Ling 120B: Syntax I

Nico(letta) Loccioni

March 28, 2022

# Part I: what is syntax?

Two notions of syntax:

- *As a system of grammar*: the part of language that allows speakers to create and understand phrases and sentences.

- *As a linguistic discipline*: the subfield of linguistics that studies the rules and properties of phrases and sentences of human languages.

What do we mean by rule?

- We do NOT mean *prescriptive* rules (= rules that prescribe how people should speak)

    Do not end sentences with a preposition!
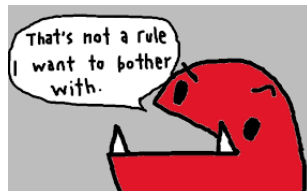
- We are interested in how people really talk. (= descriptive rules)

The syntax of what?

The syntax of *human language*:

- All and only humans have language

    There are no 'primitive' or 'simple' languages

    Even stigmatized languages and varieties of languages are complex, rule-governed systems.

- Superficially very different languages share deep similarities

- Many logically possible grammatical structures do not exist in any language

- Children acquire language similarly across different languages

- Language shares properties with innate instinctual behaviors

The innateness hypothesis: the ability to use language is part of the human biological makeup.

Universal Grammar

- If language is at least partly biologically determined, all humans start off with the same set of tools for learning the languages in our environments

- Different languages employ the same fundamental building blocks in different ways

- The ultimate goal:

    - Find the building blocks

    - Understand the learning algorithm

Universal Grammar (UG)
- The innate building blocks of language
- The learning algorithm that turns them into the grammar of a particular language.

What does a theory of syntax have to explain? I

- <u>Creativity</u>: speakers of a language can create and understand new sentences

  (1) The bald eagle bathed the banana in mustard sauce.

- <u>Order sensitivity</u>: the order of words affects:
    - whether a sentence is possible in a language

      (2) The boy arrived vs. *Boy the arrived.

    - what a sentence means

      (3) John loves Mary $\neq$ Mary loves John

- <u>Ambiguity</u>: A string of words can have multiple meanings

  (4) Mary saw the astronomer with the telescope.

What does a theory of syntax have to explain? II

- <u>Recursion</u>: Sentence structures embed inside one another, allowing infinitely many sentence structures

  (5) I know [that you said [that Mary claimed [that Bill knows [that John loves Carol]]]]

- <u>Restrictions</u>: Not every arrangement of words is a sentence

  (6) a. Mary {claimed / made the claim} that John likes Sue.
       b. Who did Mary claim that John likes ?
       c.*Who did Mary make a claim that John likes ?
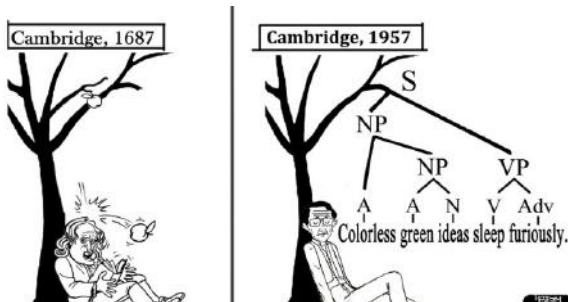
Grammaticality I

- A sentence that is part of a language is called *grammatical*, a sentence that is not is called *ungrammatical*.

- Our data = grammaticality judgements.

  (7)   The boy is tired.   (8) *The boys is tired.

- Meaningless ≠ ungrammatical:

  (9) #Colorless green ideas sleep furiously

Grammaticality II

- Ungrammaticality ≠ hard to understand

  - Ungrammatical sentences that are easy to understand:

    (10) The cookies is delicious.

  - Grammatical sentences that are hard to understand:

    (11) No two people are not on fire.   (from a children's book)

Summary

Goals of syntax: Discover the rules and processes that:

- Allow the production and comprehension of infinitely many sentences

- Capture the relationship between syntactic structure and meaning

- Capture syntactic ambiguity

- Characterize ungrammaticality, explain why some expressions with plausible meanings and structures are unacceptable

How will we do this:

- Revisit the issues and phenomena already that you saw in Ling20, but go deeper

- Flesh out concepts that were only touched briefly before

- Look at new phenomena

- Look at a variety of languages

Consequence: Our understanding of syntactic structure will change as we move along

Part II: Parts of Speech

The traditional/semantic approach I

How do we classify words?
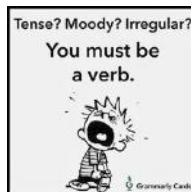
> The semantic approach
> Parts of speech are classified by their meanings.
> - *Noun*: Person, place, or thing.
> - *Verb*: Action, occurrence, or state of being.

Problems with this approach:

- The meaning of *destruction* is an action. But we identify it as a noun.

  (12) The <u>destruction</u> of the city bothered the Mongols.

The traditional/semantic approach II

- *Sincerity* is an attribute. But we identify it as a noun.

  (13) <u>Sincerity</u> is an important quality.

- Some word can change its part of speech depending on where it appears!

  (14) a. I really need to drink a cup of <u>water</u>.
       b. Tomorrow, John will <u>water</u> the plants.

- What is the meaning of *that* in (15)?

  (15) The doctor said <u>that</u> you should eat more vegetables.

How do we know that a particular word belongs to a particular category?

The Jabberwocky
Twas brillig, and the slithy toves
Did gyre and gimble in the wabe;
All mimsy were the borogoves,
And the mome raths outgrabe.

(Lewis Carol from Through the Looking-Glass and What Alice Found There, 1872)

We look at their <u>distribution</u>!

Distributional Criteria

Distibutional Definition

Parts of speech are classified by their distribution.

- Morphological distribution: Affixes appear only on certain kinds of words.
- Syntactic distribution: Position relative to other words.

- Words that have the same or strongly overlapping distributions have the same part of speech

- Words whose distribution does not overlap have different parts of speech

- Since distribution is language specific, the criteria for identifying parts of speech and the part of speech categories themselves can be language specific.

(16) I was happy to ...
   a. ... learn/leave/wander/relax.
   b.*... student/door/wandered/relaxation.
   c.*... underneath/overhead
   d.*... energetic/thoughtful/green/sad
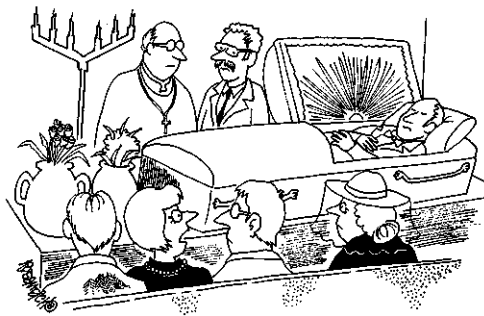
Morphemes

"Isn't it ironic that the first three letters in 'funeral' spell fun?"

Morpheme: meaningful atom = the smallest meaning bearing unit in language:

→ They can be free or bound;

   (17) *man* vs. person-*al*

→ They can be affix or root;

   (18) do-*able* vs. *do*-able

→ They can belong to an open or to a closed class;

   (19) *cat* vs. *for*

→ There are inflectional and derivational affixes.

   (20) danc-*ed* vs. symbol-*ize*

A word can be atomic (if it contains only one morpheme) or complex (if it contains more than one).

   sad vs. sad-ly.

Inflectional vs. Derivational morphology

Derivational	Inflectional
•Plays a lexical (meaningful) role	•Plays only a grammatical role
•Can change part of speech/category	•Never changes base's category
•Never required by grammatical rules	•Usually required by grammatical rules
•Typically indicate semantic relations within the word	•Typically indicate relations between different words in a sentence
•Typically occur before inflectional morphology	•Occur at the end

-able is a derivational affix.   enjoy**able**, do**able**, understand**able**

  → it combines with verbs to make adjectives.

-tion is a derivational affix.   construc**tion**, selec**tion**

  → it combines with verbs to make nouns.

-s is a inflectional affix.   John run**s** 10 miles every day.

  → it combines with verbs when the subject is third person singular.

Examples of inflectional morphology across languages

(21) The cat vs. the cat-**s** vs. zero cat-**s**.

(22) a. Maksim zašiščajt Viktor**a**
    b. Maksim Viktor**a** zašiščajt
    c. Viktor**a** Maksim zašiščajt
    d. Viktor**a** zašiščajt Maksim
    'Maxim defends Victor'   Russian

(23) a. **ma**-kono **ga**-vaga
      4-hands   4-three
      Three hands
    b. **ivi**-baga **vi**-vaga
      8-cats   8-three
      Three cats
    c. **ava**-saza **va**-vaga
      2-men   2-three
      three man
    d. **zi**-dɛka   **zi**-vaga
      10-minutes 10-three
      three minutes   Maragoli, Bantu

Nouns

<u>Derivational suffixes</u>: *-ment* (*basement*), *-ness* (*emptiness*), *-ity* (*sincerity*)...

<u>Inflectional suffixes</u>: *-s* (*cats*), *-es* (*glasses*), *-a* (*addenda*)...

<u>Syntactic distribution</u> Nouns can appear

- as the only word between determiners (the, a, etc.) and verbs

  (24) the <u>package</u> arrived vs. *the big arrived

- as the only word after determiners, adjectives, verbs, prepositions

  (25) John ate the {<u>cake</u>/ *big / *on}

- negated by no

  (26) No <u>trees</u> grow here

Verbs

Derivational suffixes: *-ify* (*solidify*), *-ize* (*regularize*)...

Inflectional suffixes: *-ing* (*Mary was running*), *-en* (*this poem was written by an American soldier*), *-s* (*Peter loves cakes*)...

Syntactic distribution Verbs can appear

- as the first word in a (non-negative) command

  (27) <u>Open</u> the door!

- as negated by 'not'/'n't'

  (28) Mary is{n't/ not} <u>coming</u>.

- modified by adverbs, but not adjectives.

  (29) Peter will {happily/ *happy} <u>come</u>.

Adjectives

<u>Derivational suffixes</u>: *-able* (*readable*), *-ing* (*the dancing cat*), *-al* (*national*)...

<u>Inflectional suffixes</u>: comparative *-er* (*bigger*), superlative *-est* (*biggest*)...

<u>Syntactic distribution</u> Adjectives can appear

- as the only word between determiners and nouns

  (30) the <u>red</u> carpet

- after comparative *more* and superlative *most*

  (31) This dress is more <u>expensive</u> than the other one.

- modified by degree adverbs like *very*

  (32) Mary is very <u>smart</u>

Adverbs

Derivational suffixes: -*ly* (*quickly, frequently...*)

Inflectional suffixes: they generally don't take inflectional morphology.

Syntactic distribution Adverbs can<u>not</u> appear

- between a determiner and a noun

  (33)*the <u>recently</u> election

- after copula 'be'

  (34)*This machine is <u>slowly</u>

Practice

Identifying the category of nonce-words (i.e. made up words) in English sentences.

Example:

a.   The dog **wugged** the ball.   *wug* occurs with past tense suffix *-ed*
b.   Sue **wuggs** balls skillfully.   *wug* combines with 3rd pers. sing. *-s*
c.   *The dog chased the **wug** cat.   it cannot occur between a D and a N
d.   *The **wug** kicked the ball.   *wug* cannot occur as the only word between a determiner and a verb.

Conclusion: I conclude that wug is a verb, based on these observations.

a.   *Every **nork** tower collapsed anyway.   _____
b.   That **nork** washed the dishes.   _____
c.   Eternal **norkdom** is upon us!   _____
d.   Few children believe in **norks**.   _____
e.   Willard may not be a **nork**.   _____
f.   *Why is everyone **norking** the dog.   _____

Conclusion: I conclude that *nork* is _____

Compositionality I

Morphemes combine in a regular way! Let's look at the internal organization of words!

Practice Let's consider the word

D E N A T I O N A L I Z A T I O N

1. The action of denationalizing, or the condition of being denationalized.
2. The action of removing (an industry, etc.) from national control and returning it to private ownership.

- How many free morphemes?

- How many bound ones?

- Which one is the root?

- How many derivational affixes?

- How many inflectional ones?

- What class do they belong to (N, V, Adj...)?

Compositionality II

Claim: only one ordering of the five morphemes in de-nation-al-iz-ation produces a English word.
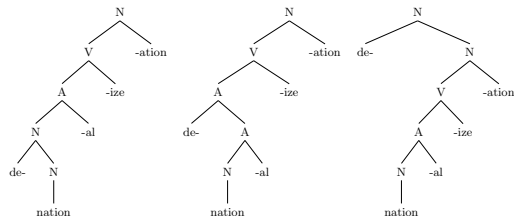
nation → national → nationalize → denationalize → denationalization

Tree structure:
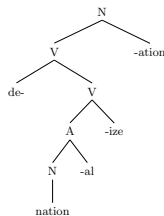
Compositionality III

The following structures are therefore excluded:

They are excluded because the affix *de-* combines with verbs, but not with nouns or adjectives.

Compositionality IV

According to what we said so far, the speaker must know:

a. how each morpheme is pronounced;
b. what kind of morpheme it is (free, prefix, affix..)
c. if an affix, what it combines with (what it c-selects);
d. if an affix, that the c-selected element must be adjacent to the affix;
e. if an affix, what kind of things result after c-selection.

The property in (d) is not in the lexical entry! It is not specific to any lexical entry!
The rest will be encoded in the lexical entry of the morpheme.

Lexical entries - first version

Lexical entry of a morpheme (=the implicit knowledge that a speaker has about it):

-er	c-selects V to form N	kill-er, dry-er ...
-able		understand-able, recycl-able
-ify		test-ify
-ize		

'c-selection', stands for 'category selection'.

The collection of lexical entries = **lexicon**

Practice

Let's try with another word:

(35) overgeneralization

(i) write down the lexical entry for each morpheme

(ii) draw the tree diagram for the whole word