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?

To change the actions of other agents
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>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention</td>
<td>S wants to inform H that ( P )</td>
</tr>
<tr>
<td>Generation</td>
<td>S selects words ( W ) to express ( P )</td>
</tr>
<tr>
<td>Synthesis</td>
<td>S utters words ( W )</td>
</tr>
<tr>
<td>Perception</td>
<td>H perceives ( W )</td>
</tr>
<tr>
<td>Analysis</td>
<td>H infers possible meanings ( P_1, \ldots, P_n )</td>
</tr>
<tr>
<td>Disambiguation</td>
<td>H infers intended meaning ( P_i )</td>
</tr>
<tr>
<td>Incorporation</td>
<td>H incorporates ( P_i ) 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 situation

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

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

The grammar is a set of rewrite rules, e.g.,
\[ S \rightarrow NP \ VP \]
\[ Article \rightarrow \text{the} | \text{a} | \text{an} | \ldots \]

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

Stages in communication (informing):

<table>
<thead>
<tr>
<th>Stage</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intention</td>
<td>S wants to inform H that ( P )</td>
</tr>
<tr>
<td>Generation</td>
<td>S selects words ( W ) to express ( P )</td>
</tr>
<tr>
<td>Synthesis</td>
<td>S utters words ( W )</td>
</tr>
<tr>
<td>Perception</td>
<td>H perceives ( W )</td>
</tr>
<tr>
<td>Analysis</td>
<td>H infers possible meanings ( P_1, \ldots, P_n )</td>
</tr>
<tr>
<td>Disambiguation</td>
<td>H infers intended meaning ( P_i )</td>
</tr>
<tr>
<td>Incorporation</td>
<td>H incorporates ( P_i ) 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 situation

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 | back | ... |
- **Pronoun**: \( me | you | I | it | ... |
- **Article**: \( the | a | an | ... |
- **Preposition**: \( to | in | on | near | ... |
- **Conjunction**: \( and | or | but | ... |
- **Digit**: \( 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 

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 | …
- **Pronoun** → me | you | I | it | S/HE | Y’ALL…
- **Name** → John | Mary | Boston | UCB | PAJC | …
- **Article** → the | a | an | …
- **Preposition** → to | in | on | near | …

Divided into closed and open classes

**Wumpus grammar**

\[
S \rightarrow NP \ VP \\
| S \ Conjunction \ S
\]

\[
NP \rightarrow Pronoun
| Noun
| Article \ Noun
| Digit \ Digit
| NP \ PP
| NP \ RelClause
\]

\[
VP \rightarrow Verb
| VP \ NP
| VP \ Adjective
| VP \ PP
| VP \ Adverb
\]

\[
PP \rightarrow \text{Preposition} \ NP
\]

\[
RelClause \rightarrow \text{that} \ VP
\]

**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
**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, 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, Ch. 23) \(O(n^3)\) 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

\[ X \rightarrow YZ \rightarrow Y(s_1) \land Z(s_2) \Rightarrow X(Append(s_1, s_2)) \]

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

\[ X \rightarrow Y \mid Z \rightarrow Y(s) \Rightarrow X(s) \quad Z(s) \Rightarrow X(s) \]

Here, \(X(s)\) means that string \(s\) can be interpreted as an \(X\)
Now it’s easy to augment the rules:

\[
\begin{align*}
\text{NP}(s_1) & \land \text{EatsBreakfast}(\text{Ref}(s_1)) \land \text{VP}(s_2) \\
& \Rightarrow \text{NP}(\text{Append}(s_1, \text{"who"}, s_2)) \\
\text{NP}(s_1) & \land \text{Number}(s_1, n) \land \text{VP}(s_2) \land \text{Number}(s_2, n) \\
& \Rightarrow \text{S}(\text{Append}(s_1, s_2))
\end{align*}
\]

Parsing is reduced to logical inference:

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

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

Generation simply requires a query with uninstantiated variables:

\[
\text{Ask}(\text{KB}, \text{S}(x))
\]

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

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

Real human languages provide many problems for NLP:

- Ambiguity
- Anaphora
- Indexicality
- Vagueness
- Noncompositionality
- Discourse structure
- Metonymy
- Metaphor
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

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
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

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

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.
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
red herring

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