September 4, 2018

1 Stuff we know

(1) a. Sue said that Mary borrowed my guitar.
    b. Sue said Mary borrowed my guitar.

(2) a. What did Sue say that Mary borrowed?
    b. What did Sue say Mary borrowed?

    a. Who did Sue say borrowed my guitar?
    b. * Who did Sue say that borrowed my guitar?

This is a problem for the idea that you (in this instance, an infant) just *generalize* from what you hear to the grammar of the language. It's more complicated than we even knew, we weren't taught this stuff. Something about either the *acquisition procedure* or the *cognitive system* or the *way it's stored in the brain* or something like that is constraining what kinds of languages there are.

This line of argument (language knowledge is more complex than is reflected in what language acquirers get in their environment) is called the *poverty of the stimulus* argument for the concept of Universal Grammar—which is itself the idea that there is something "human" about language. That by virtue of being human (and not a cat, ape, daffodil) we have this ability and mechanism to learn language. What determines that an organism will be human is that its parents are human; same deal with birds. So there's *something* genetic involved, if language is a uniquely human trait (even if it leverages other biological/cognitive systems that are not uniquely human).

2 Esperanto

Martin Haspelmath (1997). *Indefinite Pronouns*, pp. 24–25.

> Then the first important generalization that we can make is that indefinite pronouns are as a rule DERIVED forms. That this is not a trivial observation can be seen from the artificial language Esperanto, designed by Ludwik Zamenhof in 1887. As a rule, Esperanto grammar follows the typological design of Standard Average European, eliminating irregularities and 'useless' features like gender and agreement. There are two indefinite series in Esperanto, which are clearly related to interrogative pronouns, much as in many natural languages:

(27)

Esperanto	interrogative	indefinite	negative
person	kiu	iu	neniu
thing	kio	io	nenio
property	kia	ia	nenia
place	kia	ie	nenie
time	kiam	iam	neniam
manner	kiel	iel	neniel

> … However, natural languages that are structured like Esperanto in this respect are virtually unattested. I am not aware of a clear case in which an indefinite pronoun is formally unmarked with respect to a marked interrogative pronoun. With respect to its indefinite pronoun system, Esperanto is thus probably not a possible human language. …

3 Prescriptive rules

Prepositions are things you should know to never end a sentence with.

4 Parts of speech

4.1 Content words vs. function words

- **Content words:**
  - words that denote concepts, supplying the core notional content of the phrases they appear in.
  - Usually "open class"—new ones are frequently invented or borrowed.
  - Nouns, Verbs, Adjectives, Adverbs, Prepositions
- **Function words:**
  - words that have no/little conceptual content—instead, signal relations between elements in the sentence, conveying *grammatical* meanings.
  - usually closed class—it's very rare for new ones to be coined or borrowed
  - Determiners *the*, *that*, *this*, *a*
  - Modals *will*, *might*, *could*, *should*, …
  - Complementizers *whether*, *if*, *that*, *for*, …
  - Auxiliaries *have*, *be*
  - …

(Syntactic) category is a *distributional* classification. It is about where words fit in context, what words (or maybe phrases) are like what other words (or maybe phrases) in terms of where they can occur. (Nouns aren't really a "person, place, or thing"—that's more a post-hoc approximation.)

4.2 Diagnosing content word categories

4.2.1 Noun [N]

- Can be the subject of the sentence (if a mass or plural noun)
- Can be the only thing following determiners like *the*, *some*, *any*, *this*, *a*, includuing possessives like *my*, *her*.
- Can usually be pluralized with *–s* (or some irregular plural morpheme)
- Can be modified by adjectives

4.2.2 Verb [V]

- Can be marked for past tense
- Can be suffixed with *–ing* (as long as the verb doesn't denote a state)
- Can be modified by adverbs like *again*, *often*
- Can follow modals like *can*, *will*, *may*, ... (or their negative versions), and the infinitive *to*
- Some can take the prefix *–un* (meaning 'reverse a process')

4.2.3 Adjective [Adj]

- Can directly follow *seem*
- Can appear between a determiner and a noun (be careful of compounds)
- Often end in *–y*, *–ish*, or *–ous*.
- Can be modified by *very* and *extremely* (but be alert, so can adverbs)
- Some can take the prefix *un–* (meaning 'not')

4.2.4 Adverbs [Adv]

- Can modify a verb and/or an adjective
- Often end in *–ly*
- Can be modified by *very* or *extremely* (but be alert, so can adjectives)

4.2.5 Preposition [P]

- The only non-verb category that can systematically take an object directly
- Those that denote locations or directions can be modified by *right* or *straight* (though some dialects allow *right*+Adj too)

4.2.6 Modals, auxiliaries, determiners

These are the closed-class things, not as much necessary to test them as to know them. But it is still distributional, you know what these are by virtue of what they can replace in sentnecs.

- Determiners (including articles): *the*, *a*, demonstratives(?) *this*, possessives(?) *my*, quantifiers(?) *every*
- Auxiliaries: *have*, *be*, *do*, modals(?) *can*, *might*, *shall*
- Complementizers: *that*, *for*, *whether*, *if*
- Propositions? *Aboard*, *alongside*, hmm...
- Negation: *not*

4.2.7 Go

(3) a. John was killed.
b. John was skilled.
c. medic
d. medical
e. remove
f. removal
g. The bike is behind the shed.
h. magically
i. kingly
j. The exercise is boring the students.
k. The exercise is boring for the students.
l. We are galloping around the room.

5 Phrase structure rules and trees

A phrase structure rule (a.k.a. "rewrite rule") looks like:

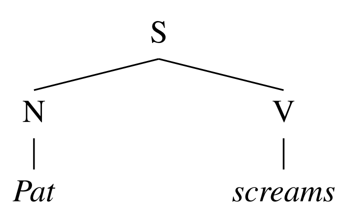
(4) $X \rightarrow Y_1\ Y_2\ Y_3 \ldots\ Y_n$

On the left is the "mother" and on the right are the "daughters."
There are two kinds of phrase structure rules we'll be concerned with.

- **Lexical rules**. Define category of a word. N → *Pat*.
- **Strucural rules**. Definte pattern of categories that form a larger expression. S → N V.

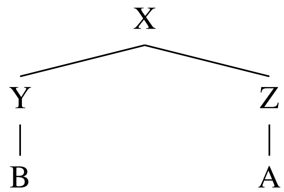
A set of phrase structure rules is said to *generate* tree diagrams.

(5) a. S → N V
b. N → *Pat*
c. V → *screams*



5.1 Names for tree parts

There are standard terms for the parts of the tree and for their relations. They are generally derived from some combination of geneology and actual trees.

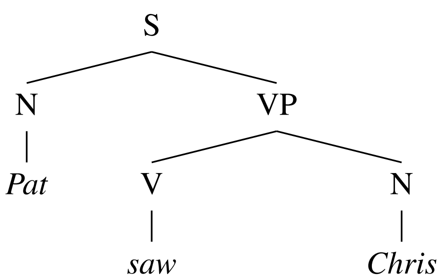


- Each point at the end of a branch is a **node** (X, Y, Z, A, B). (**Branches** connect nodes.)
- The node at the top is the **root node** (X).
- The nodes along the bottom border are the **leaf nodes** (A, B).
- The node at the other end of an upward branch is a **mother node** (Z is the mother node of A).
- The node at the other end of a downward branch is a **daughter node** (A is the daughter node of Z).
- Nodes that share the same mother node are **sister nodes** (Y and Z are sister nodes).

So are A and B sister nodes? Are B and Z sister nodes?

Note: Although you can imagine what they might refer to, nodes are not described as being "aunts" or "nieces" (or "grand-daughters"). Just mothers, daughters, sisters, roots.

(6)



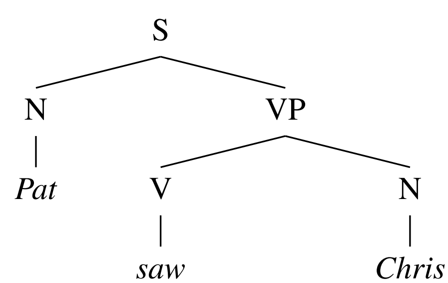
(7)
a. What is the root node?
b. What are the leaf nodes?
c. Which node is the mother of *Pat*?
d. Which node is the daughter of *Pat*?
e. Which node is the mother of the right-most N?
f. Which node is the daughter of right-most N?
g. Which node is the sister of the leftmost N?
h. Which node is the sister of *saw*?
i. Which nodes are daughters of S?

Do the rules in (8) generate the tree in (9)?

(8)

Grammar	
S →	N VP
VP →	V N
V →	*saw*
N →	*Pat*
N →	*Chris*

(9)

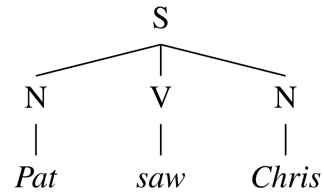


Do the rules in (10) generate the tree in (11)?

(10)

Grammar	
S →	N VP
VP →	V N
V →	*saw*
N →	*Pat*
N →	*Chris*

(11)



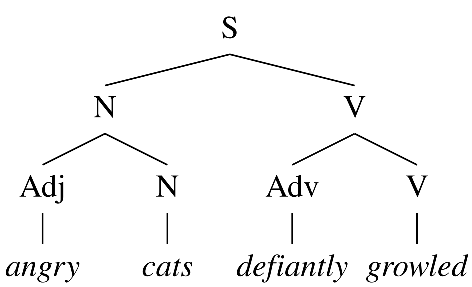
Sometimes a set of PSRs assigns more than one tree to the same string of words.
(This is a good thing, because this allows for a theory of certain kinds of ambiguity.)
How many tree diagrams does (12) generate for the sentence *Pat pats cats at Alexanderplatz*? Draw them.

(12)

Grammar	
S →	N VP
VP →	VP PP
VP →	V N
N →	N PP
PP →	P N
V →	*pats*
N →	*Pat*
N →	*cats*
N →	*Alexanderplatz*
P →	*at*

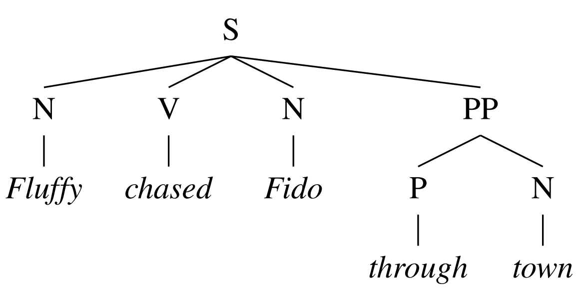
Let's walk through an example of how to go the other way, to take a tree and come up with the PSRs that can generate it. What rules generate (13)?

(13)



And what rules generate (14)?

(14)



(15) Inspect the following set of PSRs and provide the tree diagram that it generates for the sentence *Fluffy chased Fido through town*.

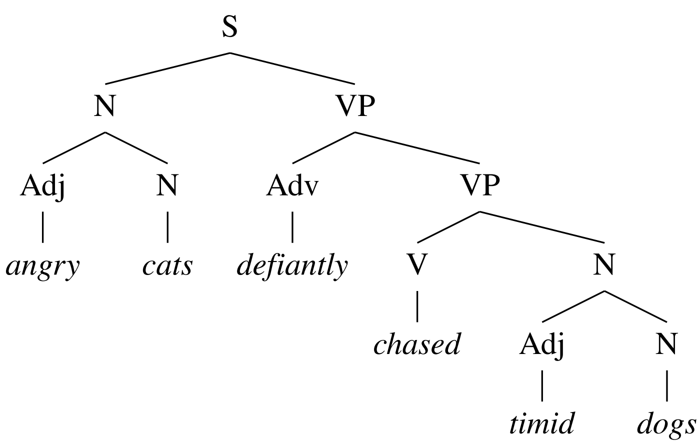
Grammar (15)	
S →	N VP
VP →	VP PP
VP →	V N
PP →	P N
P →	*through*
V →	*chased*
N →	*Fluffy*
N →	*Fido*
N →	*town*

(16) Inspect the following, different, set of PSRs, and provide the tree diagram that it generates for the same sentence: *Fluffy chased Fido through town*.

Grammar (16)	
S →	N VP
VP →	V N P N
P →	*through*
V →	*chased*
N →	*Fluffy*
N →	*Fido*
N →	*town*

Provide a set of PSRs that will generate (17):

(17)



(18) Look at the following set of PSRs and draw the tree diagram that it generates for the sentence *The young girl joyfully played her saxophone in the garden.*

Grammar (18)	
S →	NP VP
NP →	Det N
N →	Adj N
VP →	Adv VP PP
VP →	V NP
PP →	P NP
Det →	*her*
Det →	*the*
P →	*in*
V →	*played*
Adj →	*young*
Adv →	*joyfully*
N →	*girl*
N →	*garden*
N →	*saxophone*

Things to know:

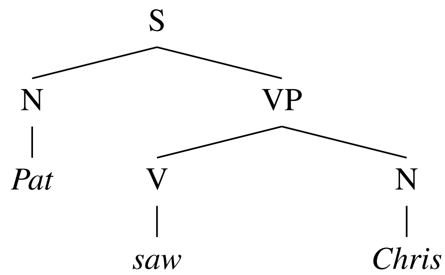
- How to identify the category of a content word using distributional evidence
- The distinction between content words and function words
- How to read and write Phrase Structure Rules
- The relationship between PSRs and tree diagrams
- How to work from PSRs to create trees and vice versa.

6 Bonus: Labeled brackets

Sometimes it is useful to represent tree structures without drawing lines. Like if you're typing an email.

A node in a tree defines a kind of sub-unit of the tree, which we will call a *constituent*.

Bracket notation represents a constituent not by the part of a tree that is under a single node, but instead by putting square brackets around it.



(19) [Pat [saw Chris]]

That's still less information than the tree has. The nodes have labels. And since paired brackets represent the nodes, we can label the pair (and provide brackets for the lexical nodes).

(20) [$_{S}$ Pat [$_{VP}$ saw Chris]]

(21) [$_{S}$ [$_{N}$ Pat] [$_{VP}$ [$_{V}$ saw] [$_{N}$ Chris]]