# Natural Language for Communication

Chapter 23

# Next week's schedule (Nov. 19, 21)

- Tuesday (Nov. 19) – Hao will continue discussion of Ch. 23
- Thursday (Nov. 21) – Hao will lead a tour of Parker's Distributed Intelligence Lab (MK629)
  - 11:10-11:45: Undergrads go to MK629 for tour
  - 11:50-12:25: Grads go to MK629 for tour

# 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):

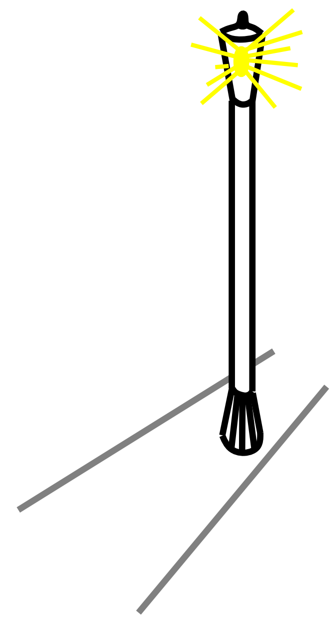
language consists of sentences that are true/false (cf. logic)

"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?



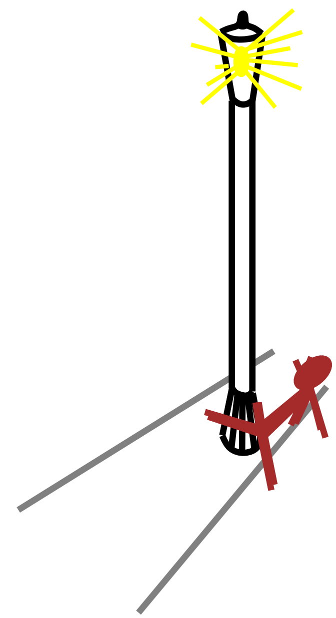
# Communication

"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

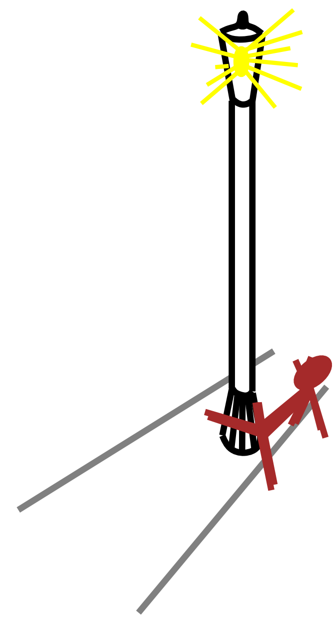
"Classical" view (pre-1953):
language consists of sentences that are true/false (cf. logic)

"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 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)

**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., 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.,

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

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: $nonterminal \rightarrow anything$

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

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

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

Divided into closed and open classes

# Wumpus lexicon

$$
\begin{aligned}
Noun &\rightarrow \textbf{\textit{stench}} \mid \textbf{\textit{breeze}} \mid \textbf{\textit{glitter}} \mid \textbf{\textit{nothing}} \\
&\quad \mid \textbf{\textit{wumpus}} \mid \textbf{\textit{pit}} \mid \textbf{\textit{pits}} \mid \textbf{\textit{gold}} \mid \textbf{\textit{east}} \mid \ldots \\
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}} \\
&\quad \mid \textbf{\textit{go}} \mid \textbf{\textit{grab}} \mid \textbf{\textit{carry}} \mid \textbf{\textit{kill}} \mid \textbf{\textit{turn}} \mid \ldots \\
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 \\
Adverb &\rightarrow \textbf{\textit{here}} \mid \textbf{\textit{there}} \mid \textbf{\textit{nearby}} \mid \textbf{\textit{ahead}} \\
&\quad \mid \textbf{\textit{right}} \mid \textbf{\textit{left}} \mid \textbf{\textit{east}} \mid \textbf{\textit{south}} \mid \textbf{\textit{back}} \mid \ldots \\
Pronoun &\rightarrow \textbf{\textit{me}} \mid \textbf{\textit{you}} \mid \textbf{\textit{I}} \mid \textbf{\textit{it}} \mid \textbf{\textit{S/HE}} \mid \textbf{\textit{Y'ALL}} \ldots \\
Name &\rightarrow \textbf{\textit{John}} \mid \textbf{\textit{Mary}} \mid \textbf{\textit{Boston}} \mid \textbf{\textit{UCB}} \mid \textbf{\textit{PAJC}} \mid \ldots \\
Article &\rightarrow \textbf{\textit{the}} \mid \textbf{\textit{a}} \mid \textbf{\textit{an}} \mid \ldots \\
Preposition &\rightarrow \textbf{\textit{to}} \mid \textbf{\textit{in}} \mid \textbf{\textit{on}} \mid \textbf{\textit{near}} \mid \ldots \\
Conjunction &\rightarrow \textbf{\textit{and}} \mid \textbf{\textit{or}} \mid \textbf{\textit{but}} \mid \ldots \\
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}
$$

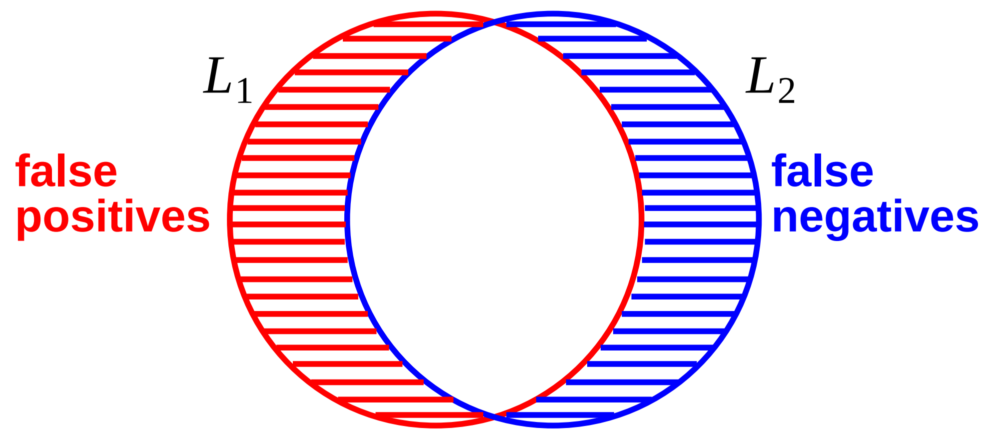
Divided into closed and open classes

# Wumpus grammar

| | | | |
|---|---|---|---|
| *S* | → | *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

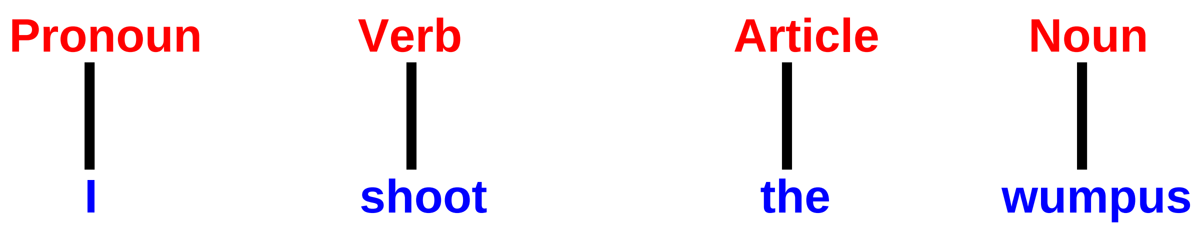
# Parse trees

Exhibit the grammatical structure of a sentence

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

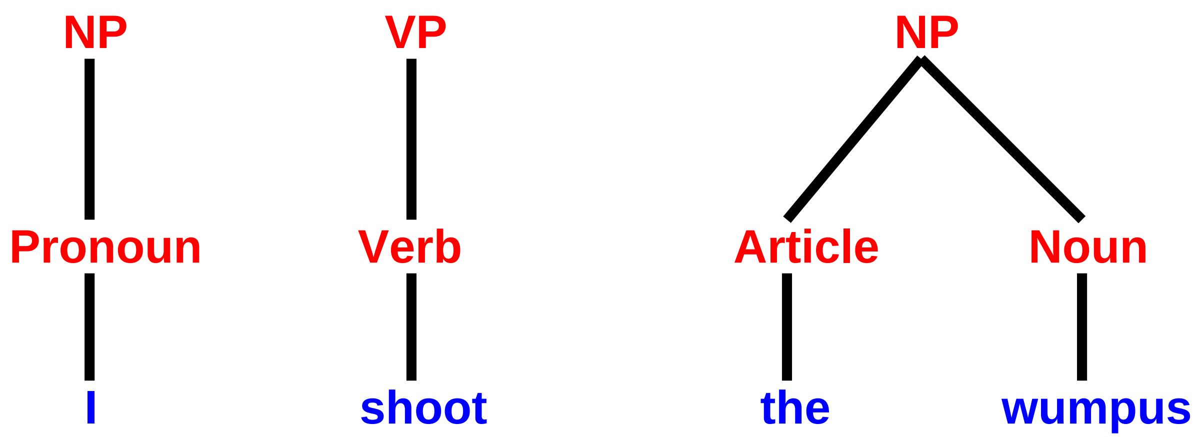
# Parse trees

Exhibit the grammatical structure of a sentence



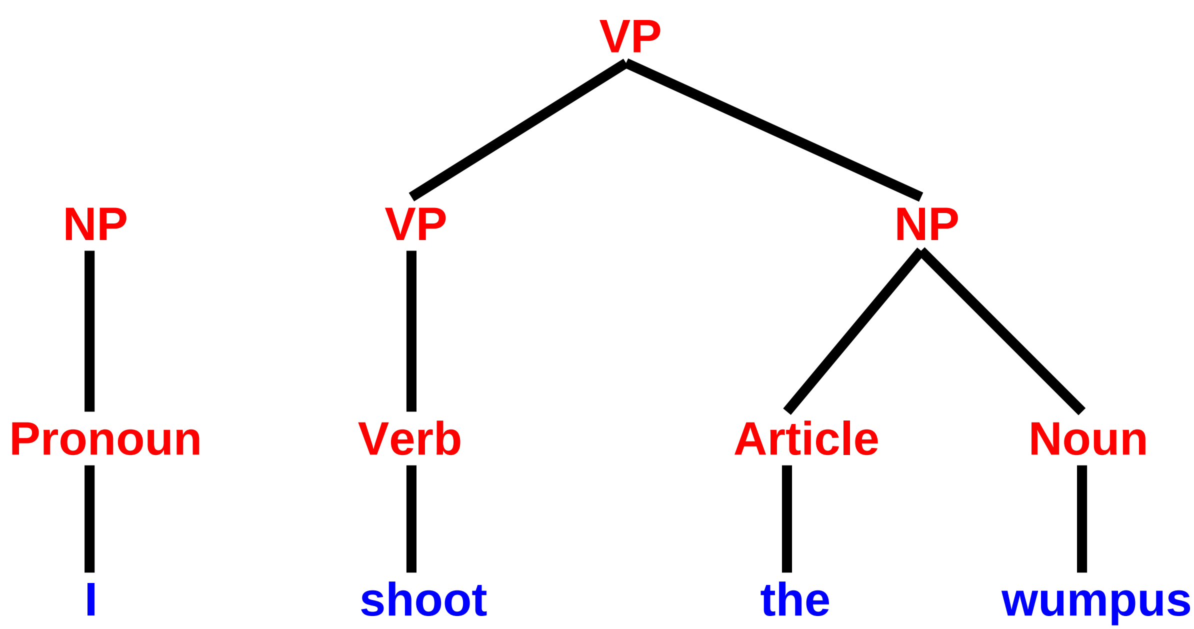
# Parse trees

Exhibit the grammatical structure of a sentence



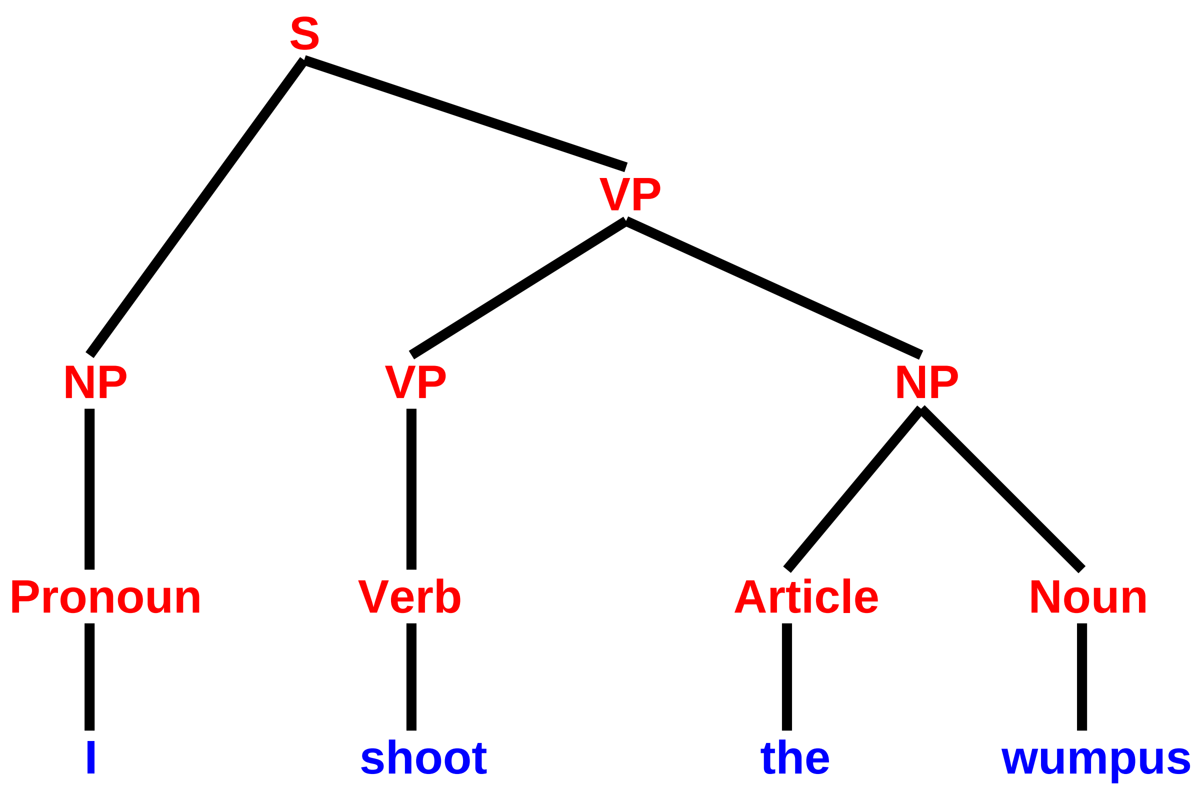
# Parse trees

Exhibit the grammatical structure of a sentence

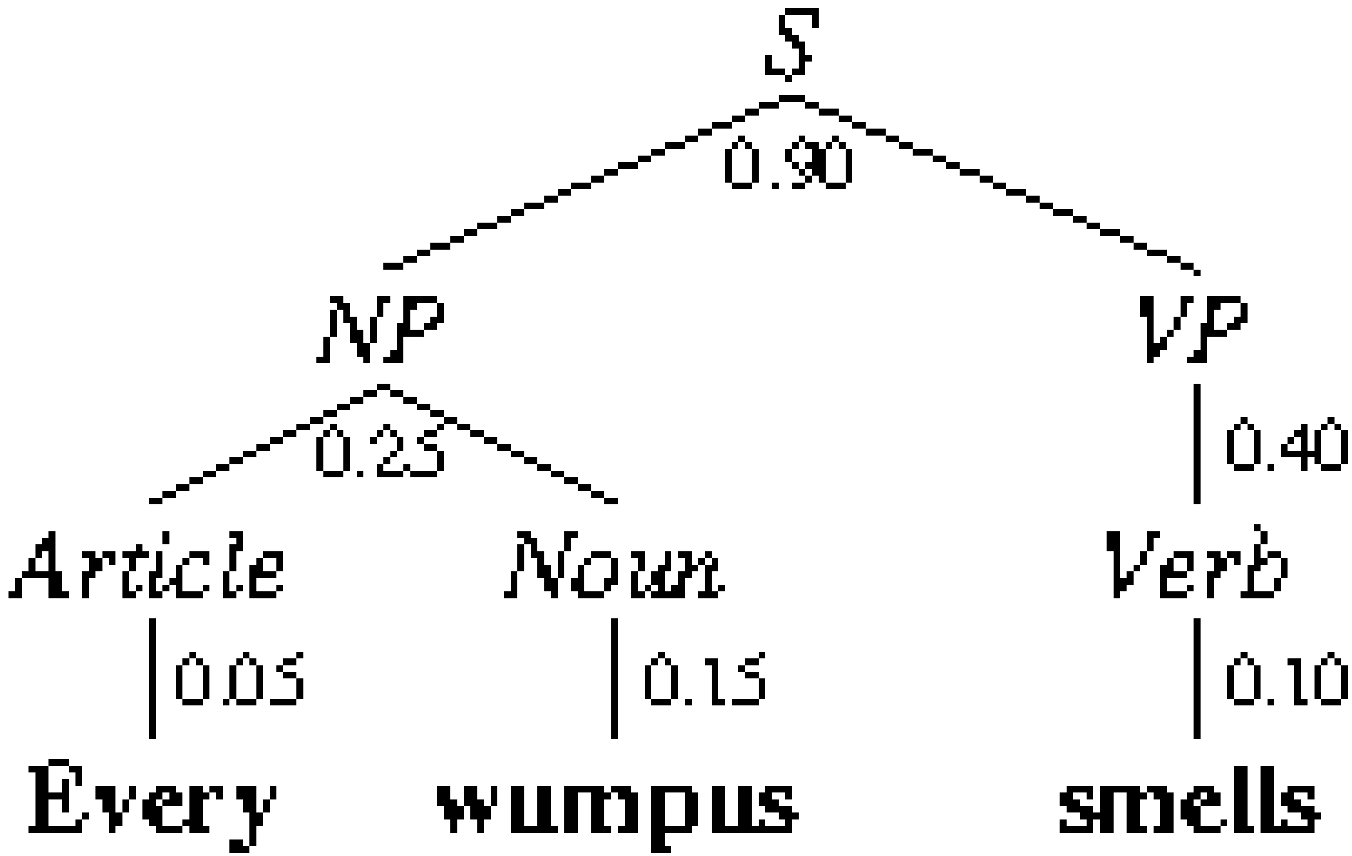


# Parse trees

Exhibit the grammatical structure of a sentence



# Parse tree for "Every wumpus smells"



# 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.”

“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 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 $\equiv$ Boolean matrix multiplication (Lee, 2002)
$\Rightarrow$ unlikely to find faster practical algorithms

# Exercise 1

- An HMM grammar is essentially a standard HMM whose state variable is *N* (nonterminal, with values such as *Adjective*, *Noun*, etc.) and whose evidence variable is W (word, with values such as *is*, *duck*, etc.). The HMM model includes a prior $P(N_o)$, a transition model $P(N_{t+1}|N_t)$, and a sensor model $P(W_t|N_t)$.
- Show that every HMM grammar can be written as a PCFG.

# Exercise 2: Consider the following PCFG for simple verb phrases

0.1: VP → Verb
0.2: VP → Copula Adjective
0.5: VP → Verb the Noun
0.2: VP → VP Adverb
0.5: Verb → is
0.5: Verb → shoots
0.8: Copula → is
0.2: Copula → seems
0.5: Adjective → **unwell**
0.5: Adjective → **well**
0.5: Adverb → **well**
0.5: Adverb → **badly**
0.6: Noun → **duck**
0.4: Noun → **well**

What is the probability of generating "is well well"?

# Exercise 2: Consider the following PCFG for simple verb phrases

0.1: VP → Verb
0.2: VP → Copula Adjective
0.5: VP → Verb the Noun
0.2: VP → VP Adverb
0.5: Verb → is
0.5: Verb → shoots
0.8: Copula → is
0.2: Copula → seems
0.5: Adjective → **unwell**
0.5: Adjective → **well**
0.5: Adverb → **well**
0.5: Adverb → **badly**
0.6: Noun → **duck**
0.4: Noun → **well**

Which of the following have nonzero probability as a VP?

i. shoots the duck well well well
ii. seems the well well
iii. shoots the unwell well badly