Ling 240

Syntax 1
agenda for today and tomorrow

• constituency tests
• phrases and categories
• phrase structure rules
• exercise
• get ready for the quiz for tomorrow!
Broader Questions

Examine properties of language

• What’s innate, and how is language acquired?

• Do we have a rule system that generates a language? If so, what properties does mental grammar have?
Syntax

Immediate goal:

- Build model of syntax which characterizes what sentences are *grammatical* (possible)
Syntax

Two fat cats jumped onto the table.

*Cats fat two onto table the jumped.
Syntax

we can produce and understand an infinite number of sentences that we have never heard before
Syntax

- Rules
- Not stored sentences
Syntax

• Rules
  – What do the rules look like?
  – What kind of units (i.e., representations) do the rules apply to?
Methodology

Examine patterns in the data to discover the nature of syntactic representations

observation (tests) → generalization → hypothesize a rule → test the rules
Generalizations

Certain groups of words can

• be questioned (and stand alone in answer to a question)

• be replaced by a pro-form (substitution)

• be moved (topicalized, clefted, etc)

• be deleted
Observation 1: Stand alone

- Certain strings of words can be questioned
Stand alone

d example sentence:

_The fat cats jumped onto the table_
Stand alone

*The fat cats jumped onto the table*

Q: Who jumped onto the table?
A: *the fat cats*
Stand alone

The fat cats jumped onto the table

Q: Where did the fat cats jump?
A: onto the table
Stand alone

The fat cats jumped onto the table

Q: What did the fat cats do?
A: jump onto the table
Stand alone

The fat cats jumped onto the table

Q: What was the relationship between the fat things and the table?

A: *cats jumped onto

NOT POSSIBLE
Stand alone

The fat cats jumped onto the table

Q: *What happened onto the table?*
A: the fat cats jumped

NOT POSSIBLE
Observation 1: Stand alone

- Certain strings of words can be questioned in a sentence and then can stand alone in answer to that question
- Not possible with other strings of words
Observation 2: Pro-forms

- Certain strings of words can be replaced by “pro-forms” and other strings of words cannot
The fat cats had a great time.

The fat cats jumped onto the table.

↑

they
The fat cats had a great time.

They jumped onto the table.
Pronouns:
he, she, it, they, him, her, them
proforms

The fat cats jumped onto the table
proforms

*The fat cats jumped onto the table*

↑↑
they

↑

it
proforms

Pro-verb = do so

*The dogs jumped onto the table, and the fat cats jumped onto the table, too.*

↑

did so
Observation 2: Pro-forms

• Certain strings of words can be replaced by “pro-forms” and other strings of words cannot
Conclusions

• some groups of words seem to form natural units within a sentence and others do not.
Units

The fat cats jumped onto the table
Units = Constituents of the sentence

The fat cats jumped onto the table
Constituent of a sentence = a word or group of words which forms a unit of the sentence.

- individual words are the smallest constituents
- the whole sentence is the largest constituent
Constituents

The fat cats jumped onto the table
Constituents

The fat cats jumped onto the table
Tree Structures

**Tree structures** are used to represent our knowledge of constituents.

**Constituents** are represented by nodes in a tree.
Syntactic representations

- representation of syntactic structure is **hierarchical**
- smaller units are combined to form larger units which are combined to form yet larger units, etc.
Tests for constituency

if a string of words is a unit, then it will behave as a unit

• **stand alone** in answer to a question
• **substitution** by a pro-form
• movement
• deletion
Test 1: Stand alone

If a group of words can stand alone in answer to a question, then the group of words is a constituent.
Test 2: substitution by proforms

If

a group of words can be replaced by a pro-form

and the resulting sentence means the same thing as the original sentence

Then

that group of words is a constituent
Constituents

The fat cats jumped onto the table
Test 3: Movement

If a group of words moves together, then it is a constituent
Test 3: Movement

Clefting:

*The fat cats jumped onto the table*

*It was ___ that

[the fat cats jumped onto the table]*
Test 3: Movement

Clefting:

*The fat cats jumped onto the table*

*It was onto the table that*

[the fat cats jumped _______________ ]
Test 3: Movement

Clefting:

The fat cats jumped onto the table

It was ____ that ....

Whatever can go in the blank is a constituent
Test 3: Movement

Clefting:

*The fat cats jumped onto the table*

*It was jumped onto the table that
[the fat cats ________________ ]*
Test 4: deletion (ellipsis)

The dogs *jumped onto the table*, and the fat cats did __________ too.

*jump onto the table*
Summary

• Passing one test is sufficient to determine constituency.
• Failing all three tests is a requirement for determining non-constituency
• Constituency is determined with respect to a particular sentence
• Meaning must be preserved
Note on Test 1: Stand alone

You have to find the right question!

the question word must be moved to the front of the sentence

The structure of the question has to maintain the structure of the original statement (except for the question word moving to the front)
exercise

- see handout
The fat cats jumped onto the table
Syntactic categories: Lexical and Phrasal

Lexical and phrasal categories are determined based on
- distribution
- internal make-up
Syntactic Categories

Defining syntactic category based on the meaning of the word:
Nouns: → usually refer to places, things, persons, concepts
Verbs: → usually refer to actions or states

But what about:

The shooting last weekend caused a long discussion.

‘shooting’ is not a verb here, though it seems to refer to an action.
Defining syntactic category based on the morphology of the word:

Nouns: → can take a plural /-s/

But what about:

The sheep ran away last night.

‘sheep’ cannot take a plural /-s/

*The sheeps ran away last night.
Defining syntactic category based on the distribution of a word.

Nouns: → can come after a determiner such as ‘the’.

But what about:

Liberty is crucial to our existence.

‘liberty’ cannot take a determiner:

*The liberty is crucial to our existence.
Syntactic Categories

Defining syntactic category based on the distribution of a word

Nouns: \(\rightarrow\) can come after a determiner such as ‘the’.

But what about:

\textbf{Liberty} is crucial to our existence.

‘liberty’ cannot take a determiner:

*The \textbf{liberty} is crucial to our existence.

\[\Rightarrow\] None of the three criteria is infallible but by applying all of them we are usually able to classify any given word.
# Syntactic categories

## Nouns:

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Morphology</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>they characterize:</td>
<td>Can serve as a base for:</td>
<td>They can occur after determiners</td>
</tr>
<tr>
<td>• individuals</td>
<td>• ___ + /-s/ (plural)</td>
<td>• Det ___</td>
</tr>
<tr>
<td>• objects</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Syntactic categories

### Verbs:

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Morphology</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>they characterize:</td>
<td>Can serve as a base for:</td>
<td>They can occur after auxiliaries and at the beginning of a request:</td>
</tr>
<tr>
<td>• actions</td>
<td>• ___ + /ed/</td>
<td>• AUX ___</td>
</tr>
<tr>
<td>• sensations</td>
<td>• ___ + /-ing/</td>
<td>• (please) ___</td>
</tr>
<tr>
<td>• states</td>
<td>• ___ + /-s/</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(third person)</td>
<td></td>
</tr>
</tbody>
</table>
## Adjectives:

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Morphology</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>They designate a property of some entity</td>
<td>Can form the comparative or superlative</td>
<td>They can occur between a determiner and a noun, after a linking verb (e.g., <em>is</em>) and after a degree word</td>
</tr>
<tr>
<td>(entity: book, property of this entity: interesting)</td>
<td>• ___ + /-er/</td>
<td>• Det ___ N</td>
</tr>
<tr>
<td></td>
<td>• ___ + /-est/</td>
<td>• linking verb ___</td>
</tr>
<tr>
<td></td>
<td>• linking verb</td>
<td>• Deg ___</td>
</tr>
</tbody>
</table>
## Syntactic categories

### Adverbs:

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Morphology</th>
<th>Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>They designate a property of some action or state</td>
<td>Are often created by adding a /-ly/ to an adjective</td>
<td>They can (optionally) occur before a verb</td>
</tr>
<tr>
<td>(action: run, property of this entity: quickly)</td>
<td>• [ [Adj] + /-ly/ ]</td>
<td>• __ V</td>
</tr>
</tbody>
</table>
Prepositions:

Examples: in, at, under, on, between, from, to, …

Distribution

Can be modified by ‘right’

• right ___ NP
Syntactic categories

**Auxiliaries:**

- **Modals** - will, can, may, must, should, could
- **and non-modals** - be, have

**Distribution**

- Can occur between a noun and a verb
- Can occur at the beginning of a question
- Can be followed by ‘not’

- NP ____ VP.
- ____ NP VP?
- ____ not
Syntactic categories

**Determiners:**

**Examples:** the, a, this, these, no, ...

**Distribution**

Can occur before a noun

- ____ N.
Syntactic categories

Conjunction:

Examples: and, or, but

Distribution

Can occur between two categories of the same type

• N ___ N
• V ___ V
• Adj ___ Adj
• Adv ___ Adv
Syntactic categories

Degree words:
Examples: too, very, more, quite, ...

Distribution
Can occur before an Adj

• ___ Adj
Syntactic categories

1) I own a very dangerous dog.
2) You have to slice the banana.
3) That boy is tall for his age.
4) We pushed the car into the street.
5) The room was empty.
6) Can you explain the point of this exercise?
7) The lions eagerly chased the pigeons.
8) The happy lion never chases the pigeon.
9) The cats dog our footsteps.
10) This is the end of the exercise.
The fat cats jumped onto the table
The fat cats jumped onto the table
Rules
the cat
*cat the
Rules

First attempt at formulating a rule:

*The word “the” must precede the word “cat”.*
<table>
<thead>
<tr>
<th>the cat</th>
<th>a rock</th>
</tr>
</thead>
<tbody>
<tr>
<td>*cat the</td>
<td>*rock a</td>
</tr>
<tr>
<td>this table</td>
<td>some ideas</td>
</tr>
<tr>
<td>*table this</td>
<td>*ideas some</td>
</tr>
</tbody>
</table>
Rules refer to categories, not individual words
• If there’s a word like “the” and a word like “cat”, then the word like “the” must precede the word like “cat”
Noun Phrases (NPs) can consist of a determiner and a noun (in that order).

NP -> det N

head of NP
Phrase Structure Rules represent two kinds of information

- what elements go into a phrase
- the order of the elements
• complete set of phrase structure rules?
Process for determining phrase structure rules

1. use **constituency tests** to determine what the constituents are
2. draw **trees**
3. formulate **rules** based on trees

**Rules reflect generalizations across trees**
Draw trees & determine PS Rules

1. Pat kissed Robin.
2. Pat pushed the stubborn horse into the barn.
3. Robin talked to the manager over the phone.
4. Robin yelled at the manager from Ohio.
5. Lee bought a nice picture of the unicorn from Robin.
Pat pushed the stubborn horse into the barn.
A set of (preliminary!) Phrase Structure Rules

S -> NP VP
NP -> (Det) (Adj) N (PP)
VP -> V (NP) (PP)
PP -> P NP

What other sentences can be generated from these rules?
So now, what does syntax look like?
Syntax: a model

PS RULES

S → NP VP
NP → (Det) N
VP → V (NP)

Phrase Structure Trees

S
  └── NP
      ├── Det
      └── N
  └── VP
      └── V (NP)

Lexicon

scare, V
the, Det
chicken, N
lion, N

Morphology

V + pst = V
N + pl = N
Testing the Model

Make sure our model of the grammar makes correct predictions

The model should generate all and only the grammatical sentences

• the model should not undergenerate or overgenerate
A set of Phrase Structure Rules

S -> NP VP
NP -> (Det) (AdjP) N (PP)
VP -> V (NP) (PP)
VP -> VP PP
PP -> P NP
Phrase Structure Rules

What other sentences can be generated from these rules?

Are there grammatical sentences that these rules can’t generate?

Does this set of rules account for the properties of language that we set out to explain?
Recursion

A rule is *recursive* if it can apply to its own output. Did you see a recursive property in our rules?

\[
\begin{align*}
  S & \rightarrow NP \ VP \\
  NP & \rightarrow (Det) (AdjP) \ N \ (PP) \\
  VP & \rightarrow V \ (NP) \ (PP) \\
  VP & \rightarrow VP \ PP \\
  PP & \rightarrow P \ NP
\end{align*}
\]
Ambiguity

ambiguous
= having more than one meaning
Ambiguity

Fred went to the bank.

Robin called the manager from Ohio.
Ambiguity

**Lexical ambiguity**: due to a single word having more than one meaning

**Structural ambiguity**: different meanings due to different structures
Ambiguity

Fred jumped from the bank.

called [the manager]

from Ohio
Structural Ambiguity

Kim smashed the statue on the rock.
Kim smashed the statue on the rock
Kim smashed the statue on the rock

cf. recall ‘unlockable’…
Structural Ambiguity

*Kim smashed the statue on the rock.*

Do our phrase structure rules generate both structures?
For tomorrow…

- **Syntax quiz**: constituency, categories and phrase structure rules. This quiz will require more thinking and skills than the phonetics quiz (which was mostly memorization)!