Syntax

LING 200: Introduction to the Study of Language

Hadas Kotek

February 2016

Outline

1. Defining syntax
   - Tacit knowledge (and how we study it)
   - A theory of syntax
   - What syntax isn't
2. Syntactic categories
   - Morphological tests for word class membership
   - Syntactic tests for word class membership
3. Constituency
   - Hierarchical structure
   - Organizing constituents

Slides credit: Jessica Coon, Rebecca Starr

Previously in LING 200...

Previously in LING 200...

- **Morphology**: structure of words.
- We learned about how different types of morphemes come together to form words.
- This week, we will start to learn about how words come together to form phrases and sentences.

Syntax

Let's start with two famous sentences...

"Colorless green ideas sleep furiously."

— Noam Chomsky, 1957

"'Twas brillig, and the slithy toves did gyre and gimble in the wabe."

— Lewis Carroll, 1871

Ask your neighbour:

- What is the difference between these sentences?
- What concept do they both demonstrate?

Syntax

Knowledge of sentence structure

- Chomsky's sentence consists of real words, but its content is strange and doesn't make much sense: "colorless green."
- Carroll's sentence is full of non-words: "brillig."
- But we still recognize that they both use grammatical English sentence structure.

Syntax

Knowledge of sentence structure

- **Conclusion 1:** there is a difference between **grammatical** and **meaningful**.
  - These sentences are not meaningful, but they are grammatical.
- **Conclusion 2:** we cannot determine **grammaticality** by comparing sentences and phrases to those we have seen before.
  - Since we have never seen "ideas sleep" before, we can't rely on past experience to tell us that it is grammatical.

Syntax

Knowledge of sentence structure

- We must have some knowledge of rules that allow us to construct new sentences.
- **Syntax** is the study of the rules underlying sentence structure.

Tacit knowledge

What do syntacticians study?

- The short answer:

  How we combine words to form sentences.

- In other words...

  The **tacit knowledge** we have about how language is structured.

Tacit knowledge

- We know things about our language that we don't know we know ...
- ...and that nobody (parents, teachers, etc.) ever explicitly taught us

➽ As syntacticians, our job is to model the structure of language.

What we don't know we know

Active–passive

(1) a. The baby chattered to us.
b. We were chattered to by the baby.

(2) a. The baby mattered to us.
b. ???

Seem vs. *hope*

(3) a. Anya seemed to speak German.
b. It seemed that Anya spoke German.

(4) a. Anya hoped to speak German.
b. ???

What we don't know we know

Ambiguity

(5) Veronica kept the car in the garage.

- Two possible meanings:
  1. Veronica retained ownership of the car which was in the garage.
  2. Veronica used the garage to store her car.

(6) Which car did Veronica keep in the garage?

- Only one possible meaning.

➽ Our theory of syntax should be able to model this difference.

What we don't know we know

Co-reference

(7) a. Ian$_i$ forgets to bring his pen every time he$_i$ goes to class.

b. He$_i$ forgets to bring his pen every time Ian$_j$ goes to class.

- *Ian* and *he* can co-refer in (7a), not in (7b).

  co-refer = refer to the same individual.

Is this just about which one comes first?

(8) a. Every time Ian goes to class, he forgets to bring his pen.

b. Every time he goes to class, Ian forget to bring his pen.

- *Ian* and *he* can co-refer in both.

➽ Our theory of syntax needs to model this too.

The learnability problem

- Every speaker of English knows the facts above.
- But how do we know this?
- Nobody ever told us: "When someone asks you: *Which car did Veronica keep in the garage?*, it only has one meaning."

- The problem isn't about whether a sentence "*makes sense*" or not.

(9) a. The judges chose a picture of Tyler.
b. Who did the judges choose a picture of?

(10) a. A picture of Tyler won first prize.
b. * Who did a picture of win first prize?

- (10b) makes perfect sense. . . it's just not **grammatical**.

The learnability problem

- Just as nobody ever told us that certain sentences aren't grammatical sentences of English, there are many sentences that we recognize as grammatical English that we've never heard before:

The learnability problem

News from The Associated Press

LONDON (AP) -- A former meerkat expert at London Zoo has been ordered to pay compensation to a monkey handler she attacked with a wine glass in a love spat over a llama-keeper.

- Though probably nobody has ever heard this sentence before, we all know exactly what it means.
- Our knowledge of what is grammatical can't be based just on sentences that we've heard and haven't heard...

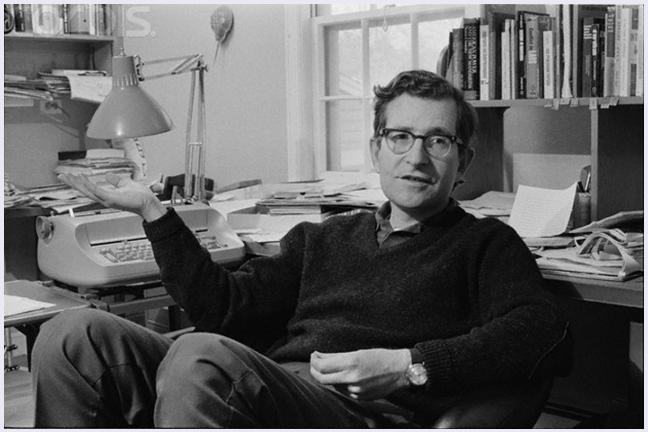
How do we know these things?

Learning vs. acquisition

- Children don't learn to speak languages the way they learn to tie their shoes or set the table.
  - Conscious knowledge (knowledge you can explain to others) is learned.
  - Unconscious knowledge is *acquired*.
- Children must acquire certain individual properties of the language they are exposed to...
- ➽ But there is reason to believe that some of our ability for language comes "hard-wired" in the brain.

Universal Grammar

Noam Chomsky



- Language is an *instinct*.
  - While individual languages (English, French, Hebrew) must be acquired, much of language is *innate*.
  - **Universal Grammar** (UG): the innate human facility for Language.

Big questions (topics of current/ongoing debate!)

- How much of language is innate?
- What are the hard-wired *principles*? and what are the *parameters* along which languages may vary?
- How much of the innateness of language is specific to language? How much follows from general cognitive principles?

UG in the news

Dan Everett and the Pirahã

- 2007 New Yorker article (*link*)
- 2009 response by Nevins, Pesetsky, & Rodrigues (*link*)
- MIT News coverage of debate (*link*)

Language diversity and UG

- 2009 Evans & Levinson article "The myth of language universals" (*library link*)
- . . . and many commentaries
- Short response article by Lisa Matthewson (*link*)

Language vs. language

Some critiques of UG confuse **i-language** and **e-language**...

Language, i-language

The ability of humans to speak; the cognitive system which allows us to acquire, produce, and interpret linguistic utterances.

language, e-language

Particular instantiations of this ability:

- utterances in particular languages
- the words on this slide
- ...

➽ As syntacticians, we will be using *e-language* in order to study the contents of *i-language*...how do we do this?

Collecting data

Corpora

- New York Times corpus (*link*)
- British National Corpus (*link*)
- CHILDES corpus (*link*)
- Google searches

Limitations

- Gives us grammatical sentences, but not *ungrammatical* sentences.
- Some sentences are grammatical, but very rare.
- If we don't find a sentence, we don't know if this is an accident, or if it is missing because it's ungrammatical.
- Many languages do not have a written tradition; no large corpora.

Collecting data

Grammaticality judgements

- A mini psychological experiment performed with a native speaker of the language you are studying (possibly yourself!)

How do these sentences sound?

(11) a. ✔ Who do you think bought tomatoes?
b. ✔ What do you think Morgan bought?
c. * Who do you wonder what bought?

(12) a. ✔ Allison is done with her homework.
b. % Allison is done her homework.

(13) a. # The toothbrush is pregnant.
b. * Toothbrush the is blue.

Developing a theory of syntax

A good theory of syntax should...

- Generate all of the grammatical sentences of a language...
- ...and none of the ungrammatical ones.

Generative Grammar

- Dominant theory of syntax developed by Noam Chomsky and colleagues beginning in the 1950's and continuing today.
- Has gone through many changes: Transformational Grammar (TG), Government & Binding Theory (GB), Principles & Parameters (P&P), and its most recent version, Minimalism.
- Non-transformational grammars: Lexical Functional Grammar (LFG); Head-Drive Phrase Structure Grammar (HPSG); Relational Grammar (RG)...

Developing a theory of syntax

Generative Grammar

- Sentences are generated by an unconscious set of procedures (=*rules*).
  - e.g., rules for how to put words together to form sentences.
- These rules are part of our cognitive abilities.
- ➽ Our jobs as syntacticians: Model these procedures!

What our theory of grammar should not model

Grammatical vs. meaningful

- A sentence can be *grammatical* without being *meaningful*.
- This distinction is illustrated by the famous pair of examples from Chomsky:

(14) a. # Colorless green ideas sleep furiously.
b. * Furiously sleep ideas green colorless.

Compare:

(15) a. The book is red.
b. # The book is exhausted.
c. * Book the is red.

What our theory of grammar should not model

Prescriptive vs. descriptive

- Linguists are not concerned with how people "*should*" speak (prescriptive), but how they actually *do* speak (descriptive).
- A common prescriptive rule in English—one you might have learned in high school English class—is not to end a sentence with a preposition:

(16) a. What are you talking about?
b. Where is Mary from?

- Yet English speakers say things like this all the time. . . if we didn't, nobody would have to tell us not to.
- A good theory of syntax should produce the sentences in (16).

What our theory of grammar should not model

Competence vs. performance

- What we actually say is influenced by all kinds of things...

  (17) I'm going to see a

- I might say (17) and then:
  - get interrupted
  - choke on my food
  - forget what I was talking about
- We don't want our theory of grammar to have to account for these factors
  - **Competence:** What we would say in a perfect world (i-language)
  - **Performance:** What we actually say (e-language)
- Our theory of syntax should only be responsible for *competence*

Competence vs. performance

- Performance also has an effect on the length of our sentences:

(18) a. Jerry's girlfriend
b. Jerry's girlfriend's neighbor
c. Jerry's girlfriend's neighbor's aunt
d. Jerry's girlfriend's neighbor's aunt's cat
e. . . .

- In principle, I could go on forever, but *performance factors* would probably stop me:
  - This class would end
  - I would get hungry
  - People would stop paying attention to me
  - I would eventually die

➽ Nonetheless, we want our syntax to produce sentences that are infinitely long.

Recap

Tacit knowledge

- We have *tacit* (innate) knowledge about languages we speak.
- UG: capacity for **L**anguage that comes hard-wired in the brain.
  - **i-language, Language**: human capacity for language (competence).
  - **e-language, language**: what we actually produce (performance).
- We use *e-language* to study *i-language*...
  - Grammaticality judgments: mini psychological experiments we conduct with native speakers to learn about grammar.

A theory of syntax

- **Generative Grammar**: our goal! A model of grammar which generates all of the grammatical sentences, but none of the ungrammatical ones

Outline

1. Defining syntax
   - Tacit knowledge (and how we study it)
   - A theory of syntax
   - What syntax isn't
2. Syntactic categories
   - Morphological tests for word class membership
   - Syntactic tests for word class membership
3. Constituency
   - Hierarchical structure
   - Organizing constituents

Syntax

Words: syntactic categories

- Reminder: when we introduced morphology, we talked about different **word classes**.
  - Noun (includes Pronoun and proper names)
  - Verb (including auxiliaries)
  - Adjective
  - Adverb
  - Preposition
  - Determiner
  - Conjunction

Syntax

Words: syntactic categories

- How do you know that a word is a noun?
- You might have learned in school:
  - "A noun is a person, place, thing, or idea."

Syntax

Morphological tests for word class membership

- In morphology, we saw that affixes only attach to certain word classes.
- For example, any word that can take plural -s is a noun:
  - businesses, dogs, teachers, theories, ...
- But: if a word cannot take plural -s, does that mean it is not a noun?
  - alumni (*alumnuses), oxen (*oxes)
- No. If a word passes a morphological test, it is a member of a category, but if it fails, we do not know whether it is a member or not.

Syntax

Practice: morphological tests

- With a neighbor, come up with an affix that tests for:
  - adjectives
  - verbs
- In each of these cases, can you come up with an adjective or verb that fails the test?

Syntax

Practice: morphological tests

- **Adjectives**:
  - **-ly**: ADJ + ly → happily, quickly, lightly...
  - Where does it fail? *longly, *farly
- **Verbs**:
  - **-ed**: V + -ed → walked, cited, created...
  - Where does it fail? *eated, *goed

Syntax

Syntactic tests for word class membership

- Here is another way to think about word classes:

  I have a _____.

- This is a **syntactic frame**. What can we put in here?

Syntax

Syntactic frames

I have a
- cat
- career
- conscience
- *furry
- *because
- *recently

- Only **nouns** can fit in this slot.
  - But: if a word cannot fit here, does that mean it is definitely not a noun?

Syntax

Syntactic frames

- Not all nouns can fit in this slot:

  *I have a sand.

- As with morphological tests, we can claim that all words which fill this slot are nouns, but failing this test doesnt mean a word is not a noun.

Syntax

Syntactic frames

- What word class is this a frame for?

He is _____ tall.

very
insanely
quite
unusually

- **Adverbs**!

Syntax

Syntactic frames

- What word class is this a frame for?

_____ cat is purring.

The
A
That
My
One

- This type of test allows us to see that words like “my” and “one” are actually **determiners**.!

Syntax

Can a word belong to more than one class?

- What does this sentence mean:

  Some fish fish.

- Some fish engage in fishing.

Syntax

Can a word belong to more than one class?

- What about...

  Some fish fish fish.

- Some fish engage in fishing for other fish.

Syntax

Can a word belong to more than one class?

- What about...

  Fish fish fish fish fish fish fish.

- Fish that are fished by other fish engage in fishing of other fish, who are themselves fished for by fish.
- Conclusion: we cannot totally rely on the form of the word to determine its function.
  - “Fish” can be either a noun or a verb, depending on its position in the sentence.

Syntax

Summary

- Only words of certain classes can fit in particular slots in syntactic frames.
- This suggests that word classes are crucial in determining the rules for how sentences are put together.

Syntax

What can we conclude from syntactic frame tests?

- What is the word class of the underlined words? Ask your neighbor...

(19) Lisa played the saxophone very loudly. — Adv

(20) Joey finds Math very difficult. — Adj

(21) Maybe the dingo ate your baby. — Det

(22) Kramer used to work at an ice cream store. — Prep

(23) George celebrated Festivus yesterday. — Adv

(24) Yesterday was unseasonably cold. — N

Reminder

Morphological tests for word class membership

- Affixes only attach to certain word classes.
- For example, any word that can take a past tense suffix -ed is a verb:
  - walked, studied, laughed, ...
- But: if a word cannot take -ed, that doesn't mean it's not a verb.
  - ate (*eated), went (*goed)

Reminder

Syntactic tests for word class membership

- We can also construct **syntactic frames** that only certain word classes can fit in:

(25) a. Kai walked yesterday.
b. Kai ran yesterday.
c. Kai sauntered yesterday.

(26) a. * Kai cats yesterday.
b. * Kai purple yesterday.
c. * Kai under yesterday.

- As with morphological tests, we can claim that all words which fill this slot are verbs, but failing this test doesn't mean a word is not a verb.

How do we turn words into sentences?

Hypothesis 1: A sentence is a collection of words

(27) The baby threw steamed broccoli on the floor.
≅ {*The, baby, threw, steamed, broccoli, on, the, floor*}

- But...

(28) a. * On broccoli the baby floor the threw steamed.
b. # The steamed broccoli threw the baby on the floor.

- Order seems to matter!

How do we turn words into sentences?

Hypothesis 2: A sentence is an *ordered* collection of words

(29) Ben threw his broccoli.
$\cong$ <Ben_1, $threw_2$, his_3, $broccoli_4$>

Is this enough?

(30) a. The baby threw steamed broccoli on the floor.
b. * On broccoli baby floor the the threw steamed.

- Hypothesis 2 tells us that the sentences in (30) are *different*...
- But it doesn't tell us *why* one is good and one is bad
- ...or even *which* is good and which is bad

Beyond linear order

- What do we know about sentences besides the order of words?

Not all substrings are equal

(31) a. The baby threw steamed broccoli on the floor.
b. The baby threw steamed broccoli on the floor.

(32) a. The baby threw steamed broccoli on the floor.
b. The baby threw the steamed broccoli on the floor.

- Intuitively, as English-speakers we know that the underlined strings in (31) form "units" in a way that those in (32) do not

  (=*tacit knowledge*)

- ➽ We call these units constituents

Hypothesis 3: Words are organized into hierarchical units

- Using constituency tests, we find that sentences have patterns like this:

(33) The baby threw steamed broccoli on the floor.

- Constituents are always *nested*

- We don't find anything like this:

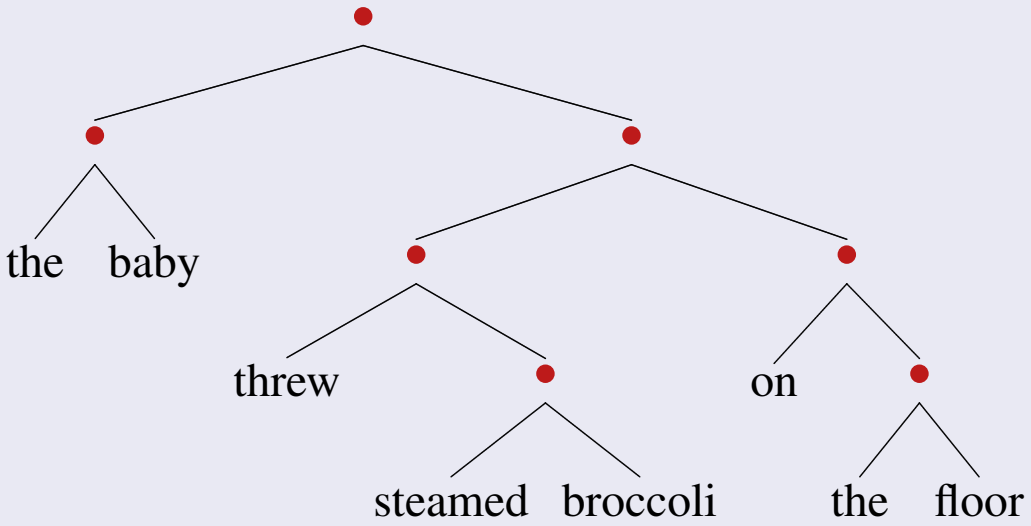
(34) * Word 1 Word 2 Word 3

Representing constituents using trees

(35) The baby threw steamed broccoli on the floor.

- We can represent this using trees:

(36)

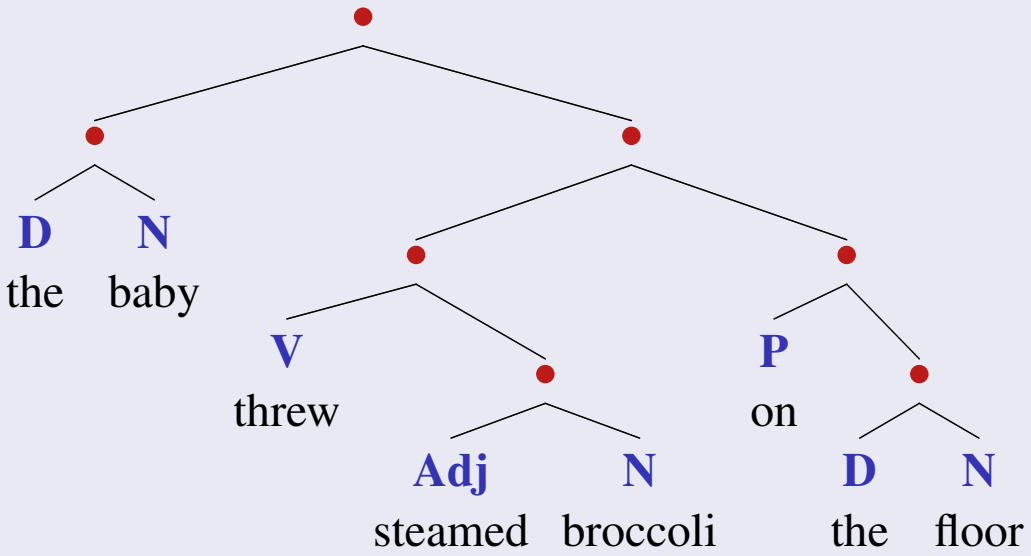


➽ Every node (= •) on the tree represents a constituent

Representing constituents using trees

- Now we can use our part of speech labels:

(37)



- What about the nodes?

Labeling nodes

Remember our distribution tests?

(38) a. Kai [$_V$ ran] yesterday.
b. Kai [threw steamed broccoli] yesterday.
c. Kai [threw steamed broccoli on the floor] yesterday.

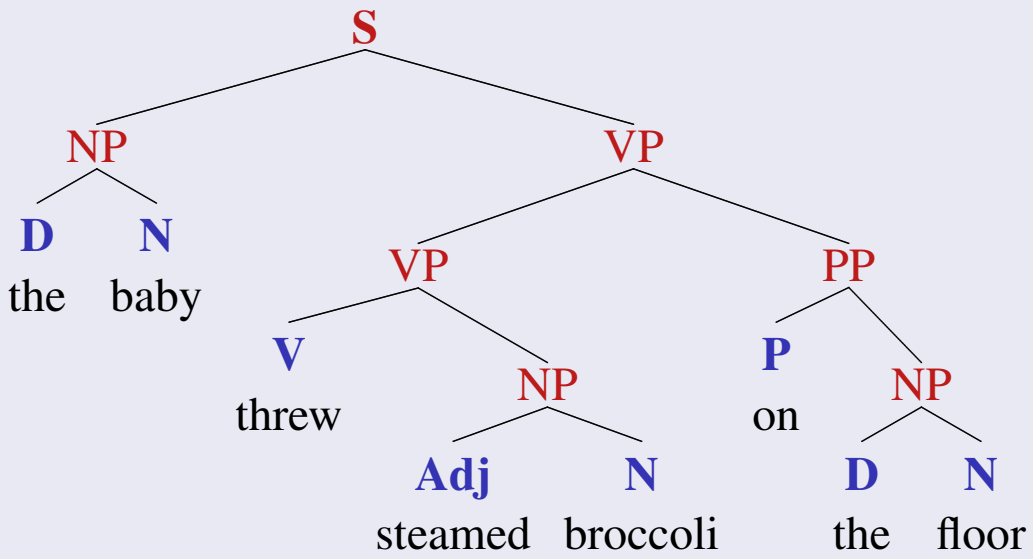
- The string *threw steamed broccoli* acts like a verb = Verb Phrase (VP).
- . . . so does *threw steamed broccoli on the floor*.

(39) a. Sophie saw [$_N$ cows].
b. Sophie saw [the baby].
c. Sophie saw [steamed broccoli].

- The constituents *the baby* and *steamed broccoli* act like nouns = Noun Phrase (NP).

Representing constituents using trees

(40)



For next time. . .

- **Assignment 4** has been posted, due next Wednesday (February 24).
- **Read**: Mihalicek & Wilson "Language Files", chapter 5.4 (pages 215-221), in course pack.