anywhere, the name has to be in the first spot and the verb has to be in the second spot.

We learned something—this has told us something non-trivial. There are classes of things that go in particular places in an English sentence. In this type of sentence at least, anything that is a "name" can go in the first spot, but not the second, and anything that is in the class of words that so far includes *screams* and *sleeps* (let's call them "verbs") can go in the second spot.

To codify this in the grammar, we do this:

(3) a. S → N V
b. N → *Chris*
c. N → *Pat*
d. N → *Tracy*
e. N → *One*
f. N → *Two*
g. N → *Three*
...
h. N → *Two-Hundred-Forty-Eight*
i. V → *screams*
j. V → *sleeps*



The two types of rules (the lexical and structural rules) really seem to have different conceptual status. The lexical rules are mostly just examples, we can easily extend them (particularly the ones corresponding

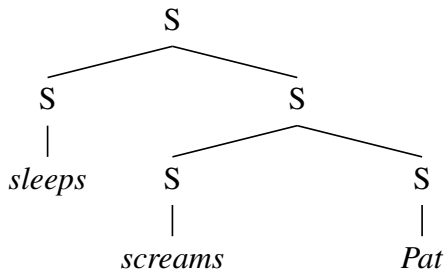
to the open-class categories like nouns and verb) without really changing our understanding of the grammar. Extending the grammar to handle *One hated Three* is different in kind from just coming up with a new name that can go in an N spot. The real *grammar* is those structural rules, the *lexicon* is distinct (though we can use the same formalism to describe it).

The procedure we'll be essentially follwing is:

- Observe a new sentence pattern that our existing grammar doesn't predict
- Modify the grammar so it does predict the new pattern
- See if this allows for any generalizations that can simplify the grammar
- See what predictions the new grammar makes
- Check to see if those predictions are borne out, repeat

Given that, let me provide an alternative and terrible grammar.

(4) a. S → S S
b. S → *Chris*
c. S → *Pat*
...
d. S → *Two-Hundred-Forty-Eight*
e. S → *screams*
f. S → *sleeps*



Suppose that we've plugged in all of the English words into rules of the S → *word* lexical rules.

Now: try to find a sentence that this grammar does not generate. You can't.

Why is this grammar terrible? Why is this not the best grammar? It predicts that every English sentence there is is a grammatical sentence of English.

The tree above for *Sleeps screams Pat* shows one problem with it. It does successfully generate all the sentences of English. But it also generates all kinds of sentences that are not English. It is not a good model of what English speakers know about English syntax. It can't differentiate between English sentences and non-English sentences (though, perhaps it can differentiate between strings of words that are English words vs. those that are not English words).

So, success for a grammar is measured both by the degree to which it generates sentences that English speakers say are English, but also by the degree to which it fails to generate sentences that English speakers say are not English.

2 Subcategories

We have categorized words as verbs and nouns (and we've already explored some of the other parts of speech, tests for them, etc.).

(5)
a. Bart ran.
b. Homer sleeps.
c. Maggie crawls.
d. Homer chased Bart.
e. Bart saw Maggie.
f. Maggie petted SLH.
g. Homer handed Lisa Maggie.
h. Marge sent Bart SLH.

(6)
a. * Ran Maggie.
b. * Crawls Homer.
c. * Chased Bart Homer.
d. * Sent Marge Bart SLH.
e. * Marge Bart SLH sent.

(7)

Grammar	
S →	N V
S →	N V N
S →	N V N N

V →	*ran*
V →	*sleeps*
V →	*crawls*
V →	*chased*
V →	*saw*
V →	*petted*
V →	*sent*
V →	*handed*

N →	*Homer*
N →	*Marge*
N →	*Lisa*
N →	*Bart*
N →	*Maggie*
N →	*SLH*

(8)
a. * Bart crawls Maggie.
b. * Maggie sleeps Bart.
c. * Homer ran Bart.
d. * Maggie handed.

What to do?

3 Better and worse

(9)
a. Homer sleeps and Maggie crawls.
b. Homer sleeps or Maggie crawls.
c. Bart ran and Homer chased Bart.
d. Bart ran or Homer chased Bart.
e. Maggie petted SLH and Bart saw Maggie.
f. Homer handed Lisa Maggie and Marge handed Bart SLH.

(10)

Add... A	
S →	N Vi *and* N Vi
S →	N Vi *or* N Vi
S →	N Vi *and* N V N
S →	N Vi *or* N V N
S →	N V N *and* N V N N
S →	N V N N *and* N V N N

Add... B	
S →	S Conj S
Conj →	*and*
Conj →	*or*

Why is A better?