

# Comunicação



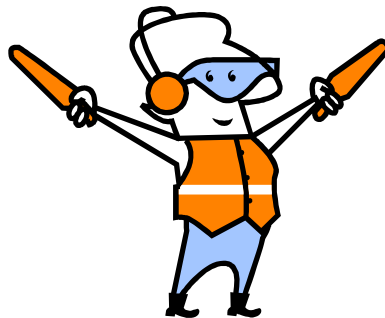
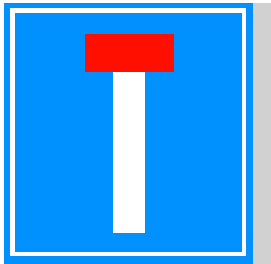
## Capítulo 22

# Sumário

- Comunicação
- Gramáticas
- Análise sintáctica (*parsing*)

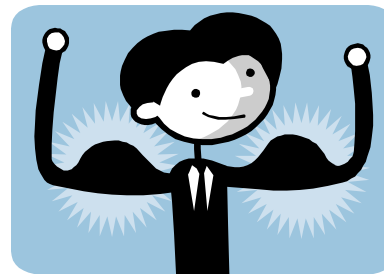
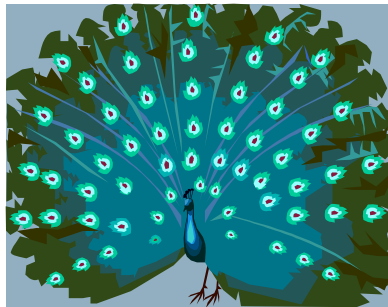
# Comunicação

- Comunicar consiste em trocar informação através da **produção e percepção de sinais**, pertencentes a um sistema partilhado de sinais convencionais



# Comunicação

- Grande parte dos animais usa sinais para representar mensagens importantes

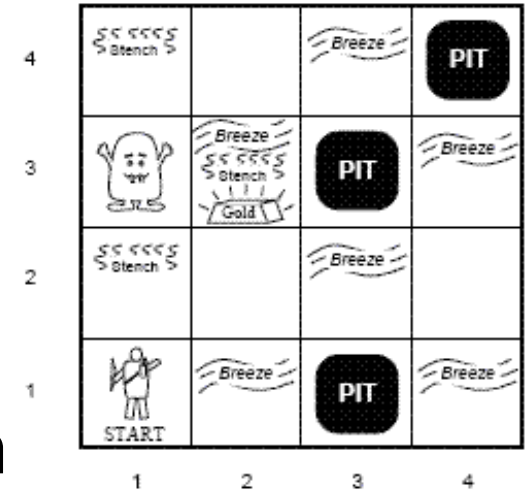


# Comunicação

- Num mundo **parcialmente observável**, a comunicação pode ajudar os agentes, pois podem aprender informação que é observada ou inferida por outros agentes

# Comunicação

- Por exemplo, no mundo dos Wumpus um grupo de agentes exploradores pode ganhar (como grupo e como indivíduos) se forem capazes de:
  - Inquirir outros agentes sobre particularidades do mundo
    - Encontraram wumpus algures?
  - Informar sobre o estado do mundo
    - Cheira bem na posição 3 4
  - Pedir a outros agentes para executar acções
    - Ajudas-me a carregar o ouro?
  - Etc.



# Comunicação

- Um agente pode produzir linguagem: a isto chamamos **actos de discurso**
  - Enviar um mail, usar linguagem gestual, são tudo actos de discurso

# Fundamentos da linguagem

- Linguagem formal:
  - Conjunto de palavras (concatenação de símbolos terminais), eventualmente infinito
  - Linguagens formais têm definições matemáticas rigorosas
    - e.g.  $P \wedge Q$  em LP é uma frase;  $PQ \wedge$  não é
- Linguagem natural:
  - Português, Inglês, Chinês, ...
  - Não têm definições rigorosas
  - Objectivo deste capítulo: estudar linguagem natural de um modo formal



# Fundamentos da Linguagem

- Sintaxe / Gramática
  - Conjunto de regras de uma linguagem
    - e.g. se X e Y são expressões válidas, então  $X+Y$  é uma expressão válida
    - e.g. *Cuidado com o degrau* é uma expressão sintacticamente válida em Português. *Degrau com cuidado o* não é uma expressão válida.

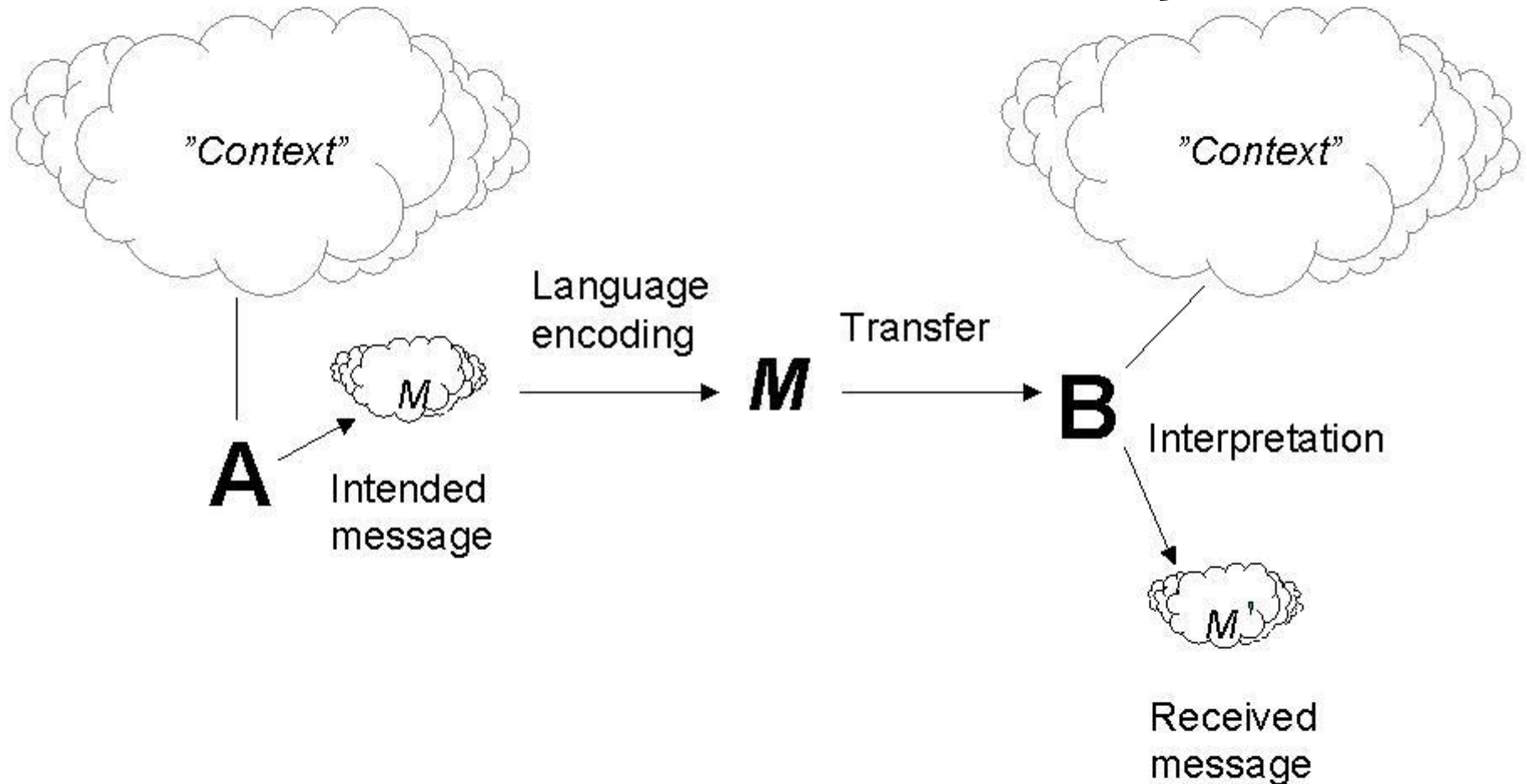
# Fundamentos da Linguagem

- Semântica
  - Significado da linguagem
    - e.g.  $X+Y$  denota a soma de  $X$  com  $Y$
    - e.g. O que quer dizer *Cuidado com o degrau*? Quer dizer (interpretação literal) que devemos ter cuidado com um degrau. Mas podem existir outras interpretações...

# Fundamentos da Linguagem

- Pragmática
  - Significado em função do contexto: nova interpretação
    - O que quer dizer *Cuidado com o degraú* se soubermos que *degraú* é o nome de um cão?

# Modelo de Comunicação



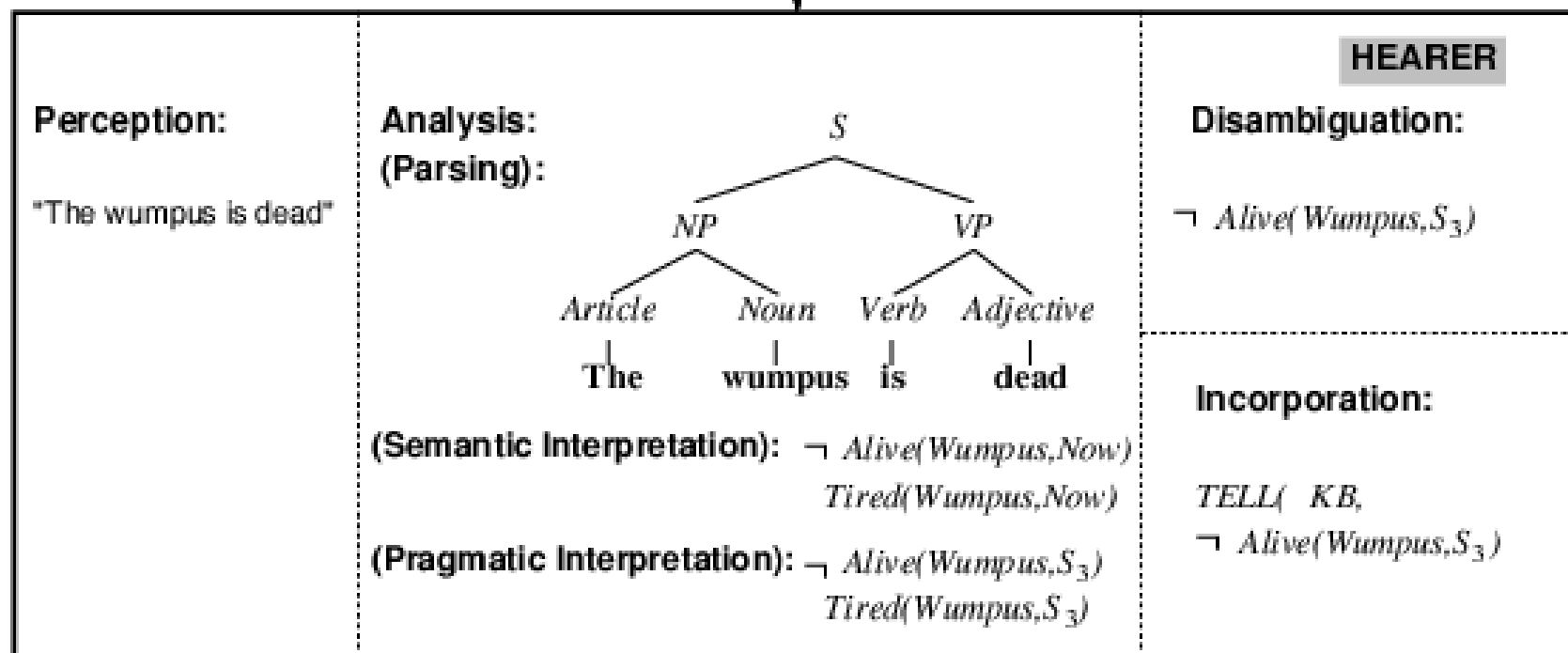
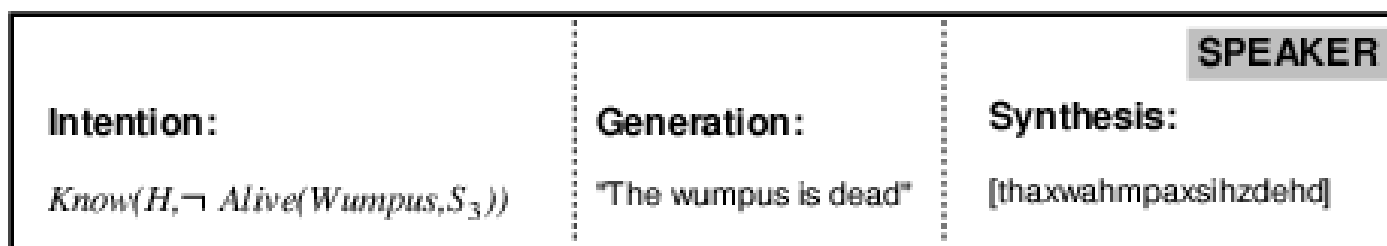
# Etapas na Comunicação

- Intenção:
  - S (*speaker*) quer informar H (*hearer*) sobre P (*proposition*)
    - A Maria quer pedir ao Pedro que abra a porta à sua amiga Berta
- Geração:
  - S selecciona palavras W (*words*) para exprimir P no contexto C
    - A Maria escolhe as palavras a usar para fazer o seu pedido
- Síntese:
  - S emite palavras W
    - A Maria diz: “Pedro, por favor, abre a porta à Berta”
    - ....

# Etapas na Comunicação

- Percepção:
  - H percebe  $W'$  no contexto C
    - Pedro percebe: Abre a porta aberta
- Análise:
  - H infere possíveis significados  $P_1, \dots, P_n$ 
    - Pedro fica a pensar o que a Maria quer dizer: abrir a porta aberta ou... Abrir a porta à Berta.
  - Esta fase inclui interpretação sintáctica (parsing), interpretação semântica e interpretação pragmática*
- Desambiguação:
  - H infere significado pretendido  $P_i$ 
    - A porta está fechada e estão à espera da Berta, pelo que o Pedro percebe que a Maria lhe pediu para abrir a porta à Berta
- Incorporação:
  - H incorpora  $P_i$  na BC (Base de Conhecimento)
    - Pedro compreende que lhe foi pedido para abrir a porta à Berta

# Comunicação: exemplo



# Sumário

- Comunicação
- Gramáticas
- Análise sintáctica (Parsing)



# Gramática Formal

- $G = (N, T, S_0, P)$ 
  - N é o conjunto de símbolos não terminais
  - T é o conjunto de símbolos terminais
  - $S_0$  é símbolo inicial
  - P é um conjunto de regras de produção

# Gramática Formal

- A maior parte dos formalismos gramaticais são baseados na ideia de que certas sequências de palavras formam sequências com diferentes categorias (sintagmas, *phrases*) que combinados formam frases (S, *Sentence*)
  - Considerando o formato BNF (Bachus-Naur Form) uma regra de uma gramática poderá ter a forma:
    - $S \rightarrow NP VP$ 
      - » S é o símbolo de frase
      - » S, NP e VP são símbolos não terminais
      - » NP, *Noun Phrase*
      - » VP, *Verb Phrase*
    - $NP \rightarrow \text{Article Name}$
    - $VP \rightarrow \text{Verb}$

# Construção de Gramáticas

- Léxico: palavras permitidas na linguagem
  - Agrupadas em classes (ou categorias): pronomes, artigos, nomes, verbos, etc.
    - Classes abertas:
      - Não é possível listar todos os elementos ou novos elementos são adicionados frequentemente
        - » Ex: kunami
      - Nomes, verbos, adjetivos, advérbios
    - Classes fechadas:
      - Os seus elementos podem ser enumerados; evolução a longo prazo
      - Pronomes, artigos, proposições, conjunções

# Léxico do Wumpus (linguagem $\varepsilon_0$ )

*Noun*  $\rightarrow$  *stench* | *breeze* | *glitter* | *nothing*  
| *wumpus* | *pit* | *pits* | *gold* | *east* | ...

*Verb*  $\rightarrow$  *is* | *see* | *smell* | *shoot* | *feel* | *stinks*  
| *go* | *grab* | *carry* | *kill* | *turn* | ...

*Adjective*  $\rightarrow$  *right* | *left* | *east* | *south* | *back* | *smelly* | ...

*Adverb*  $\rightarrow$  *here* | *there* | *nearby* | *ahead*  
| *right* | *left* | *east* | *south* | *back* | ...

*Pronoun*  $\rightarrow$  *me* | *you* | *I* | *it* | *S/HE* | *Y'ALL*...

*Name*  $\rightarrow$  *John* | *Mary* | *Boston* | *UCB* | *PAJC* | ...

*Article*  $\rightarrow$  *the* | *a* | *an* | ...

*Preposition*  $\rightarrow$  *to* | *in* | *on* | *near* | ...

*Conjunction*  $\rightarrow$  *and* | *or* | *but* | ...

*Digit*  $\rightarrow$  *0* | *1* | *2* | *3* | *4* | *5* | *6* | *7* | *8* | *9*

# Gramática do Wumpus

$S \rightarrow NP VP$	I + feel a breeze
$S Conjunction S$	I feel a breeze + and + I smell a wumpus
$NP \rightarrow Pronoun$	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
$RelClause \rightarrow \textit{that} VP$	that + is smelly

# Classificação de gramáticas

- Gramáticas são classificadas em função da capacidade gerativa (linguagens que podem representar)
- Supondo
  - $\alpha, \beta, \delta$  sequência de símbolos terminais e não terminais
  - $A, B$  símbolos não terminais
  - $a, b$  símbolos terminais
  - As gramáticas **regulares** têm regras da forma:
    - $A \rightarrow B a \mid a B \mid \epsilon$ 
      - ex:  $S \rightarrow aS \mid a$
  - As gramáticas **livres de contexto** têm regras da forma:
    - $A \rightarrow \alpha$ 
      - ex:  $S \rightarrow aSb \mid aS \mid Sb \mid a$
  - As gramáticas **sensíveis ao contexto** têm regras da forma:
    - $\alpha A \beta \rightarrow \alpha \delta \beta$
  - As gramáticas **recursivamente enumeráveis** não têm restrições nas regras
- Foi demonstrado que quase todas as línguas naturais podem ser caracterizadas por **gramáticas livres de contexto**

# Avaliação de Gramáticas

- Problemas
  - Supergeração: geração de frases incorrectas
    - *Me go Boston*
  - Subgeração: não é possível gerar frases que estão correctas
    - *I think the wumpus is smelly*

# Sumário

- Comunicação
- Gramática
- Análise sintáctica (Parsing)



# Análise sintáctica: parsing

- Processo que gera uma árvore de parsing para os dados de entrada
  - PARSE(“I shoot the wumpus”,  $\varepsilon_0$ , S)
    - $\varepsilon_0$  é o léxico
    - S é o símbolo inicial
- Parsing (básico)
  - Top-down
    - Começar com S como raiz da árvore; gerar árvore com palavras como folhas
  - Bottom-up
    - Começar com palavras como folhas

# Parsing top-down

- Estado inicial
  - [S: ?]
- Função sucessores
  - Selecciona o nó aberto mais à esquerda na árvore
  - Procura regras cujo símbolo do lado esquerdo corresponda a esse nó
  - Por cada regra cria um sucessor onde ? é substituído pelo lado direito da regra correspondente
- Teste objectivo
  - Folhas correspondem exactamente aos dados de entrada
  - Nenhuma palavra dos dados de entrada está em aberto ou é desconhecida

# Parsing top-down

- Exemplo
  - Considerando a gramática do Wumpus, temos
    - $S \rightarrow NP VP \mid S \text{ Conjunction } S$
  - Para qualquer input, o primeiro passo de um algoritmo top-down é:
    - $[S: [NP: ?][VP: ?]]$
    - $[S: [S: ?][Conjunction: ?][S: ?]]$
  - Nota:
    - $[A: ?]$  representa um símbolo não terminal ainda em aberto

# Parsing bottom-up

- Estado inicial:
  - Lista de palavras nos dados de entrada
- Função sucessores
  - Procura emparelhar lado direito de uma regra com cada posição na lista
- Teste objectivo:
  - Estado com uma única árvore com S na raiz

# Parsing bottom-up

- Exemplo:
  - *I shoot the wumpus*
    - *[I,shoot,the,wumpus]*
  - Dado que na gramática temos
    - *Pronoun* → *I*
  - Cria-se a entrada
    - *[[Pronoun: I],shoot,the,wumpus]*
    - ...
    - *[S:[NP:[Pronoun: I]][VP:[VP:[Verb:shoot]][NP:[Article:the][Noun:wumpus]]]]*

# Árvore de parsing

- Após uma análise **top-down** ou **bottom-up**, se a frase em análise pertencer à linguagem aceita pela gramática em causa, é criada pelo menos uma **árvore de análise/parsing**.

# Árvore de Parsing

(construída por um algoritmo bottom-up)

I

shoot

the

wumpus

# Árvore de Parsing

(construída por um algoritmo bottom-up)

Pronoun

I

Verb

shoot

Article

the

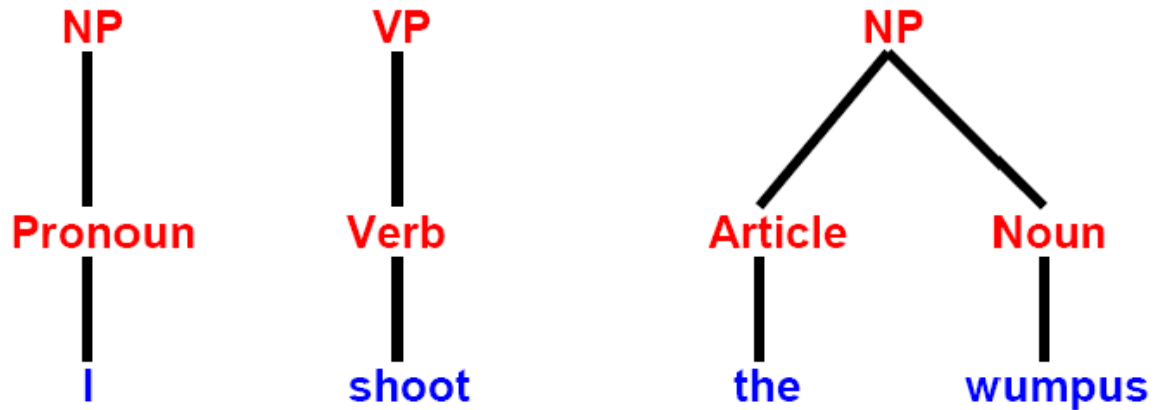
Noun

wumpus



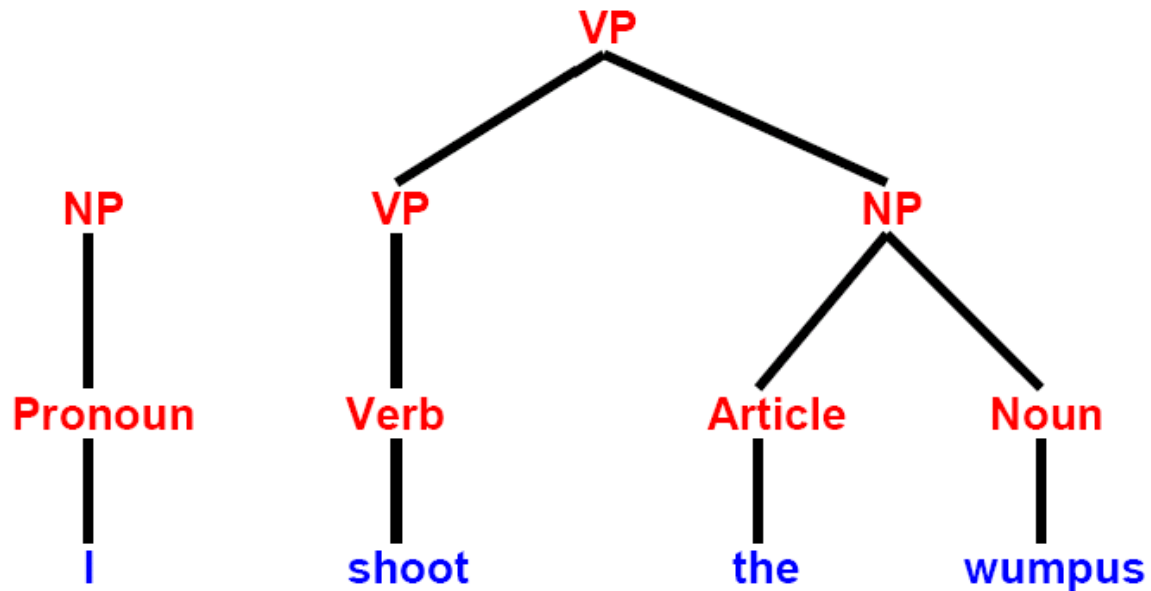
# Árvore de Parsing

(construída por um algoritmo bottom-up)



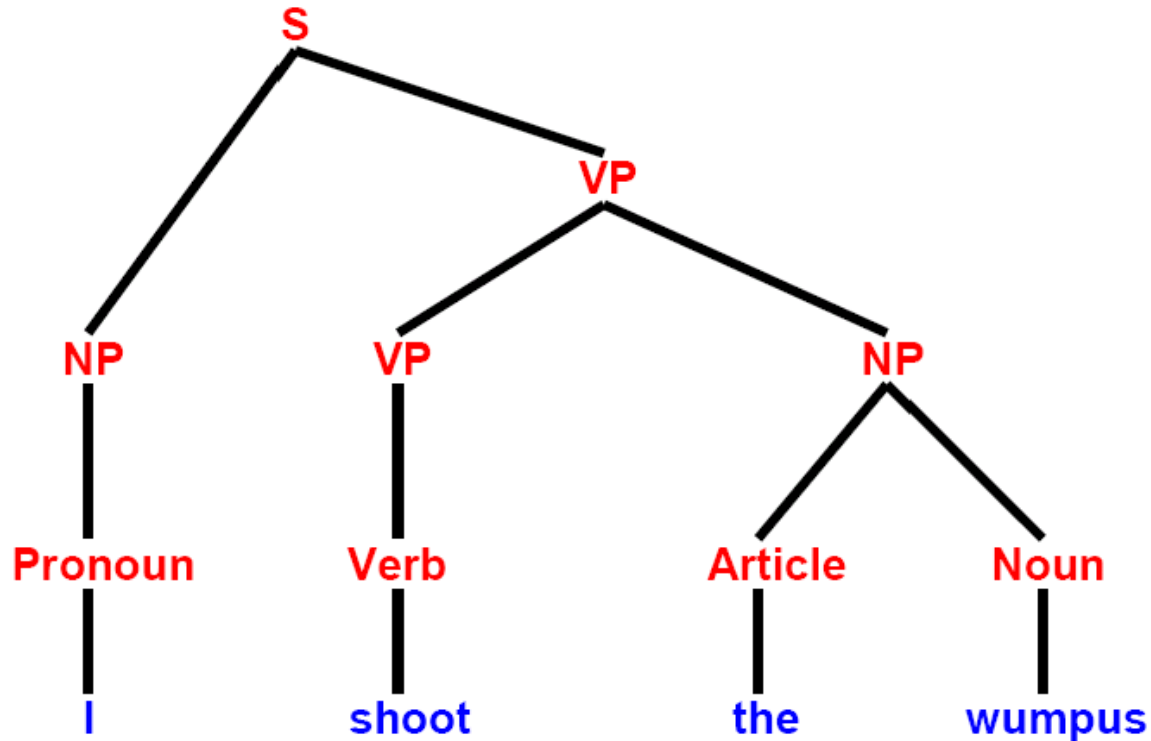
# Árvore de Parsing

(construída por um algoritmo bottom-up)



# Árvore de Parsing

(construída por um algoritmo bottom-up)



# Análise sintáctica

- Quer a análise bottom-up quer a top-down têm os seus problemas:
  - A bottom-up pode estar a construir sub-árvores que não levam ao símbolo inicial
  - A top-down pode estar a construir sub-árvores que não levam aos símbolos terminais
    - O parser canto esquerdo sempre resolve este problema
  - Em qualquer dos casos podem haver **redundâncias** nos cálculos, pois sub-árvores que já foram construídas uma vez, voltam a sê-lo

# Extensão do formalismo gramatical

- Considere-se a seguinte gramática para um pequeno fragmento do Português:
  - Frase  $\rightarrow$  SN SV // SN sintagma nominal, SV verbal
  - SN  $\rightarrow$  nome | det nome // det determinante
  - SV  $\rightarrow$  vi | vt SN // vi verbos intransitivos, vt transitivos
  - nome  $\rightarrow$  sumo | Maria
  - det  $\rightarrow$  a
  - vt  $\rightarrow$  bebe
  - vi  $\rightarrow$  fugiu
  - vi  $\rightarrow$  fugiram

# Extensão do formalismo gramatical

- Fazem parte desta gramática frases como:
  - A Maria bebe sumo
  - A Maria fugiu
- Mas também:
  - A Maria bebe a sumo
  - A Maria fugiram
- Casos como **A Maria bebe a sumo** e **A Maria fugiram** não são correctos porque:
  - No primeiro caso não há concordância em **género** entre o determinante **a** (feminino) e o nome comum **sumo** (masculino)
  - No segundo caso não há concordância em **número** entre o sujeito **A Maria** (singular) e o predicado **fugiram** (plural).

# Extensão do formalismo gramatical

- Uma maneira de resolver este problema é:
  - Adicionar novas categorias:
    - Por exemplo, **nome** passa a **nome-fs**, **nome-ms**, **nome-fp**, **nome-mp**
      - f/m para feminino/masculino, s/p para singular/plural
  - E reescrever as regras:
    - Por exemplo, **SN** → **nome | det nome** passa a
      - **SN-ms** → **nome-ms | det-ms nome-ms**
      - **SN-mp** → **nome-mp | det-mp nome-mp**
      - Etc.
- Só que esta abordagem multiplica o número de regras e símbolos e torna a gramática muito complicada

# Extensão do formalismo gramatical

- Uma maneira alternativa consiste em estender o formalismo gramatical (**augmented grammars**):
  - Definite Clause Grammars (DCGs)
  - Por exemplo,
    - $SN \rightarrow \text{nome} \mid \text{det nome}$
  - passa simplesmente a
    - $SN(NUM, GEN) \rightarrow \text{nome}(NUM, GEN) \mid \text{det}(NUM, GEN) \text{ nome}(NUM, GEN)$
  - Indicando que:
    - O nome e o determinante têm de concordar em género e número
    - O SN tem NUM como número e GEN como género.



# Conclusões

- A área de comunicação (processamento de língua natural) é uma importante área do ramo de IA.
- Sintaxe, semântica, pragmática são alguns dos ingredientes em jogo.
- A ambiguidade da língua natural faz com que formalismos/técnicas usadas para línguas formais não tenham grande sucesso.
- São inúmeras as aplicações nesta área:
  - Sistemas de pergunta/resposta
  - Sistema de diálogo
  - Sistemas de tradução automática
  - Sumarização
  - ...

- Slides adaptados a partir de:
  - [www.inf.ed.ac.uk/teaching/courses/aipp/lecture\\_slides/11\\_PS\\_DCGs.ppt](http://www.inf.ed.ac.uk/teaching/courses/aipp/lecture_slides/11_PS_DCGs.ppt)
  - [www.inf.ed.ac.uk/teaching/courses/aipp/lecture\\_slides/10\\_DCG.ppt](http://www.inf.ed.ac.uk/teaching/courses/aipp/lecture_slides/10_DCG.ppt)
  - <http://www.cse.nd.edu/courses/cse571/www/presentations/chap22.pdf>