Week 1. Introduction

Syntax

- The scientific study of how sentences are structured in human language.
- Cf. Semantics, Phonology, Phonetics, Pragmatics, Historical Linguistics, …

Things we know

- What we are studying is what we know about language.
- Perhaps this sounds pointless…
- However, there are things we know but don’t know we know. Knowledge of language is largely unconscious knowledge, and our task is to figure out what that knowledge is.

Things we know

- A native speaker of English knows right away which of the following sentences are “English” and which are not.
  - Pat the book lifted.
  - Pat lifted the book.
  - Lifted Pat the book.
  - Pat book the lifted.

Things we know

- A native speaker of English knows right away which of the following sentences are “English” and which are not.
  - *Pat the book lifted.
  - Pat lifted the book.
  - *Lifted Pat the book.
  - *Pat book the lifted.

The * indicates “ungrammatical”
Language is complex

1) Tony threw out the couch.
2) Tony threw the couch out.

- Prepositions can go on either side of the object.

3) Tony stormed out the door.
4) * Tony stormed the door out.

And yet people know this stuff…

5) What did Mary say John bought?
6) What did Mary say that John bought?

- Ok, *that* is optional.

7) Who did Mary say bought coffee?
8) *Who did Mary say *that* bought coffee?

Speakers of English know…

9) Bill thinks Mary is a genius.
10) Her mother thinks Mary is a genius.
11) She thinks Mary is a genius.
12) I asked Mary to buy coffee.
13) What did you ask Mary to buy?
14) I saw the book about aliens on the table.
15) *What did you see the book about on the table?

How do people know this?

- Everyone native speaker of English knows this.
- No native speaker of English (growing up) “You can’t question a subject in a complement embedded with *that*” or “You can’t use a proper name as an object if the subject is co-referential.”

Notice also…

- What people eventually end up with is a system with which they can produce (and rate) sentences. A grammar. Even if you’ve never heard these before, you know which one is “English” and which one isn’t:

16) Eight very lazy elephants drank brandy.
17) Eight elephants very lazy brandy drank.

Positive and negative evidence

- Adults know if a given sentence S is grammatical or ungrammatical. This is part of the knowledge kids gain through language acquisition.
- Kids hear grammatical sentences (positive evidence)
- Kids are not told which sentences are ungrammatical (no negative evidence)
Positive and negative evidence

- One of the striking things about child language is how few errors they actually make.
- For negative feedback to work, the kids have to make the errors (so that it can get the negative response).
- But they don’t make the errors.

Conclusion: People have language

- A linguistic capacity is part of being human.
- Like having two arms, ten fingers, a vision system, humans have a language faculty.
- The language faculty (tightly) constrains what kinds of languages a child can learn.
- =“Universal Grammar” (UG).

So, how come we don’t all speak the same language?

- Languages differ.
- But in light of the learnability problem (and from empirical observation) they must differ only in limited ways.

Word Order

- English, French: Subject Verb Object (SVO)
  - John ate an apple.
  - Pierre a mangé une pomme.
- Japanese, Korean: Subject Object Verb (SOV)
  - Taroo-wa ringo-o tabeta.
  - Chelswu-ka sakwa-lul mekessta.
- Irish, Arabic (VSO), Malagasy (VOS), …

Word order—adverbs

- English: Adverbs before verbs
  - Mary quickly eats an apple.
  - (also: Mary ate an apple quickly)
  - *Mary eats quickly an apple.
- French: Adverbs after verbs
  - Geneviève mange rapidement une pomme.
  - *Geneviève rapidement mange une pomme.

Parameters

- We can categorize languages in terms of their word order: SVO, SOV, VSO.
- This is a parameter by which languages differ.
- The dominant formal theory of first language acquisition holds that children have access to a set of parameters by which languages can differ; acquisition is the process of setting those parameters.
Our tasks

- We’re trying to characterize knowledge of language. What are these things we know but don’t know we know?
- We each have knowledge of this sort of our native language.

The enterprise

- The data we will primarily be concerned with are native speaker intuitions.
- Native speakers, faced with a sentence S, know whether the sentence S is part of their language or isn’t. These intuitions are highly systematic.
- We want to uncover the system (which is unconscious knowledge) behind the intuitions of native speakers—their knowledge of language.

I-language

- Notice: What we are studying is what one person knows. We are studying what the system is behind one person’s pattern of intuitions.
- Of course, speakers growing up in the same community have knowledge of language that is very similar, but language is an individual thing. This is sometimes referred to as a person’s I-language.
- A characterization of the I-languages of a whole community of people is sometimes referred to as an E-language. It is external, it does not characterize any one person’s knowledge, but is a generalization over many people’s I-languages. For example, Parisian French.

Competence

- We are also concerned with what a person knows. What characterizes a person’s language competence. We are in general not concerned here with how a person ends up using this knowledge (performance).
- You still have your language competence when you are sleeping, in the absence of any performance.

Prescriptive rules

- Another thing we need to be cautious of are prescriptive rules. Often prescriptive rules of “good grammar” turn out to be impositions on our native grammar which don’t actually reflect our native competence. After all, why did they need to be rules in the first place?

Prescriptive rules

- Prepositions are things you don’t end a sentence with.
- We want to successfully complete this course.
- Remember: Capitalize the first word after a colon.
- Don’t be so immodest as to say I and John left; say John and I left instead.
- Impact is not a verb.
Prescriptive rules

• When making grammaticality judgments (or when asking others to make grammaticality judgments), we must do our best to factor out prescriptive rules (learned explicitly in school).

Judgments

• Another complicating fact is that a sentence can be bad for any number of reasons, only some of which we are interested in at a given point.
  * Student the meditated happily.
  The pebble meditated happily.
  A Sun rose in the East.
  John wondered who to go with.

Syntax as science

• Here, we will study syntax scientifically. This means, in particular, approaching syntax using the scientific method.
  • Step 1: Gather observations (data)
  • Step 2: Make generalizations
  • Step 3: Form hypotheses
  • Step 4: Test predictions made by these hypotheses, returning to step 1.

A simple introductory example

1) Bill kissed himself.
2) *Bill kissed herself.
3) *Sally kissed himself.
4) Sally kissed herself.

• Try these out. Which ones sound good, which ones don’t?

A simple introductory example

1) Bill kissed himself.
2) *Bill kissed herself.
3) *Sally kissed himself.
4) Sally kissed herself.

• Hypothesis: An anaphor must have an antecedent which agrees with it in gender.
Hypothesis: An anaphor must have an antecedent which agrees with it in gender.

- Let’s test the hypothesis against more data.

5) The robot saw itself in the mirror.
6) *John and Bill saw himself in the mirror.
7) *The boys saw himself in the mirror.
8) *Mary and Jane saw herself in the mirror.
9) John and Bill saw themselves in the mirror.
10) Mary and Jane saw themselves in the mirror.
11) The boys saw themselves in the mirror.

- Our hypothesis only explains (5). What is the generalization?

Hypothesis: An anaphor must have an antecedent which agrees with it in gender.

5) The robot saw itself in the mirror.
6) *John and Bill saw himself in the mirror.
7) *The boys saw himself in the mirror.
8) *Mary and Jane saw herself in the mirror.
9) John and Bill saw themselves in the mirror.
10) Mary and Jane saw themselves in the mirror.
11) The boys saw themselves in the mirror.

- The grammatical ones agree in number, the ungrammatical ones do not. So, we need to revise our hypothesis.

Hypothesis: An anaphor must agree in gender and number with its antecedent

12) The executives gave themselves a raise.
13) *I gave himself a cookie.
14) I gave myself a cookie.
15) You gave myself a cookie.
16) *You gave herself a cookie.
17) *You gave himself a cookie.
18) You gave yourself a cookie.

- Again, our hypothesis doesn’t successfully predict which of these are grammatical and which aren’t. What’s the generalization?

Person

- I is “first person singular”
- You is “second person singular” (you left) or “second person plural” (you left = y’all left)
- He, She is “third person”
- We is “first person plural”
- They is “third person plural”

- Anaphors seem to agree with person. Myself for first person singular, ourselves for first person plural, yourself for second person, himself, herself, or itself for third person singular, themselves for third person plural.

Hypothesis about anaphors

- An anaphor must agree in gender, number, and person with its antecedent.

- This is the hypothesis we will end with, although there are lots of other things about anaphors that this hypothesis doesn’t predict. (We will come back to it).
Levels of adequacy

• If our hypotheses can predict the existence of the grammatical sentences in a corpus (a set of grammatical sentences), it is **observationally adequate**.
• If our hypotheses can predict the native-speaker intuitions about which sentences are grammatical and which are ungrammatical, it is **descriptively adequate**.
• If we can take a descriptively adequate set of hypotheses one step further and account not only for the native speaker judgments but also for how children come to have these judgments, our hypotheses are **explanatorily adequate**.
• It’s this last level that we are hoping to achieve.

Refresher on syntax

• Words can be grouped into **categories** by **part of speech** like noun, verb, adjective, preposition, …
• Parts of speech are determined **distributionally** (traditional “semantic” definitions don’t work)
  – The yinkish dripner blorked quastofically.
  – *Yinkish* is an adjective, *dripner* is a noun, *to blork* is a verb, *quastofically* is an adverb.

Constituents

• The words that make up a sentence like…
  – The students did their syntax assignment.
• …are grouped together into component parts, **constituents**, which function together as a unit.
• Among them, [the students], the do-ers, and [their syntax assignment], the done.

Phrases

• An important type of constituent is the **phrase**, which has at least a central core word (the **head** of the phrase) and often other words or phrases related to the head.
• The category of the head determines the category of the phrase.
• *The happy students* is a noun phrase, headed by the noun *students*. *Happy* modifies *students*, *the* specifies which students.

Sentences

• Complete sentences need to have a subject and a verb.
  – John left.
  – *John.
  – *Left.
  – The happy students left speedily.
• So sentences are made of noun phrases and verb phrases.
Trees

- We can start by drawing the structure of a sentence like this, which means: “John left is a Sentence composed of a Noun Phrase (composed of John) and a Verb Phrase (composed of left).”

(S
  NP  |  VP
  |  N  |  V
  John | left)

(Wait a minute, that’s not an X-bar structure!)

- You may recall from previous courses some discussion of the X-bar schemata that says that all kinds of phrases (NPs, VPs, …, XPs) have a certain consistent shape. We will get to X-bar structures soon, but we are going to work our way up to it and provide some evidence for it.

Finding constituents

- How do we find constituents in a sentence? For many of them, we can guess, but a guess isn’t evidence. If sentences and phrases have structure, we should be able to test for this structure.

Replacement test

- A constituent is a group of words which function as a unit. If you can replace part of the sentence with another constituent (the smallest constituent being a single word), this tells us that the replaced section of the sentence is a constituent.
- This isn’t foolproof, but it usually works if you try to keep the meaning as close as possible.

Replacement test

- The students left.
- They left.
- *The students is a constituent.
  - The students ate the sandwiches.
  - They ate the sandwiches.
  - The students ate them.
  - The students dined.
- [The students] [ate [the sandwiches]].

Sentence fragment test

- Generally, only constituents can be used in the fragmentary response to a question.
- Who ate the sandwiches?
  - The students.
  - *The.
- What did the students do?
  - Ate the sandwiches.
  - *Ate the.
- What did the students eat?
  - The sandwiches.
- [The students] [ate [the sandwiches]].
**Movement tests—clefiting**

- If you can move a sequence of words together to another part of a sentence that means roughly the same thing, that sequence of words is a constituent.

- **Clefting** involves creating a sentence of the shape *It was — who/that — out of your sentence.*

<table>
<thead>
<tr>
<th>Movement tests—preposing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preposing involves creating a sentence by putting a constituent at the beginning of the sentence of this form:</td>
</tr>
<tr>
<td>— is/are who/what/where/… —</td>
</tr>
<tr>
<td>- The students ate the sandwiches.</td>
</tr>
<tr>
<td>- [The students] are who ate the sandwiches.</td>
</tr>
<tr>
<td>- [The sandwiches] are what the students ate.</td>
</tr>
<tr>
<td>- [Eat the sandwiches] is what the students did.</td>
</tr>
<tr>
<td>- *[The] is what students ate the sandwiches.</td>
</tr>
</tbody>
</table>

**Coordination test**

- Generally you can replace a constituent of a certain type X with another constituent of type [X and X].
- [The students] ate the sandwiches.
- [[John] and [the students]] ate the sandwiches.
- The students ate [the sandwiches].
- The students ate [[the sandwiches] and [the eggrolls]].

This shows two things: 1) *The students* is the same kind of constituent as *John* and 2) that *the students and John* are each constituents (as is *John and the students*).

<table>
<thead>
<tr>
<th>When constituency tests fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>- [The students] ate [the sandwiches]].</td>
</tr>
<tr>
<td>- But consider:</td>
</tr>
<tr>
<td>- John prepared and the students ate the sandwiches.</td>
</tr>
<tr>
<td>- [[John prepared] and [the students ate]] the sandwiches?</td>
</tr>
<tr>
<td>- The coordination test failed to reveal the structure.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>When constituency tests fail</th>
</tr>
</thead>
<tbody>
<tr>
<td>- <strong>Moral:</strong> Don’t rely on just one constituency test. Use several tests, assuming that occasionally any given test might yield an anomalous result.</td>
</tr>
</tbody>
</table>
Trees and constituency

• [The students] [ate [the sandwiches]]

Trees and constituency

• [The students] [ate [the sandwiches]]

Trees and constituency

• [The students] [ate [the sandwiches]]

Trees and constituency

• [The students] [ate [the sandwich]]

Phrases and constituents

• The constituents we have identified are the noun phrases and the verb phrase, which have internal pieces as well.

• *The* is a determiner (D).

• *Students* and *sandwiches* are nouns, heading the Noun Phrase (NP).

• *Ate* is a verb, heading the Verb Phrase (VP)

Inside the NP

• The NPs we found have a D and an N
  – The students
  – The sandwiches

• An NP needs to have a *head noun*.

• An NP can have a determiner:
  – NP: (D) N

• But there are lots of other kinds of NPs…
Inside the NP

- The enthusiastic students ate the sandwiches.
- The enthusiastic syntax students ate the sandwiches.
- The enthusiastic syntax students in LX522 ate the sandwiches.

Inside the NP

- The enthusiastic students…
- Enthusiastic is an adjective (modifying students).
- So we need to revise our hypothesis about the components of NP.
- NP: (D) (Adj) N

Inside the NP

- The enthusiastic syntax students…
- We know enthusiastic is an adjective modifying students. What is syntax?
- Syntax seems to also be an adjective modifying students. You can have two adjectives in an NP.
- We need to revise our hypothesis again.

Inside the NP

- The enthusiastic syntax students…
- NP: (D) (Adj) (Adj) N
  - The big red fluffy dog barked.
  - The excited big red fluffy dog barked.
- NP: (D) (Adj) (Adj) (Adj) N
  - This seems to miss a generalization. Instead…
- NP: (D) (Adj+) N

\[ \Rightarrow \text{means "repeat as many times as necessary"} \]

Inside the NP

- The enthusiastic students in LX522…
- We can have in LX522 in an NP too.

- Intermission: What’s in LX522?
- It is a prepositional phrase (PP), with a head preposition (P) in. To a pretty close approximation:
- PP: P NP
  - In the tree… in the big green tree…

Inside the NP

- The enthusiastic students in LX522…
- Back to the problem, revising our hypothesis:

- NP: (D) (Adj+) N (PP)
Modifiers

- **Golden Rule of Modifiers:**
  Modifiers are always attached within the phrase they modify.

- \[ \text{NP, The enthusiastic syntax students in LX522} \]

The bird in the tree on the hill...

This is an NP...
- The bird in the tree on the hill left.
- I left.
- The bird in the tree on the hill and I left.

Trick question: Should we revise our NP rule to NP: (D) (Adj+) N (PP+)?

Answer to the trick question:
Not based on this evidence!

The actual structure...

- We can draw the same information in a tree.

Trees
Trees

• We can draw the same information in a tree.

\[\text{NP} \rightarrow \text{D} \rightarrow \text{N} \rightarrow \text{P} \rightarrow \text{PP} \rightarrow \ldots\]

- The book
- of poems
- with the blue cover

Modifiers

• The book of poems with the blue cover...

• Now should we revise our NP rule to
  NP: (D) (Adj+) N (PP+)?

• Of poems modifies book.

• With the blue cover modifies book.

• Answer: Yes, this is evidence for the new
  NP rule.

Trees

• And the tree…

- The book
- of poems
- with the blue cover

Inside the VP

• A VP always has a head verb (V).
  - Pat left.
  - VP: V

• A VP can sometimes have an adverb.
  - Pat left quickly.
  - VP: V (Adv)

Inside the VP

- Pat quickly left.
- Pat often left early.
- Pat cleverly rarely shouts loudly twice.

• A VP can have any number of adverbs
  before or after the verb.
  - VP: (Adv+ V (Adv+))

Inside the VP

- The students ate the sandwiches.
  - The students ate the sandwiches hungrily.

• VP: (Adv+ V (NP) (Adv+))
  - Chris ate pizza at the café.
  - Chris ate pizza at the café hungrily.
  - Pat bought peanuts at the café for a dollar.
  - Pat bought peanuts at the café for a dollar on Tuesday triumphantly.

• VP: (Adv+ V (NP) (PP+) (Adv+))
What we’ve got…

- NP: (D) (Adj+) N (PP+)
- PP: P (NP)
- VP: (Adv+) V (NP) (PP+) (Adv+)

- The very happy students left.
- [D Adv Adj N] V

- What should we do now?

AdjP and AdvP

- The very happy students left.
- The **Golden Rule of Modifiers** says that modifiers must attach inside the phrase they modify.
  - Very is modifying happy.
  - Very must be inside an Adjective Phrase (AdjP)
    - AdjP: (Adv) Adj
- The students left very quickly.
  - Very is modifying quickly (an adverb).
  - Very must inside an Adverb Phrase (AdvP)
    - AdvP: (Adv) Adv

What we’ve got now…

- NP: (D) (AdjP+) N (PP+)
- PP: P (NP)
- VP: (AdvP+) V (NP) (PP+) (AdvP+)
- AdjP: (Adv) Adj
- AdvP: (Adv) Adv

- Let’s digest this a bit.
- Every phrase has one required element.
- This one required element is the head.
- Every phrase has only one head.

What we’ve got now…

- NP: (D) (AdjP+) N (PP+)
- PP: P (NP)
- VP: (AdvP+) V (NP) (PP+) (AdvP+)
- AdjP: (AdvP) Adj
- AdvP: (AdvP) Adv

- Given this, AdvP looks somewhat suspicious.
- Which Adv is the head in very quickly?
- Suppose: Modifiers are always phrases.
- This requires a revision to get us closer…

What we’ve got now…

- NP: (D) (AdjP+) N (PP+)
- PP: P (NP)
- VP: (AdvP+) V (NP) (PP+) (AdvP+)
- AdjP: (AdvP) Adj
- AdvP: (AdvP) Adv

- Hypothesis:
  - Phrases only consist of one head and modifier phrases.
  - There’s still one non-conformist in our midst.
  - Should D be a DP?
  - Put this on hold. Leave this is the sole exception.
For next time:

• Read:
  – Chapter 1-2
  – (Chapter 2 contains some material we didn’t cover this time but will address next time)

• Homework:
  – Chapter 1: problem 1.
  – Chapter 2: problems 1, 2, and 6.