Natural Language for Communication

Chapter 23.1-23.3

Outline

◊ Phrase Structure Grammars

◊ Syntactic Analysis (Parsing)

◊ Augmented Grammars and Semantic Interpretation

◊ 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

Need a deeper understanindg of language

Grammar

Address Ch 22 data sparsity through generalization (categories)

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., 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.,

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

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

Grammar types

Regular: *nonterminal* → ***terminal***[*nonterminal*]

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

Context-free: *nonterminal* → *anything*

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

Context-sensitive: more nonterminals on right-hand side

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

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

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

Divided into closed and open classes

Wumpus lexicon

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

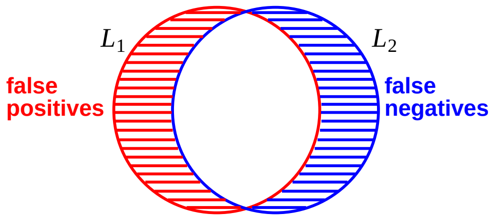
Divided into closed and open classes

Wumpus grammar

S	→	*NP VP*	I + feel a breeze
	|	*S Conjunction S*	I feel a breeze + and + I smell a wumpus
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

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

Probabilistic CFGs

$Noun \rightarrow \textbf{\textit{stench}}(.05) \mid \textbf{\textit{breeze}}(.10) \ldots$

$S \rightarrow NP\ VP$ (.9) I + feel a breeze
$\mid\ S\ Conjunction\ S$ (.1) I feel a breeze + and + I smell a wumpus

- Sum of the probabilities for each category is 1

Parse trees

Exhibit the grammatical structure of a sentence

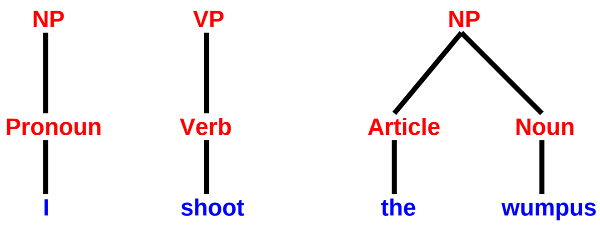
I shoot the wumpus

Parse trees

Exhibit the grammatical structure of a sentence

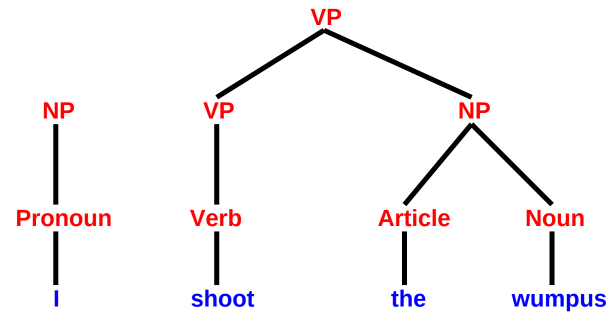
Parse trees

Exhibit the grammatical structure of a sentence



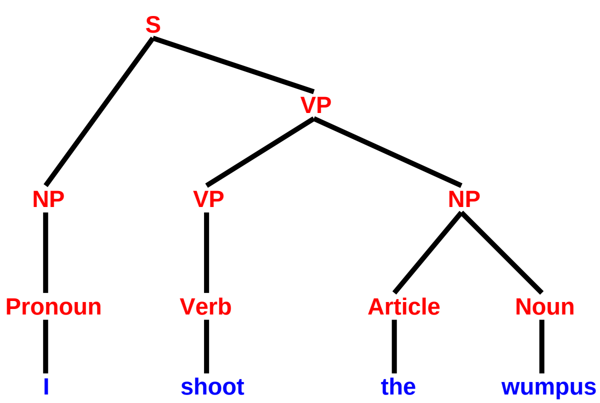
Parse trees

Exhibit the grammatical structure of a sentence



Parse trees

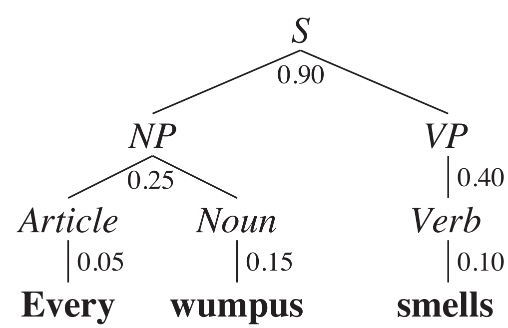
Exhibit the grammatical structure of a sentence



Parse trees - probabilistic

Each interior node is labeled with its probability.

The probability of the tree as a whole is .9 * .25 * .05 * .15 * .4 * .1



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."

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 RHS of a rule with the rule's LHS

Efficient algorithms (e.g., chart parsing (Ch. 23.2) - normal forms, dynamic programming again!)

Learning probabilities for PCFGs - treebanks

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 contd.

Now it's easy to augment the rules

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

Parsing is reduced to logical inference:
$\text{Ask}(KB, S([\text{“}\boldsymbol{I}\text{” “}\boldsymbol{am}\text{” “}\boldsymbol{a}\text{” “}\boldsymbol{wumpus}\text{”}]))$
(Can add extra arguments to return the parse structure, semantics)

Generation simply requires a query with uninstantiated variables:
$\text{Ask}(KB, S(x))$

Augmented grammars contd.

Lexicalized PCFGs

- VP(v) -> Verb(v) NP(n) [P1(v,n)]
- VP(v) -> Verb(v) [P2(v)]
- ...
- Noun(banana) -> banana [pn]

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

Ambiguity

Squad helps dog bite victim
Helicopter powered by human flies

Ambiguity

Squad helps dog bite victim
Helicopter powered by human flies
American pushes bottle up Germans

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

Ambiguity

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

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

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

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.

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**

Metonymy

Using one noun phrase to stand for another

I've read **Shakespeare**

Chrysler announded 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

Noncompositionality

basketball shoes
baby shoes

Noncompositionality

basketball shoes
baby shoes
alligator shoes

Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes

Noncompositionality

basketball shoes
baby shoes
alligator shoes
designer shoes
brake shoes

Noncompositionality

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

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