Last update: April 29, 2010

Communication and Language

CMSC 421: Chapter 22

Outline

- Communication
- Grammar
- Syntactic analysis
- Problems

Communication

"Classical" view (pre-1953):
language consists of sentences that are true/false (cf. logic)

"Modern" view (post-1953):
language is a form of action

Wittgenstein (1953) *Philosophical Investigations*
Austin (1962) *How to Do Things with Words*
Searle (1969) *Speech Acts*

Why?

Communication

"Classical" view (pre-1953):
language consists of sentences that are true/false (cf. logic)

"Modern" view (post-1953):
language is a form of action

Wittgenstein (1953) *Philosophical Investigations*
Austin (1962) *How to Do Things with Words*
Searle (1969) *Speech Acts*

Why? **To change the actions of other agents**

Speech acts

Speech acts are attempts to achieve the speaker's goals:

Inform	"There's a pit in front of you."
Query	"Can you see the gold?"
Request	"Could you open my file for me?"
Command	"Pick it up."
Promise	"I'll share the gold with you."
Acknowledge	"OK"

Planning a speech act requires knowledge of

- Situation
- Semantic and syntactic conventions
- Hearer's goals, knowledge base, and rationality

Stages in communication (informing)

Intention	S wants to inform H that P
Generation	S selects words W to express P in context C
Synthesis	S utters words W
Perception	H perceives W' in context C'
Analysis	H infers possible meanings $P_1, \ldots P_n$
Disambiguation	H infers intended meaning P_i
Incorporation	H incorporates P_i into KB

How could this go wrong?

Stages in communication (informing)

Intention	S wants to inform H that P
Generation	S selects words W to express P in context C
Synthesis	S utters words W
Perception	H perceives W' in context C'
Analysis	H infers possible meanings $P_1, \ldots P_n$
Disambiguation	H infers intended meaning P_i
Incorporation	H incorporates P_i into KB

How could this go wrong?

- Insincerity (S doesn't believe P)
- Speech wreck ignition failure
- Ambiguous utterance
- Differing understanding of current context ($C \neq C'$)

Grammar

Vervet monkeys, antelopes etc. use isolated symbols for sentences

$\Rightarrow$ restricted set of communicable propositions, no *generative capacity*

(Chomsky (1957): *Syntactic Structures*)

Grammar specifies the compositional structure of complex messages

e.g., text (linear), speech (linear), music (multi-dimensional)

A *formal language* is a set of *strings* of *terminal symbols*

Each string in the language can be analyzed/generated by the grammar

The grammar is a set of *rewrite rules*, e.g.,

$$S \rightarrow NP\ VP$$
$$Article \rightarrow \boldsymbol{the} \mid \boldsymbol{a} \mid \boldsymbol{an} \mid \ldots$$

Here S is the *sentence* symbol, NP and VP are *nonterminals*

Grammar types

Regular: $nonterminal \rightarrow \boldsymbol{terminal}[nonterminal]$

$$S \rightarrow \boldsymbol{a}S$$
$$S \rightarrow \Lambda$$

Context-free: $nonterminal \rightarrow string$

$$S \rightarrow \boldsymbol{a}S\boldsymbol{b}$$

Context-sensitive: $thing_1\ nonterminal\ thing_2 \rightarrow thing_1\ string\ thing_2$

$$ASB \rightarrow AA\boldsymbol{a}BB$$

Recursively enumerable: no constraints

Natural languages probably context-free, parsable in real time

Wumpus lexicon

$$
\begin{aligned}
\textit{Noun} &\rightarrow \textbf{\textit{stench}} \mid \textbf{\textit{breeze}} \mid \textbf{\textit{glitter}} \mid \textbf{\textit{nothing}} \\
&\mid \textbf{\textit{wumpus}} \mid \textbf{\textit{pit}} \mid \textbf{\textit{pits}} \mid \textbf{\textit{gold}} \mid \textbf{\textit{east}} \mid \ldots \\
\textit{Verb} &\rightarrow \textbf{\textit{is}} \mid \textbf{\textit{see}} \mid \textbf{\textit{smell}} \mid \textbf{\textit{shoot}} \mid \textbf{\textit{feel}} \mid \textbf{\textit{stinks}} \\
&\mid \textbf{\textit{go}} \mid \textbf{\textit{grab}} \mid \textbf{\textit{carry}} \mid \textbf{\textit{kill}} \mid \textbf{\textit{turn}} \mid \ldots \\
\textit{Adjective} &\rightarrow \textbf{\textit{right}} \mid \textbf{\textit{left}} \mid \textbf{\textit{east}} \mid \textbf{\textit{south}} \mid \textbf{\textit{back}} \mid \textbf{\textit{smelly}} \mid \ldots \\
\textit{Adverb} &\rightarrow \textbf{\textit{here}} \mid \textbf{\textit{there}} \mid \textbf{\textit{nearby}} \mid \textbf{\textit{ahead}} \\
&\mid \textbf{\textit{right}} \mid \textbf{\textit{left}} \mid \textbf{\textit{east}} \mid \textbf{\textit{south}} \mid \textbf{\textit{back}} \mid \ldots \\
\textit{Pronoun} &\rightarrow \textbf{\textit{me}} \mid \textbf{\textit{you}} \mid \textbf{\textit{I}} \mid \textbf{\textit{it}} \mid \ldots \\
\textit{Name} &\rightarrow \textbf{\textit{John}} \mid \textbf{\textit{Mary}} \mid \ldots \\
\textit{Article} &\rightarrow \textbf{\textit{the}} \mid \textbf{\textit{a}} \mid \textbf{\textit{an}} \mid \ldots \\
\textit{Preposition} &\rightarrow \textbf{\textit{to}} \mid \textbf{\textit{in}} \mid \textbf{\textit{on}} \mid \textbf{\textit{near}} \mid \ldots \\
\textit{Conjunction} &\rightarrow \textbf{\textit{and}} \mid \textbf{\textit{or}} \mid \textbf{\textit{but}} \mid \ldots \\
\textit{Digit} &\rightarrow \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \mathbf{4} \mid \mathbf{5} \mid \mathbf{6} \mid \mathbf{7} \mid \mathbf{8} \mid \mathbf{9}
\end{aligned}
$$

Divided into *closed* and *open* classes

Wumpus lexicon

$$
\begin{aligned}
\textit{Noun} &\rightarrow \textbf{stench} \mid \textbf{breeze} \mid \textbf{glitter} \mid \textbf{nothing} \\
&\quad \mid \textbf{wumpus} \mid \textbf{pit} \mid \textbf{pits} \mid \textbf{gold} \mid \textbf{east} \mid \ldots \\
\textit{Verb} &\rightarrow \textbf{is} \mid \textbf{see} \mid \textbf{smell} \mid \textbf{shoot} \mid \textbf{feel} \mid \textbf{stinks} \\
&\quad \mid \textbf{go} \mid \textbf{grab} \mid \textbf{carry} \mid \textbf{kill} \mid \textbf{turn} \mid \ldots \\
\textit{Adjective} &\rightarrow \textbf{right} \mid \textbf{left} \mid \textbf{east} \mid \textbf{south} \mid \textbf{back} \mid \textbf{smelly} \mid \ldots \\
\textit{Adverb} &\rightarrow \textbf{here} \mid \textbf{there} \mid \textbf{nearby} \mid \textbf{ahead} \\
&\quad \mid \textbf{right} \mid \textbf{left} \mid \textbf{east} \mid \textbf{south} \mid \textbf{back} \mid \ldots \\
\textit{Pronoun} &\rightarrow \textbf{me} \mid \textbf{you} \mid \textbf{I} \mid \textbf{it} \mid \textbf{they} \mid \textbf{y'all} \mid \ldots \\
\textit{Name} &\rightarrow \textbf{John} \mid \textbf{Mary} \mid \textbf{CollegePark} \mid \textbf{UMD} \mid \ldots \\
\textit{Article} &\rightarrow \textbf{the} \mid \textbf{a} \mid \textbf{an} \mid \ldots \\
\textit{Preposition} &\rightarrow \textbf{to} \mid \textbf{in} \mid \textbf{on} \mid \textbf{near} \mid \ldots \\
\textit{Conjunction} &\rightarrow \textbf{and} \mid \textbf{or} \mid \textbf{but} \mid \ldots \\
\textit{Digit} &\rightarrow \mathbf{0} \mid \mathbf{1} \mid \mathbf{2} \mid \mathbf{3} \mid \mathbf{4} \mid \mathbf{5} \mid \mathbf{6} \mid \mathbf{7} \mid \mathbf{8} \mid \mathbf{9}
\end{aligned}
$$

Divided into *closed* and *open* classes

Wumpus grammar

$S \rightarrow$	$NP\ VP$	I + feel a breeze
$\mid$	$S\ Conjunction\ S$	I feel a breeze + and + I smell a wumpus
$NP \rightarrow$	$Pronoun$	I
$\mid$	$Noun$	pits
$\mid$	$Article\ Noun$	the + wumpus
$\mid$	$Digit\ Digit$	3 4
$\mid$	$NP\ PP$	the wumpus + to the east
$\mid$	$NP\ RelClause$	the wumpus + that is smelly
$VP \rightarrow$	$Verb$	stinks
$\mid$	$VP\ NP$	feel + a breeze
$\mid$	$VP\ Adjective$	is + smelly
$\mid$	$VP\ PP$	turn + to the east
$\mid$	$VP\ Adverb$	go + ahead
$PP \rightarrow$	$Preposition\ NP$	to + the east
$RelClause \rightarrow$	$\boldsymbol{that}\ VP$	that + is smelly

Parse trees

Exhibit the grammatical structure of a sentence

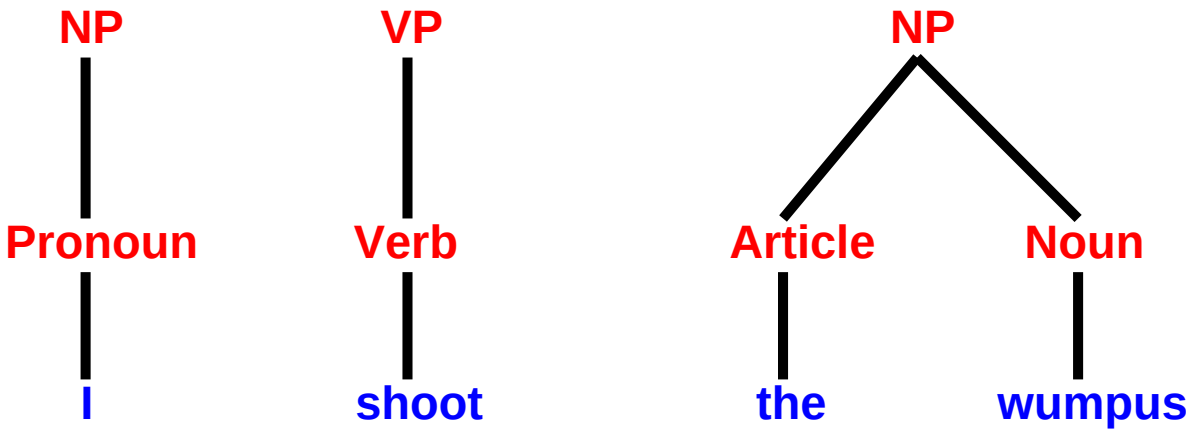
I **shoot** **the** **wumpus**

Parse trees

Exhibit the grammatical structure of a sentence

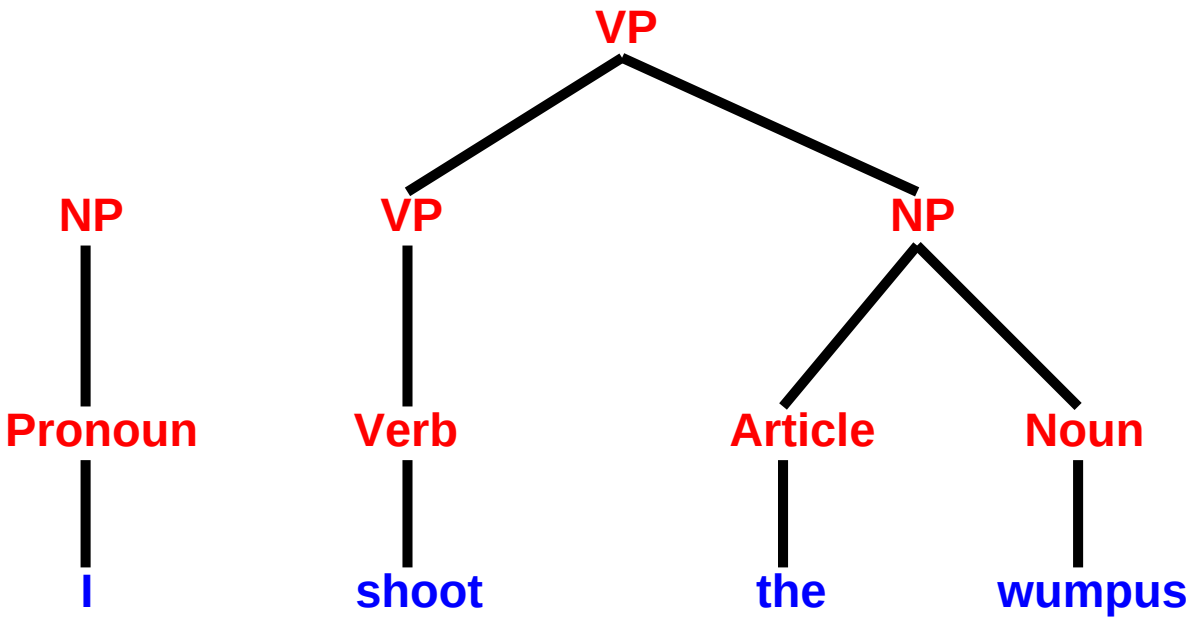
Parse trees

Exhibit the grammatical structure of a sentence



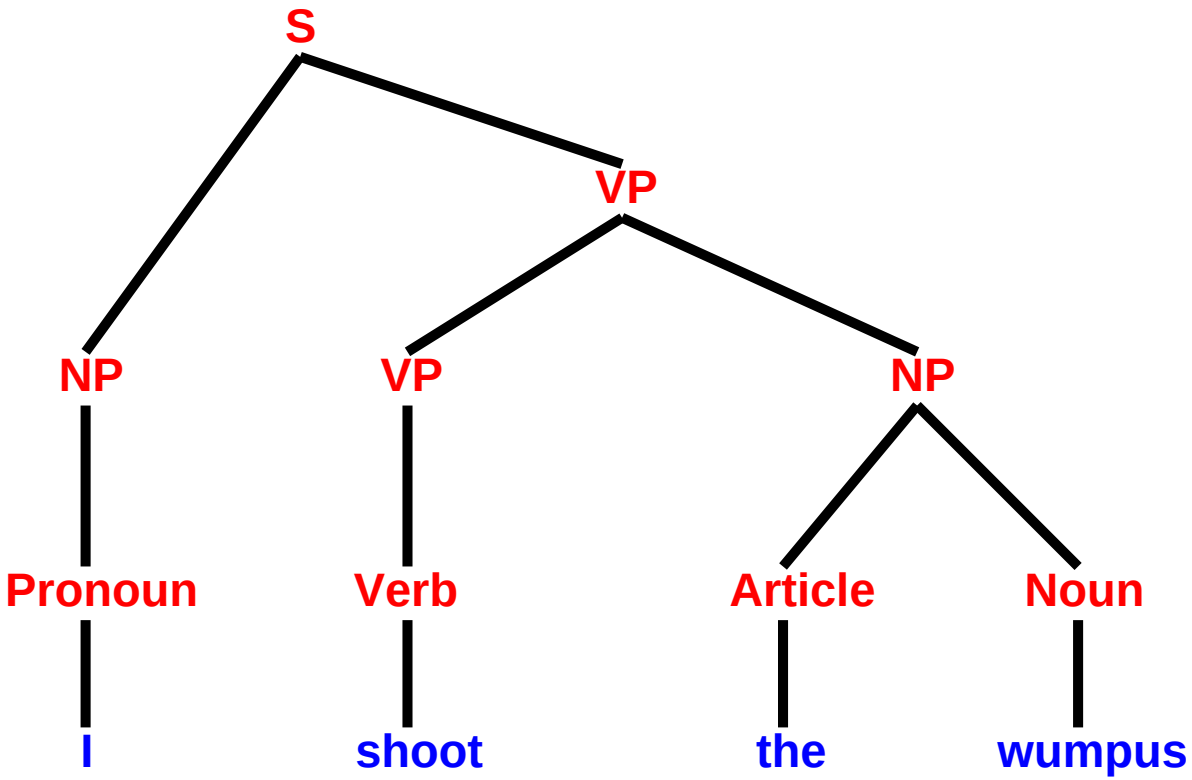
Parse trees

Exhibit the grammatical structure of a sentence



Parse trees

Exhibit the grammatical structure of a sentence



Context-free parsing

Bottom-up parsing works by replacing any substring that matches the RHS of a rule with the rule's LHS

e.g., replace $VP\ NP$ with VP

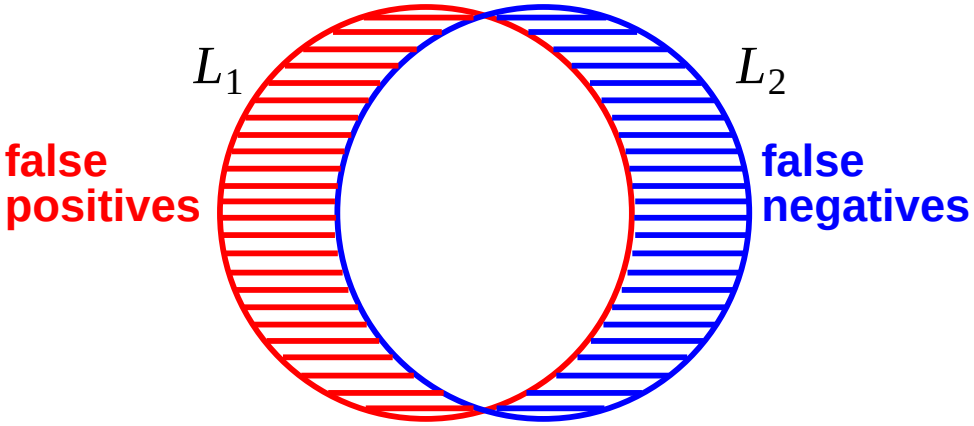
Efficient algorithms (e.g., chart parsing, Section 22.3) $O(n^3)$ for context-free, run at several thousand words/sec for real grammars

Context-free parsing $\equiv$ Boolean matrix multiplication (Lee, 2002)

$\Rightarrow$ unlikely to find faster practical algorithms

Grammaticality judgements

Formal language L_1 may differ from natural language L_2



Adjusting L_1 to agree with L_2 is a learning problem!

* the gold grab the wumpus
* I smell the wumpus the gold
  I give the wumpus the gold
* I donate the wumpus the gold

Intersubjective agreement somewhat reliable, independent of semantics
Real grammars 10–500 pages, insufficient even for "proper" English

Logical grammars

BNF notation for grammars too restrictive:

- difficult to add "side conditions" (number agreement, etc.)
- difficult to connect syntax to semantics

Idea: express grammar rules as logic

$X \rightarrow YZ$	becomes	$Y(s_1) \wedge Z(s_2) \Rightarrow X(Append(s_1, s_2))$
$X \rightarrow \boldsymbol{word}$	becomes	$X([\text{"}\boldsymbol{word}\text{"}])$
$X \rightarrow Y \mid Z$	becomes	$Y(s) \Rightarrow X(s) \quad Z(s) \Rightarrow X(s)$

Here, $X(s)$ means that string s **can be interpreted** as an X

Logical grammars, continued

Now it's easy to augment the rules

$$NP(s_1) \ \wedge \ Human(Ref(s_1)) \wedge VP(s_2) \\ \Rightarrow \ NP(Append(s_1, [\text{“}\boldsymbol{who}\text{”}], s_2))$$

$$NP(s_1) \ \wedge \ Number(s_1, n) \wedge VP(s_2) \wedge Number(s_2, n) \\ \Rightarrow \ S(Append(s_1, s_2))$$

Parsing is reduced to logical inference:

ASK(KB, $S([\text{“}\boldsymbol{I}\text{”} \ \text{“}\boldsymbol{am}\text{”} \ \text{“}\boldsymbol{a}\text{”} \ \text{“}\boldsymbol{wumpus}\text{”}])$)

(Can add extra arguments to return the parse structure, semantics)

Generation simply requires a query with uninstantiated variables:

ASK(KB, $S(x)$)

If we add arguments to nonterminals to construct sentence semantics, NLP generation can be done from a given logical sentence:

ASK(KB, $S(x, At(Robot, [1, 1]))$)

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;
"Mary hit John" $\neq$ "John hit Mary"

Can you figure out the meaning of the following?

> "And since I was not informed—as a matter of fact, since I did not know that there were excess funds until we, ourselves, in that checkup after the whole thing blew up, and that was, if you'll remember, that was the incident in which the attorney general came to me and told me that he had seen a memo that indicated that there were no more funds."

–President Ronald Reagan, April 28, 1987
http://www.reagan.utexas.edu/archives/speeches/1987/042887e.htm

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;
"Mary hit John" $\neq$ "John hit Mary"

Can you figure out the meaning of the following?

> "And **since** I was not informed—as a matter of fact, **since** I did not know that there were excess funds **until** we, ourselves, in that checkup after the whole thing blew up, and that was, if you'll remember, that was the incident in which the attorney general came to me and told me that he had seen a memo that indicated that there were no more funds."

–President Ronald Reagan, April 28, 1987
http://www.reagan.utexas.edu/archives/speeches/1987/042887e.htm

since I was not informed, *I did what?*
until we, ourselves, *did what?*

Syntax in NLP, continued

On the other hand, syntax by itself isn't sufficient to provide meaning.

Here's a sentence that's nonsensical but is grammatically correct:

Colorless green ideas sleep furiously.
–Noam Chomsky, 1957

Real language

Real human languages provide many problems for NLP:

- *ambiguity*
- *anaphora*
- *indexicality*
- *vagueness*
- *discourse structure*
- *metonymy*
- *metaphor*
- *noncompositionality*

Lexical Ambiguity

Lexical ambiguity: same syntactic structure, several word meanings

Lexical ambiguity can produce semantic ambiguity:

He wore a light suit.
The fisherman went to the bank
Red Tape Holds Up New Bridge
Police Begin Campaign to Run Down Jaywalkers

Lexical Ambiguity

Syntactic structure can sometimes resolve lexical ambiguity:

The first one won one one-dollar bill.

Lexical Ambiguity

Syntactic structure can sometimes resolve lexical ambiguity:

The first one won one one-dollar bill.

Flies like a flower.

Lexical Ambiguity

Syntactic structure can sometimes resolve lexical ambiguity:

The first one won one one-dollar bill.

Flies like a flower.

John, where Mary had had "had," had had "had had."

Lexical Ambiguity

Syntactic structure can sometimes resolve lexical ambiguity:

The first one won one one-dollar bill.

Flies like a flower.

John, where Mary had had "had," had had "had had."

Dogs dogs dog dog dogs.
I.e., dogs whom other dogs dog, themselves dog other dogs.

Lexical Ambiguity

Syntactic structure can sometimes resolve lexical ambiguity:

The first one won one one-dollar bill.

Flies like a flower.

John, where Mary had had "had," had had "had had."

Dogs dogs dog dog dogs.
I.e., dogs whom other dogs dog, themselves dog other dogs.

Buffalo buffalo buffalo buffalo buffalo.
– William Rapaport (a professor at SUNY Buffalo)

Buffalo buffalo Buffalo buffalo Buffalo buffalo Buffalo buffalo Buffalo buffalo.
– William Rapaport again

Syntactic Ambiguity

Syntactic ambiguity: more than one possible syntactic structure.

This usually also involves lexical ambiguity.

It usually creates semantic ambiguity.

> The chicken is ready to eat.
> They are hunting dogs.
> Squad Helps Dog Bite Victim
> Helicopter Powered By Human Flies

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

For the honeymoon, **they** went to Hawaii

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

For the honeymoon, **they** went to Hawaii

Mary saw a ring through the window and asked John for **it**

Anaphora

Using pronouns to refer back to entities already introduced in the text

After Mary proposed to John, **they** found a preacher and got married.

For the honeymoon, **they** went to Hawaii

Mary saw a ring through the window and asked John for **it**

Mary threw a rock at the window and broke **it**

Anaphora

Using pronouns to refer back to entities already introduced in the text

Referential ambiguity can lead to semantic ambiguity:

Bob waved to Jim in the hallway between class. He smiled.
    Who smiled? Bob or Jim?

Anaphora

Using pronouns to refer back to entities already introduced in the text

Referential ambiguity can lead to semantic ambiguity:

Bob waved to Jim in the hallway between class. He smiled.
    Who smiled? Bob or Jim?

After they finished the exam, the students and lecturers left.
    Who finished the exam? The students? The lecturers? Both?

Anaphora

Using pronouns to refer back to entities already introduced in the text

Referential ambiguity can lead to semantic ambiguity:

Bob waved to Jim in the hallway between class. He smiled.
    Who smiled? Bob or Jim?

After they finished the exam, the students and lecturers left.
    Who finished the exam? The students? The lecturers? Both?

Sue and Lisa gave John and Mark some photos because they liked them.
    Sue and Lise liked the photos?
    John and Mark liked the photos?
    Sue and Lise liked John and Mark?

Indexicality

Indexical sentences refer to utterance situation (place, time, S/H, etc.)

I am over **here**

Why did **you** do **that**?

Metonymy

Using one noun phrase to stand for another

I've read **Shakespeare**

Chrysler announced record profits

The **ham sandwich** on Table 4 wants another beer

Metaphor

"Non-literal" usage of words and phrases, often systematic:

I've tried killing the process but it won't die. Its parent keeps it alive.

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

small moon

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

small moon
large molecule

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

small moon
large molecule
mere child

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

small moon
large molecule
mere child
alleged murderer

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

small moon
large molecule
mere child
alleged murderer
real leather

Noncompositionality

How does a phrase's meaning relate to the meanings of its parts?

baby shoes
basketball shoes
alligator shoes
designer shoes
brake shoes

red book
red pen
red hair
red herring

small moon
large molecule
mere child
alleged murderer
real leather
artificial grass

Part-of-speech tagging

A single English word may be any of several parts of speech.

"Flies like a flower."

- ♢ *Flies*: noun or verb?
- ♢ *like*: preposition, adverb, conjunction, noun, or verb?
- ♢ *a*: article, noun, or preposition?
- ♢ *flower*: noun or verb?

One way to help resolve lexical ambiguity is **part-of-speech tagging** . . .