“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 achieve the speaker’s goals:
- Inform: “There’s a pit in front of you”
- Query: “Can you see the gold?”
- Command: “Pick it up”
- Promise: “I’ll share the gold with you”
- Acknowledge: “OK”

Speech act planning requires knowledge of:
- Situation
- Semantic and syntactic conventions
- Hearer’s goals, knowledge base, and rationality

Stages in communication (informing)

<table>
<thead>
<tr>
<th>Intention</th>
<th>S wants to inform H that P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation</td>
<td>S selects words W to express P in context C</td>
</tr>
<tr>
<td>Synthesis</td>
<td>S utters words W</td>
</tr>
<tr>
<td>Perception</td>
<td>H perceives W’ in context C’</td>
</tr>
<tr>
<td>Analysis</td>
<td>H infers possible meanings P₁,...,Pₙ</td>
</tr>
<tr>
<td>Disambiguation</td>
<td>H infers intended meaning Pᵢ</td>
</tr>
<tr>
<td>Incorporation</td>
<td>H incorporates Pᵢ into KB</td>
</tr>
</tbody>
</table>

How could this go wrong?
- Insincerity (S doesn’t believe P)
- Speech wreck ignition failure
- Ambiguous utterance
- Differing understanding of current context (C ≠ C’)

Grammar

Vervet monkeys, antelopes etc. use isolated symbols for sentences
⇒ restricted set of communicable propositions, no generative capacity (Chomsky (1957): Syntactic Structures)

Grammar specifies the compositional structure of complex messages
  e.g., speech (linear), text (linear), music (two-dimensional)

A formal language is a set of strings of terminal symbols

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

S → NP VP
Article → the | a | an | ...

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

Grammar types

Regular: nonterminal → terminal | nonterminal
  S → aS
  S → λ

Context-free: nonterminal → anything
  S → aSb

Context-sensitive: more nonterminals on right-hand side
  ASB → AAaBB

Recursively enumerable: no constraints

Related to Post systems and Kleene systems of rewrite rules

Natural languages probably context-free, parsable in real time

Wumpus lexicon

<table>
<thead>
<tr>
<th>Noun</th>
<th>stench</th>
<th>breeze</th>
<th>glitter</th>
<th>nothing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wumpus</td>
<td>pit</td>
<td>pits</td>
<td>gold</td>
</tr>
<tr>
<td>Verb</td>
<td>is</td>
<td>see</td>
<td>smell</td>
<td>shoot</td>
</tr>
<tr>
<td></td>
<td>go</td>
<td>grab</td>
<td>carry</td>
<td>kill</td>
</tr>
<tr>
<td>Adjective</td>
<td>right</td>
<td>left</td>
<td>east</td>
<td>back</td>
</tr>
<tr>
<td>Adverb</td>
<td>here</td>
<td>there</td>
<td>nearby</td>
<td>ahead</td>
</tr>
<tr>
<td>Pronoun</td>
<td>me</td>
<td>you</td>
<td>I</td>
<td>it</td>
</tr>
<tr>
<td>Name</td>
<td>John</td>
<td>Mary</td>
<td>Boston</td>
<td>UCB</td>
</tr>
<tr>
<td>Article</td>
<td>the</td>
<td>a</td>
<td>an</td>
<td>...</td>
</tr>
<tr>
<td>Preposition</td>
<td>to</td>
<td>in</td>
<td>on</td>
<td>near</td>
</tr>
<tr>
<td>Conjunction</td>
<td>and</td>
<td>or</td>
<td>but</td>
<td>...</td>
</tr>
<tr>
<td>Digit</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Divided into closed and open classes
Wumpus lexicon

Noun → stench | breeze | glitter | nothing | wumpus | pit | pits | gold | east | ...
Verb → is | see | smell | shoot | feel | stinks | go | grab | carry | kill | turn | ...
Adjective → right | left | east | south | back | smelly | ...
Adverb → here | there | nearby | ahead | right | left | east | south | back | ...
Preposition → to | in | on | near | ...
Digit → 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9

Divided into closed and open classes

Chapter 22 13

Wumpus grammar

S → NP VP | S Conjunction S
  | I + feel a breeze
NP → Pronoun | I
  | Noun | pits
  | Article Noun | the + wumpus
  | Digit Digit | 3 4
  | NP PP | the wumpus + to the east
  | NP RelClause | the wumpus + that is smelly
VP → Verb | stinks
  | VP NP | feel + a breeze
  | VP Adjective | is + smelly
  | VP PP | turn + to the east
  | VP Adverb | go + ahead
PP → Preposition NP | to + the east
RelClause → that VP | that + is smelly

Chapter 22 14

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

Chapter 22 15

Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus

Chapter 22 16

Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus

Chapter 22 17

Parse trees

Exhibit the grammatical structure of a sentence

I shoot the wumpus

Chapter 22 18
Parse trees

Exhibit the grammatical structure of a sentence

Syntax in NLP

Most view syntactic structure as an essential step towards meaning;
“Mary hit John” ≠ “John hit Mary”

“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.”

“Wouldn’t the sentence ‘I want to put a hyphen between the words Fish
and And and And and Chips in my Fish-And-Chips sign’ have been clearer
if quotation marks had been placed before Fish, and between Fish and and,
and and And, and And and, and, and and and, and and and Chips, as well as after Chips?”

Context-free parsing

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

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 ≡ Boolean matrix multiplication (Lee, 2002)
⇒ unlikely to find faster practical algorithms

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 $Y(s_1) \land Z(s_2) \Rightarrow X(Append(s_1, s_2))$

$X \rightarrow word$ becomes $X(["word"])$

$X \rightarrow Y \mid Z$ 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 contd.

Now it’s easy to augment the rules

\[ NP(s_1) \land EatsBreakfast(Ref(s_1)) \land VP(s_2) \Rightarrow NP(Append(s_1, \text{"who"}, s_2)) \]

\[ NP(s_1) \land Number(s_1, n) \land VP(s_2) \land Number(s_2, n) \Rightarrow S(Append(s_1, s_2)) \]

Parsing is reduced to logical inference:

Ask\( (KB, S([\text{"I am a wumpus"}] )) \)

(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\( (KB, S(x; At(\text{Robot}, [1, 1])) ) \)

Real language

Real human languages provide many problems for NLP:

\[ \text{ambiguity} \]
\[ \text{anaphora} \]
\[ \text{indexicality} \]
\[ \text{vagueness} \]
\[ \text{discourse structure} \]
\[ \text{metonymy} \]
\[ \text{metaphor} \]
\[ \text{noncompositionality} \]

Ambiguity

Squad helps dog bite victim

Helicopter powered by human flies

American pushes bottle up Germans

I ate spaghetti with meatballs
Ambiguity

Squad helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans
I ate spaghetti with meatballs
   salad
abandon

Ambiguity can be lexical (polysemy), syntactic, semantic, referential

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
Mary saw a ring through the window and asked John for it
Mary threw a rock at the window and broke it

Indexicality
Indexical sentences refer to utterance situation (place, time, S/H, etc.)
I am over here
Why did you do that?

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

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

basketball shoes
baby shoes
alligator shoes
designer shoes

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair

Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes
red book
red pen
red hair
red herring
small moon
large molecule
mere child
Noncompositionality

basketball shoes
baby 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