

These notes were shamelessly taken from <http://aima.eecs.berkeley.edu/slides-tex/> with some modifications.

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

Why?

Language is used to affect the actions of other agents.

Speech acts

Speech acts achieve the speaker's goals:

<i>Inform</i>	“There's a pit in front of you”
<i>Query</i>	“Can you see the gold”
<i>Command</i>	“Pick it up”
<i>Promise</i>	“I'll share the gold with you”
<i>Acknowledge</i>	“OK”

Speech act planning requires knowledge of

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

Stages in communication (informing)

<i>Intention</i>	S wants to inform H that P
<i>Generation</i>	S selects words W to express P
<i>Synthesis</i>	S utters words W
<i>Perception</i>	H perceives W'
<i>Analysis</i>	H infers possible meanings P_1, \dots, P_n
<i>Disambiguation</i>	H infers intended meaning P_i
<i>Incorporation</i>	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 situation

Grammar

Grammar specifies the structure of messages.

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.,

$$S \rightarrow NP VP$$

$$Article \rightarrow the \mid a \mid an \mid \dots$$

Here S is the *sentence* symbol, NP and VP are *non-terminals*

Wumpus lexicon

Noun \rightarrow stench \mid breeze \mid glitter \mid nothing
 \mid wumpus \mid pit \mid pits \mid gold \mid east \mid ...

Verb \rightarrow is \mid see \mid smell \mid shoot \mid feel \mid stinks
 \mid go \mid grab \mid carry \mid kill \mid turn \mid ...

Adjective \rightarrow right \mid left \mid east \mid south \mid back \mid smelly \mid ...

Adverb \rightarrow here \mid there \mid nearby \mid ahead
 \mid right \mid left \mid east \mid south \mid back \mid ...

Pronoun \rightarrow me \mid you \mid I \mid it \mid ...

Name \rightarrow John \mid Mary \mid Boston \mid UCB \mid PAJC \mid ...

Article \rightarrow the \mid a \mid an \mid ...

Preposition \rightarrow to \mid in \mid on \mid near \mid ...

Conjunction \rightarrow and \mid or \mid but \mid ...

Digit \rightarrow 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9

Wumpus grammar

S	\rightarrow	$NP VP$	I + feel a breeze
		$S Conjunction S$	I feel a breeze + and + I smell a wumpu
NP	\rightarrow	$Pronoun$	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	\rightarrow	$Verb$	stinks
		$VP NP$	feel + a breeze
		$VP Adjective$	is + smelly
		$VP PP$	turn + to the east
		$VP Adverb$	go + ahead
PP	\rightarrow	$Preposition NP$	to + the east
RelClause	\rightarrow	that VP	that + is smelly

Grammaticality judgements

Formal language L_1 may differ from natural language L_2

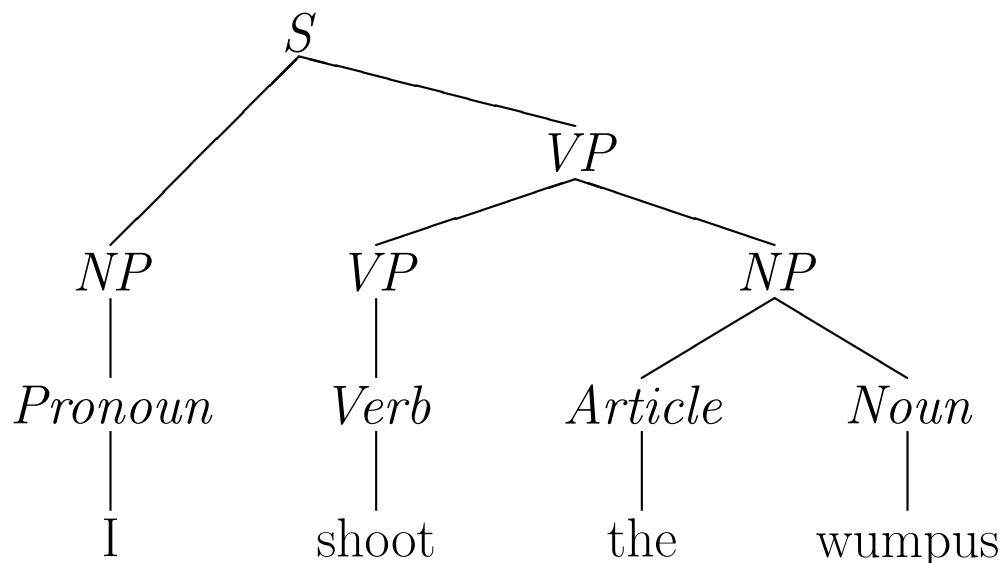
Adjusting L_1 to agree with L_2 is a learning problem!

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

Real grammars are 10–500 pages, insufficient even for “proper” English.

Parsing

A parse tree exhibits the grammatical structure of a sentence.



Syntax in NLP

Most view syntactic structure as an essential step towards meaning;

“Mary hit John” \neq “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.”

Context-free parsing

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

Efficient algorithms (e.g., chart parsing, Ch. 23) are $O(n^3)$ for context-free grammars and run at several thousand words/sec for real grammars.

Context-free parsing \equiv Boolean matrix multiplication (Lee, 2002)
 \rightarrow faster practical algorithms are unlikely.

Logical grammars

BNF notation for grammars makes it difficult:

- to add “side conditions” (number agreement, etc.)
- to connect syntax to semantics

Idea: express grammar rules as logic

$X \rightarrow YZ$ becomes $Y(s_1) \wedge Z(s_2) \rightarrow X(\text{Append}(s_1, s_2))$

$X \rightarrow \text{word}$ becomes $X([\text{“word”}])$

$X \rightarrow Y \mid Z$ becomes $Y(s) \rightarrow X(s)$ and $Z(s) \rightarrow X(s)$

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

Logical grammars contd.

Now it's easier to augment the rules:

$$NP(s_1) \wedge Agent(Ref(s_1)) \wedge VP(s_2) \\ \rightarrow NP(Append(s_1, ["who"], s_2))$$

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

Parsing is reduced to logical inference:

$$ASK(KB, S(["I" "am" "a" "wumpus"]))$$

Generation is a query with uninstantiated variables:

$$ASK(KB, S(x))$$

Extra arguments can be added to return/construct the parse tree and semantics.

Real language

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

a fork

a friend

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*

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:

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

Noncompositionality

basketball shoes	red book	small moon
baby shoes	red pen	large molecule
alligator shoes	red hair	mere child
designer shoes	red herring	alleged murderer
brake shoes		real leather
		artificial grass