Fundamentals of Grammar
Lexical and Grammatical Categories

Joshua Crowgey

Ling 100
Department of Linguistics
University of Washington

22 June 2012
Review: model of grammar

semantics
  ↓
syntax
  ↓
morphology
  ↓
phonetics, phonology

meaning
  ↑
sentences, phrases
  ↑
words, affixes
  ↑
sounds, sound symbols
Syntax

Definition
Syntax is the study of the organization of words into larger units (e.g., *phrases* and *sentences*), how these units *function*, and the *rules* that govern them.
Syntactic Categories

Definition
Syntactic categories are the classes that words can be divided into based on their characteristics. Words from the same category will behave/function the same way.

You shall know a word by the company it keeps.
– J.R. Firth (1957)

► The __ is on the table.
  ► (book, cup, lemon, pencil ...)
  ► (*lovely, *faster, *detect, *and, *were ...)

► I will __ the book.
  ► (read, write, buy, devour ...)


Syntactic Categories

- **syntactic categories**
  - **lexical**
    - noun
    - verb
    - adjective
    - adverb
  - **grammatical**
    - determiners
    - quantifiers
    - conjunctions
    - auxiliaries
Syntactic categories

- 3 tests (or “clues”) for categorization:
  1. **semantic**: what type of meaning
  2. **morphological**: how the word is structured
  3. **syntactic**: how the word functions, where it occurs in the sentence

- caveat: usually, no single test is enough to determine a word’s category. Tests have exceptions. Use several tests and consider the “preponderance of the evidence”
Using tests to determine lexical/grammatical categories

**semantic test**
Does the word have semantic content, or meaning of its own? Or does the word contribute a functional meaning?

- if YES → lexical category
- if NO → grammatical category

<table>
<thead>
<tr>
<th>Word</th>
<th>Part of Speech</th>
<th>YES/NO</th>
<th></th>
<th>Word</th>
<th>Part of Speech</th>
<th>YES/NO</th>
</tr>
</thead>
<tbody>
<tr>
<td>meaningless</td>
<td>adj</td>
<td>YES</td>
<td>a</td>
<td>det</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>dunk</td>
<td>v</td>
<td>YES</td>
<td>the</td>
<td>det</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>banana</td>
<td>n</td>
<td>YES</td>
<td>is</td>
<td>aux</td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td>quickly</td>
<td>adv</td>
<td>YES</td>
<td>and</td>
<td>conj</td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>

= lexical   = grammatical
Using tests to determine lexical/grammatical categories

**semantic test**
Does the word have semantic content, or meaning of its own? Or does the word contribute a functional meaning?

- lexical categories are the principle content of a sentence
  - in telegraphic speech (seen in headlines), grammatical categories are often dropped
    “TWO CONVICTS EVADE NOOSE, JURY HUNG”

- grammatical categories are the ‘glue’ that hold a sentence together and make it ‘grammatical’
Using tests to determine lexical/grammatical categories

morphological test
Can the word undergo processes of derivation to create new words?

- lexical categories typically do
- grammatical categories typically do not

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>cat</td>
<td>→</td>
<td>catlike</td>
<td>YES</td>
</tr>
<tr>
<td>walk</td>
<td>→</td>
<td>walker</td>
<td>YES</td>
</tr>
<tr>
<td>happy</td>
<td>→</td>
<td>unhappy</td>
<td>YES</td>
</tr>
<tr>
<td>quick</td>
<td>→</td>
<td>quickly</td>
<td>YES</td>
</tr>
<tr>
<td></td>
<td>=</td>
<td>lexical</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Using tests to determine lexical/grammatical categories

morphological test
Can the word undergo processes of derivation to create new words?

- lexical categories are ‘open’ class, they can add new members easily
- grammatical categories are ‘closed’ class, their membership is finite and limited
Using tests to determine lexical/grammatical categories

**syntactic test**
How does the word function? Where does it occur in the sentence?

- lexical categories tend to serve as *heads* of phrases
- grammatical categories tend to serve as *introducers* of phrases, they often modify heads
Syntactic categories

traditional means of determining category

- noun: “a person, place or thing”
- verb: “an action, event, state of being”

noun or verb?

- The **action** of the defendant surprised the judge. (‘action’ is a noun!)
- The gamblers **place** their bets on the table. (‘place’ is a verb!)
- **Swimming** is fun. (noun)
- I am **swimming**. (verb)

Let’s try out our new tests and set the traditional method aside...
Syntactic categories

We’ve used our tests to determine the category “type”, but we can also use them to determine the specific category.

**semantic test**

Insert other items of similar semantic content to see if they fit in this location.

- The *parrot* was green.
  n. (parakeet, bird, finch) , *v ( *repeat, *mimic)
- The child would *parrot* her teacher.
  v. (repeat, echo, mimic)
  *n. (*parakeet, *bird, ...)

But what about these examples:

- Charlene said *that* she was tired.
- The *dripner* blorked frabjously.
Syntactic categories

morphological tests

How is the word formed? What sort of affixes can it take?

- nouns can typically occur with the plural suffix -s
  - action actions
  - event events

- BUT, not all nouns can ...
  - serenity *serenities
  - happiness *happiniesses

- ...and some verbs can too.
  - enjoy *enjoys
  - appear *appears
Syntactic categories

syntactic tests
How does the word function (is it a head)? Where can it occur in the sentence?

- **Running** makes me tired.
  - *n.* → subject; precedes verb
- The jarny blored the **nibnik** frabjously.
  - *n.* → nouns can be introduced by *the*
Can *snicker* be used as a noun?

**semantic test**
Synonyms? *laugh, giggle, snort*
can be N. or V.

**morphological test**
Can we add -s? $\rightarrow$ *snickers*, yes

**syntactic test**
Can it be a subject, or follow ‘the’? $\rightarrow$
*The snicker I heard made me laugh too.* yes
Syntactic categories

summary

subclasses

**lexical categories**
- have semantic content
- can undergo derivation
- are ‘open’ class
- are typically heads

**grammatical categories**
- provide functional content
- do not typically undergo derivation
- are ‘closed’ class
- are introducers of heads
Grammar is a system in which tout se tient. [everything hangs together].
–Antoine Meillet (or Ferdinand de Saussure)

We can also apply our tests to determine the specific category of a word in question. Moving forward, we’ll be stepping through the syntactic categories of English and begin to build up a model of English phrase structure. But first, we will introduce the framework that we will use to formalize questions of phrase structure. The next topic: formal languages and phrase structure grammars!
Formal models of language

- we’ve got a theory of words, now we need a theory of how words combine
- from the point of view of formal grammar, a language is a set of strings defined over some alphabet.
- what’s a string?
- what’s a set?
strings and sets

- a set is a collection of elements which:
  - has a size (a “cardinality”, written $|S|$)
  - can be empty ($|S| = 0$, often written $\emptyset$)
  - can be finite or non-finite
  - contains no inherent order
  - contains no “repeated elements”
  - $\{a,b,c\} = \{b,a,c\} = \{c,a,b\}$

- a string is an ordered sequence of symbols
  - “abc” $\neq$ “bac” $\neq$ “cab”
  - a string has no internal structure other than the ordering of its symbols
a formal language is a set of strings

- a language is a set of strings defined over some alphabet
- an alphabet is a (finite!) set of symbols
- ∅ is a language
- {‘a’} is a language
- {‘who are you’, ‘i am sam’} is a language
- but the set of strings of English (all natural languages) is non-finite. how can we define this set?
sets can be defined in two ways:

- by listing (or enumerating) the elements:
  - \{ ‘apples’, ‘bananas’, ‘papaya’\}
  - only useful for finite sets

- by predication, by providing a pattern:
  - \{x | x is a fruit\}
  - \{x | x is a number\}

natural languages are not finite! so we’ll have to find a way to describe the pattern if we want to characterize the set of English sentences
complexity classes, chomsky hierarchy

- a formal grammar (as opposed to mental grammar) is a mathematical object for characterizing sets of strings
- a formal grammar has at least these components:
  - a set of “production rules” of the form $A \rightarrow B$ where $A$ and $B$ are strings of symbols
  - a set of “nonterminal symbols”
  - a set of “terminal symbols” (the words of the alphabet)
  - a “start” symbol (a distinguished nonterminal)
- The Chomsky hierarchy groups formal grammars according to complexity classes

![Chomsky Hierarchy Diagram]

- recursively enumerable
- context-sensitive
- context-free
- regular
application to natural language

- a **language** is a set of **strings** defined over some **alphabet**
- the relevant **alphabet** for defining the strings of English is the **words** of English as provided by the morphology
- the syntactic categories we’ve been discussing will be part of the set of nonterminal symbols
some examples

A toy grammar:

\[
S \rightarrow A
\]

\[
A \rightarrow Ab
\]

\[
A \rightarrow \epsilon
\]

What language does this grammar “generate”?

\{‘b’, ‘bb’, ‘bbb’, ...\}
a derivation tree

Grammar:

\[ S \rightarrow A \]
\[ A \rightarrow Ab \]
\[ A \rightarrow \epsilon \]

Deriving the string ‘b’ from this grammar:

```
  S
   |
  --
   A
   |
  --
  A b
   |
  --
  \epsilon
```
some examples

Another toy grammar:

<table>
<thead>
<tr>
<th>nonterminals</th>
<th>terminals</th>
<th>rules</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>‘cat’</td>
<td>S → NP VP</td>
</tr>
<tr>
<td>NP</td>
<td>‘milk’</td>
<td>NP → Det N</td>
</tr>
<tr>
<td>VP</td>
<td>‘drank’</td>
<td>VP → V NP</td>
</tr>
<tr>
<td>V</td>
<td>‘sleeps’</td>
<td>VP → V</td>
</tr>
<tr>
<td>N</td>
<td>‘the’</td>
<td>N → ‘cat’</td>
</tr>
<tr>
<td>Det</td>
<td></td>
<td>N → ‘milk’</td>
</tr>
</tbody>
</table>

What strings are generated by this grammar?
Draw the phrase structure trees to prove it.