LING 2010Q – Spring 2017

5 - Syntax

Roberto Petrosino roberto.petrosino@uconn.edu

Oct 19 - Nov 1, 2017

Contents

1 Basic concepts (back to week 1)

(1) Recall the basic questions of linguistic theory:
   a. What exactly does a native speaker *know* about their language? ←
   b. How is that knowledge acquired?
   c. How is that knowledge put to use?

   In this class we focus mostly on the first question, but the other two are important as well.

(2) A challenge in answering the first question is that speakers do not have direct access to the knowledge they have of their language.

1.1 Activity 1: Tacit knowledge.

Consider the following sentences.

1. John expected to surprise him.
2. I wonder who John expected to surprise him.
3. I wonder who John expected to surprise.

In each case, who is surprising whom? Do you know what principles you use to decide this? Did anyone teach you these?

(3) Many properties of language are **recursive**, which means that they can be reiterated over and over again. For example:
   a. [Amanda]
   b. [[Amanda]'s boyfriend]
   c. [[[Amanda]'s boyfriend]'s sister]
   d. [[[[Amanda]'s boyfriend]'s sister]'s son]
   e. [[[[[Amanda]'s boyfriend]'s sister]'s son]'s toy] *...etc*

(4) Another example:
   a. [Dad bought a cat].
   b. [Rose said [(that) Dad bought a cat]].
   c. [Jon thinks [(that) Rose said [(that) Dad bought a cat]]].
   d. [Mark believes [(that) Jon thinks [(that) Rose said [(that) [Dad bought a cat]]]]].
   e. [Luke wonders [whether Mark believes [(that) Jon thinks [(that) Rose said [(that) Dad bought a cat]]]]].

(5) **Recursion** is theoretically **unlimited**; however, our processing memory is limited, so our recursive abilities are limited.

(6) ??[The dog [the boy [the mom [the cat [...] scratched] scolded] found] ran away].

(7) When we learn our first language, our *knowledge of that language* is:

a. **tacit**, namely not explicit: we know things we don't know how we know
b. **complex**: we can form pretty complicated sentences when we are pretty young
c. **untaught**: nobody really *taught* us how to express ourselves in our first language
d. **the poverty of stimulus**: at the age of 7 we typically become *native* in our language, despite impoverished and noisy evidence.

Why is that? Noam Chomsky theorized that we are able to acquire languages because, as humans, we are equipped with a **Universal Grammar** (UG). We can imagine UG as a blueprint of language communication, children refer to when acquiring a language; we all are endowed with it innately. Evidence for that mainly comes from the following:

a. Acquisition is typically rapid and successful across children and across languages.
b. Deep analysis reveals many similarities between superficially very different languages.
c. Any human can learn any human language.

2 Syntax

(8) syntax < Gr. *syn* 'with' + *taxis* 'order'.
It refers to how discrete elements may combine into complex entities, according to a limited set of rules.

(9) The first thing that syntax does is **categorizing**, i.e. dividing the basic building blocks of the combinations into different categories on the basis of shared properties.

SYNTACTIC CATEGORIES	
LEXICAL CATEGORIES	FUNCTIONAL CATEGORIES
nouns	articles (determiners)
adjectives	auxiliaries (be, have)
verbs	modals (can, would, shoud, ...)
adverbs	particles (at, in, against, up, with, ...)
prepositions	...

(10) Let's revise syntactic categories one more time. There are a few tests you can use to identify each of those.

a. Nouns: the _____
b. Verbs: Mary will _____
c. Adjectives: the _____ girl
d. Adverbs: Mary _____ left

e. Prepositions: dance _____ it

(11) After categorizing, syntax needs to seek out how to arrange each of the categories in a coherent way. For example in English:

a. Bart ran.
N V

b. *Ran Bart.
V N

c. *Chased Bart John.
V N N

d. John gave Maggie Lisa.
N V N N

(12) So, to wrap up:

a. Acceptable sentences of English:
(i) N V
(ii) N V N
(iii) N V N N

b. Unacceptable as sentences of English:
(i) N N
(ii) V N
(iii) V N N

2.1 Activity 2: Patterns and categories

State the categories found in sentences below and the patterns involved. Then give another sentence that fits that pattern using completely different words.

1. John came home tired.
2. John heard Maggie clearly.
3. Lisa picked Maggie up.
4. Mary thinks Bart chased Lisa.

3 Syntactic rules and trees

(13) Let's play a bit. Read the following string of words aloud to yourself a couple of times.

a. A enjoy delicious I cereal night of often at bowl.

Done? Now, cover up the paper and try to repeat it back out loud. How did it go? Try to do the same with the following string of words:

b. I often enjoy a delicious bowl of cereal at night.

(14) Cover it up and repeat (13b) out loud — better, right? Why do you think it was so much easier?

- If sentences were just words put together linearly, then we wouldn't have any feeling for meaningful subparts of sentences.
- Every time we produce or understand a sentence, we are grouping words into phrases, and then *phrases* into bigger phrases (yes, that's recursion!)
- Therefore, when we speak, we organize words and phrases in complex, hierarchical structures. Syntacticians aim at discovering those structures, and understand how they work, by coming up with *grammars*.

(15) A grammar is a set of rules and principles that constitute a scientific theory about human linguistic knowledge. These are rewriting rules:

a. S → NP VP
b. S → NP VP NP
c. S → NP VP NP NP

They mean: the symbol *S* (input) may be rewritten as the symbols following the arrow (ouput; in the example above and below, *NP* stands for Noun Phrase, *VP* for Verb Phrase). In case of multiple rewriting rules having the same input:

a. NP → John
b. NP → Lisa
c. NP → Maggie
d. NP → Bart

They may be abbreviated as:
NP → John | Lisa | Maggie | Bart

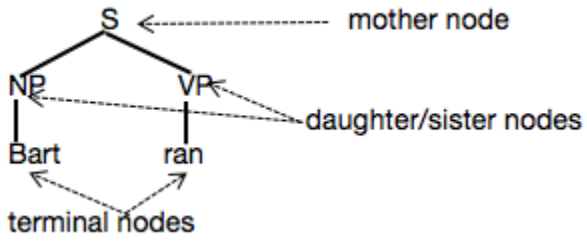
(16) Computations for sentences in English begin with the initial symbol *S*.

'Bart ran'

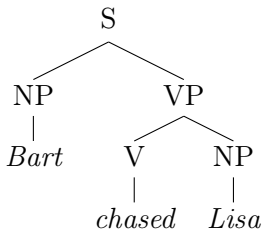
Start: S	S → NP VP
Step 1: NP VP	NP → Bart
Step 2: Bart VP	VP → ran
Step 3: Bart ran	

(17) These are also called **phrase structure rules** (PSR). Compare this rules with the rules that we have used in phonology; the latter were context-sensitive rules, the former are context free.

(18) A convenient way of representing the structure of a sentence is a tree diagram, or phrase marker.



(19) Trees and rules for *Bart chased Lisa*:



S → NP VP

NP → N

N → Bart | Lisa

VP → V NP

V → chased

3.1 Activity 3: Rewriting rules and tree diagrams

Write the syntactic rules and the corresponding tree diagrams for the following sentences. Remember that a sentence always starts with S.

a. Emma kicked the ball.	
Tree Diagram	Rules

b. Christos ate the whole cake.	
Tree Diagram	Rules
c. Steven wrote an article.	
Tree Diagram	Rules
d. Claudia put the book on the desk.	
Tree Diagram	Rules

e. Vanessa went to the Philippines.	

4 Semantic Ambiguity

(20) A sentence is *semantically ambiguous* if it has more than one meaning. In some cases, the semantic ambiguity can be traced to a syntactic ambiguity.

4.1 Activity 4: Ambiguity

For each of the following sentences, let's consider whether they can have one or more interpretations.

1. Old men and women received the medication.
2. Michael hit the man with the halibut.

Syntactic relations among constituents may help disambiguate the meaning. Let's try this out together.

a. Michael hit [the man with halibut]	
Tree Diagram	Rules
Interpretation: _____	Interpretation: _____

b. Michael [hit [the man] with halibut]	
Tree Diagram	Rules
Interpretation: _____	Interpretation: _____

5 Studying grammars

(21) Syntacticians aim at proposing grammars to account for observations made about a given language and test the predictions made by these grammars.

(22) Syntacticians collect data from native speakers, by asking them **judgments** about the well-formedness of sentences of their language.

(23) Judging grammaticality is not simple: naÃŕve native speakers give judgments of acceptability, which may be based on comprehensibility, notions of social correctness, ease of processing, etc. Syntacticians need to take care to decide which judgments reflect intuitions of (un-)grammaticality. Consider the following examples:

a. Colorless green ideas sleep furiously.
b. Colorless sleep furiously ideas green.

What kind of observation can we make here?
Your answer:

(24) In evaluating a grammar, we must see if the grammar predicts judgments of the speakers whose grammatical knowledge we are trying to account for. This includes *both* judgments of grammaticality and ungrammaticality.

a. John ate.
b. *Ate John.

It is not enough for a grammar to generate a), and it must not generate the ungrammatical sentence b).

(25) Typically a grammar constructed for a small set of data will generate sentences beyond those the initial small set of data. These are **predictions** of the grammar.

(26) If grammar G generates sentence S, G predicts that S is *grammatical*.

(27) If grammar G doesn't generate sentence S, G predicts that S is *ungrammatical*.

(28) If a grammar predicts that a grammatical sentence is ungrammatical, then the grammar **undergenerates**. If a grammar predicts that an ungrammatical sentence is ungrammatical, then the grammar **overgenerates**.

(29) The ideal grammar of a language must have the following properties:

a. full coverage of the facts: the more facts explained the better.
b. simplicity: the simpler the theory — fewer rules and symbols — the better.
c. fertility: a good theory makes testable predictions.
d. depth of understanding: a good theory tells us why.

6 Constituency

(30) What is a **constituent?** Intuitively, it is a string of words that we can manipulate (i.e., move around, replace, delete, ?) a single chunk.

(31) We can make use of several **constituency tests**, to identify syntactic constituents. They're pretty intuitive, but it is crucial how to perform them. The reasoning behind all of them is always the same: if a a string of words meaningfully passes

most of them, that string of words can be deemed as a coherent syntactic unit.

(32) **Conjunction Test**. If a string of words can be conjoined, then they are a constituent.

a. Bill talked [to Mary].

We want to find out whether [to Mary] is a constituent. To do that, we can conjoined that phrase with a similar one, e.g.:

a' Bill talked [to Mary] *and* [to John].

b. Another example: John talked to Mary [on Friday].

b' John talked to Mary [on Friday] *and* [on Monday].

(33) **Proform Replacement Test**. If a string of words can be replaced by a proform, they form a constituent.

a. John saw [the cop].

a'. Mary talked to *him* (i.e., the cop).

b. Bill arrived [on Tuesday].

b'. Sen also arrived *then*.

c. Lisa [criticized Alex on Monday].

c'. Lisa [criticized Alex on Monday] and Kate *did so too*.

(34) **Ellipsis Test**. If a string of words can be elided (i.e., left unpronounced), then they form a constituent.

a. Adam will [buy a donut]

a'. Adam will [buy a donut] and Bill won't ~~buy a donut~~.

b. Mary took two [pictures of Greg].

b'. Mary took two [pictures of Greg] and Bill took four ~~pictures of Greg~~.

(35) **Dislocation Test**. If a string of words can be dislocated, then it is a constituent.

a. John will [read a book].

a'. *Read the book*, John will.

b. Mary loves [cats].

b'. *Cats*, Mary loves.

7 Recap

(36) We don't have a direct access to how we correctly speak a language ? we just know we have learnt (**tacit knowledge**) it when we were kids, without being explicitly taught. This led Chomsky to argue that we are endowed with a *Universal Grammar*, a set of principles children use to acquire the language(s) they are being exposed to.

(37) **Syntax** aims at catching how words are coherently organized in phrases, i.e. syntactic chunks. Phrases can be grouped together in order to form bigger phrases, such as sentences.

(38) Syntax deals with organizing an incredible amount of words in a coherent way. However, it doesn't particularly care about the exact words, rather the **syntactic category** each of them belongs to. By looking at how syntax of a language merges words, we can come up with patterns a given language always sticks to.

(39) A **grammar** is the set of rules a language uses to forms phrases. Syntacticians aim at explaining data collected from native speakers through *acceptability judgments*. A grammar is good, if it makes correct *predictions* about what is grammatical and what is not for a native speaker.

(40) *Phrases* can be represented in **tree diagrams**; we can identify phrases by using **constituency tests**.