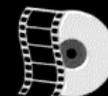


PROCESSAMENTO DE LINGUAGEM NATURAL

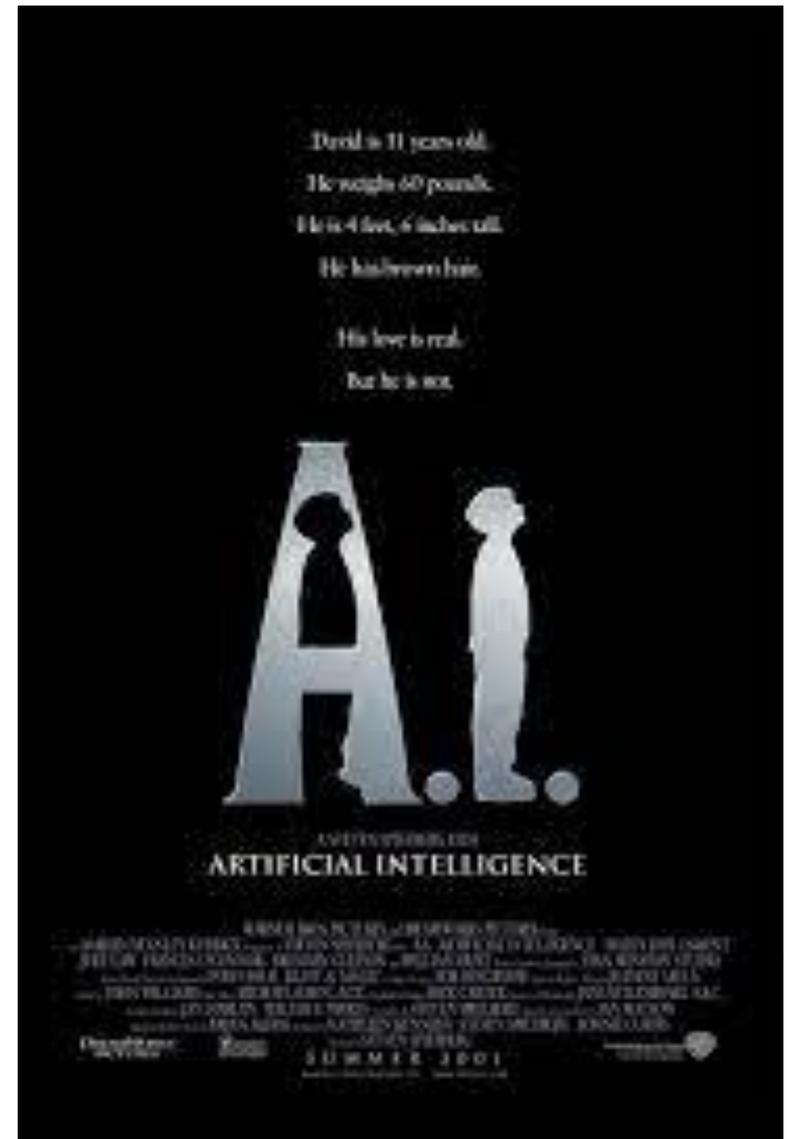


DAVID



A.I.: INTELIGÊNCIA ARTIFICIAL

- Filme de 2001
 - Projeto de Stanley Kubrick, que faleceu antes da realização do filme
 - Dirigido por Steven Spielberg
- Dividiu opiniões



A TRAGÉDIA E A BUSCA



SIGNIFICADO DE SER HUMANO

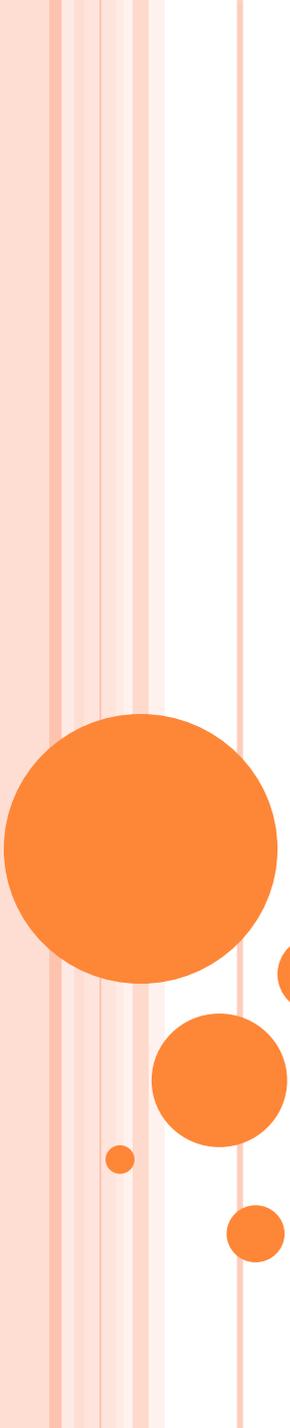


OS RUMOS DA IA

- A realização humana de criar máquinas inteligentes
 - Humanos como “seres divinos”
- Mais do que “ler e escrever”, máquinas capazes de amar e que sofrem com a realidade do mundo
 - A questão existencial: o que significa ser real?
- A inversão dos papéis: após o fim da humanidade, as máquinas recriam os humanos
 - Máquinas se tornam “seres divinos”
- Máquinas buscando entender o significado da vida humana

O AVANÇO DO PLN

- Semântica em todos os seus níveis
 - Significado
 - Ligação com o “mundo real”
 - Sentimentos
- E os desafios existenciais da IA e do PLN
 - Máquinas sencientes?
 - Máquinas conscientes?
 - Queremos realmente isso?



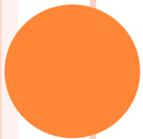
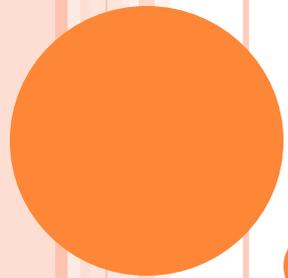
SINTAXE

SCC5908 Introdução ao Processamento de Língua Natural

SCC0633 Processamento de Linguagem Natural

RELEMBRANDO

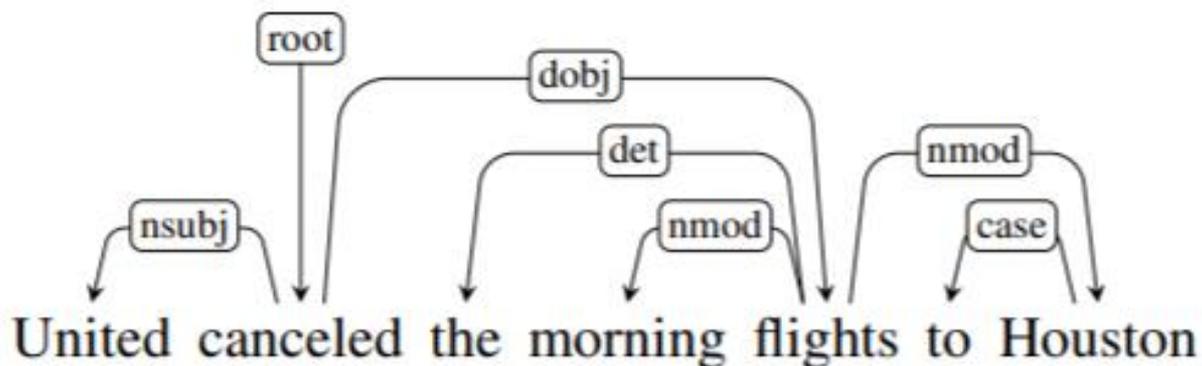
- Importância da sintaxe
- Análise de constituintes e de dependências
- Gramática e treebanks
- Parsing completo e parcial
- Método CKY
- Probabilidades



PARSING DE DEPENDÊNCIA

UMA NOVA ÓTICA

- Busca por pares de elementos e suas relações
- Tendência cada vez mais forte em PLN



UNIVERSAL DEPENDENCIES

- Conjunto “universal” de relações de dependências
 - Atualmente em sua versão 2.8

Clausal Argument Relations	Description
NSUBJ	Nominal subject
DOBJ	Direct object
IOBJ	Indirect object
CCOMP	Clausal complement
XCOMP	Open clausal complement
Nominal Modifier Relations	Description
NMOD	Nominal modifier
AMOD	Adjectival modifier
NUMMOD	Numeric modifier
APPOS	Appositional modifier
DET	Determiner
CASE	Prepositions, postpositions and other case markers
Other Notable Relations	Description
CONJ	Conjunct
CC	Coordinating conjunction

DEFINIÇÕES

- Pares *head*/cabeça-dependente

Relation	Examples with <i>head</i> and dependent
NSUBJ	United <i>canceled</i> the flight.
DOBJ	United <i>diverted</i> the flight to Reno. We <i>booked</i> her the first flight to Miami.
IOBJ	We <i>booked</i> her the flight to Miami.
NMOD	We took the morning <i>flight</i> .
AMOD	Book the cheapest <i>flight</i> .
NUMMOD	Before the storm JetBlue canceled 1000 <i>flights</i> .
APPOS	<i>United</i> , a unit of UAL, matched the fares.
DET	The <i>flight</i> was canceled. Which <i>flight</i> was delayed?
CONJ	We <i>flew</i> to Denver and drove to Steamboat.
CC	We flew to Denver and <i>drove</i> to Steamboat.
CASE	Book the flight through <i>Houston</i> .

UDPIPE

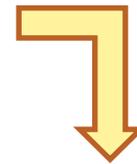
- Um dos parsers mais populares com base em UD, para diversas línguas
- Para testar:
<https://lindat.mff.cuni.cz/services/udpipe/>
- Straka (2018): artigo para os interessados
 - É interessante ver o esforço em larga escala de demonstrar o bom desempenho e a “universalidade” do modelo

REPRESENTAÇÃO GRAMATICAL

- Também é possível
- Usualmente, tem-se uma quádrupla (R, T, C, F)
 - R: regras de dependência
 - T: símbolos terminais
 - C: categorias não terminais
 - F: funções de associação entre T e C

- Exemplo

- $R = \{*(V), V(N,*,N), N(Det,*), N(*), Det(*)\}$
- $T = \{\text{loves, woman, John, a}\}$
- $C = \{V, N, Det\}$
- $F(\text{loves}) = V$
- $F(\text{woman}) = N$
- $F(\text{John}) = N$
- $F(\text{a}) = Det$



Notação

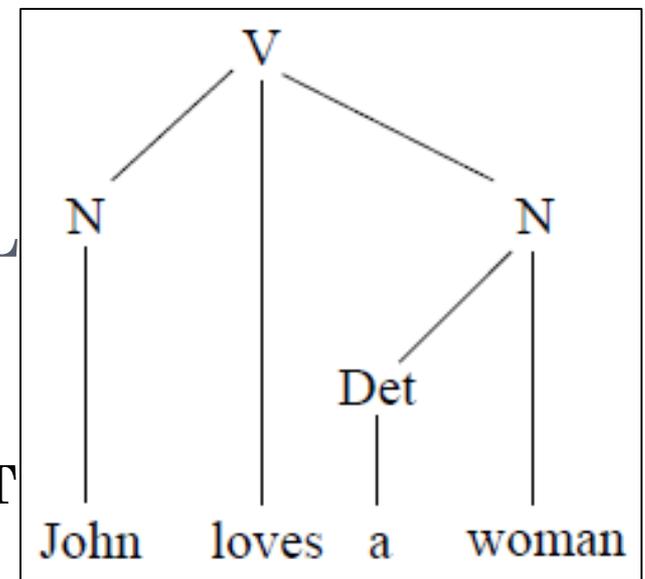
$x(p_1, \dots, *, \dots, p_k)$: $p_1 \dots p_k$ são dependentes de x

$x(*)$: x é folha

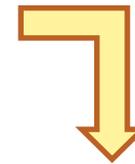
$*(x)$: x é raiz da árvore

REPRESENTAÇÃO GRAMATICAL

- Também é possível
- Usualmente, tem-se uma quádrupla (R, T, C, F)
 - R: regras de dependência
 - T: símbolos terminais
 - C: categorias não terminais
 - F: funções de associação entre T e C



- Exemplo
 - $R = \{*(V), V(N,*,N), N(Det,*), N(*), Det(*)\}$
 - $T = \{\text{loves, woman, John, a}\}$
 - $C = \{V, N, Det\}$
 - $F(\text{loves}) = V$
 - $F(\text{woman}) = N$
 - $F(\text{John}) = N$
 - $F(a) = Det$



Notação

$x(p_1, \dots, *, \dots, p_k)$: $p_1 \dots p_k$ são dependentes de x

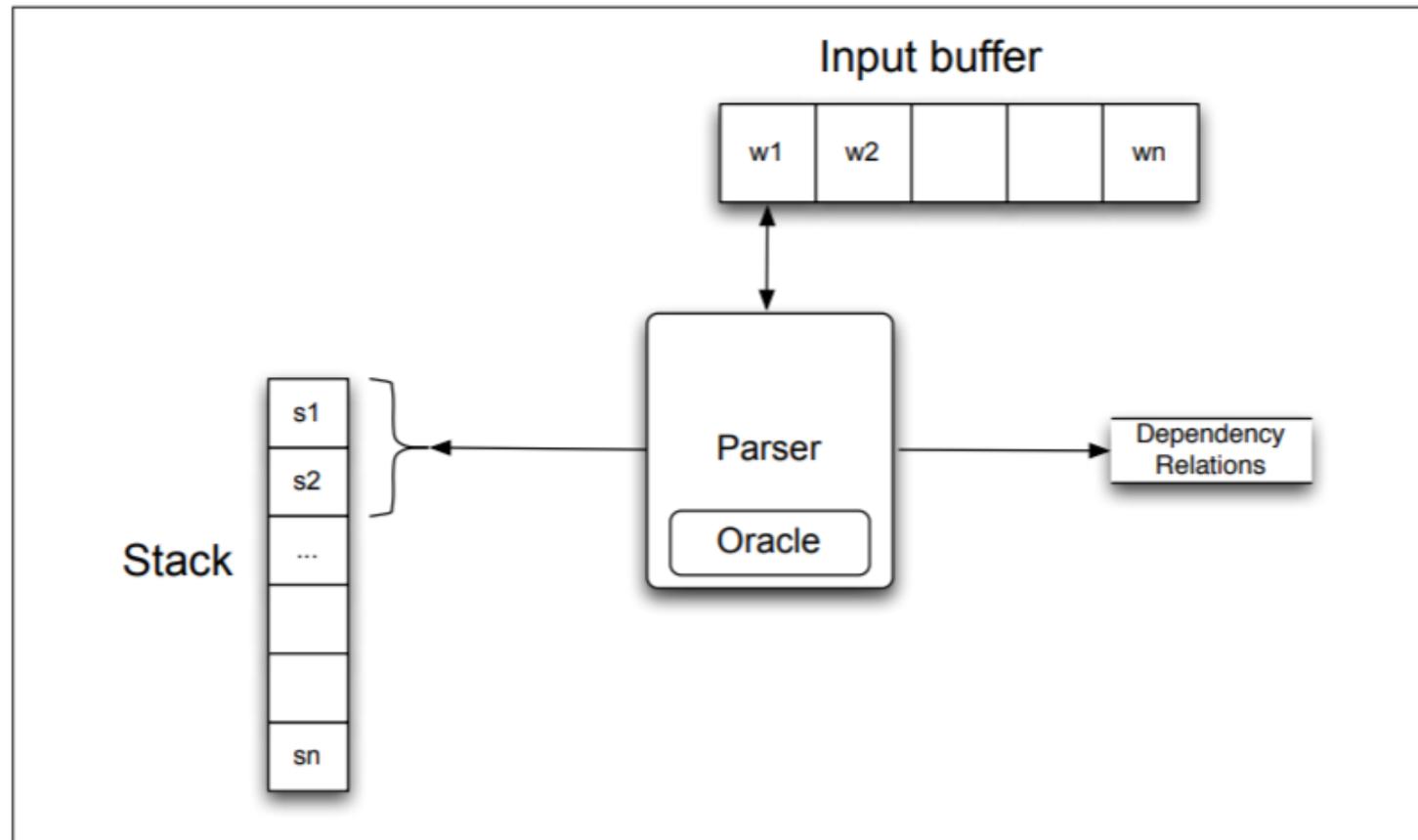
$x(*)$: x é folha

$*(x)$: x é raiz da árvore

MÉTODO

- Baseado na estratégia **shift-reduce** normalmente utilizada em compiladores
 - Palavras são empilhadas (*shift* do buffer de entrada para a pilha)
 - Buscam-se relações entre os elementos empilhados (fazendo-se a “redução” da pilha)
 - Em cada momento, há uma “configuração” do sistema, portanto
 - Estado da pilha
 - Buffer com a entrada
 - Relações já encontradas
 - O **segredo do parsing**: encontrar a operação apropriada em cada configuração (com base em um “oráculo” que orienta o processo)

MÉTODO BASEADO EM TRANSIÇÕES (*ARC TRANSITION*)



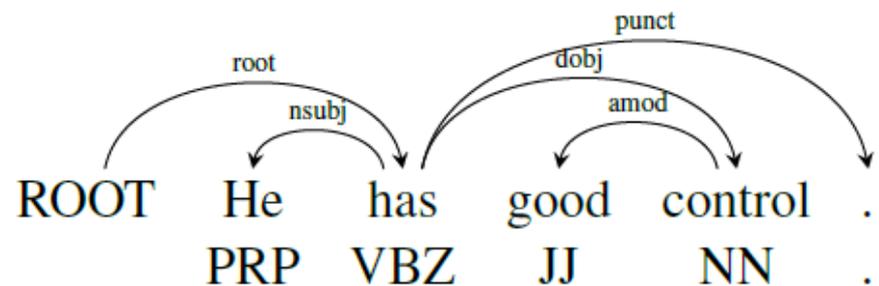
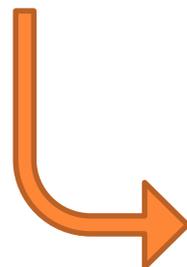
MÉTODO

- Operações possíveis
 - Shift: a palavra da frente do buffer é empilhada (no topo da pilha)
 - Formato de pilha deitada, para facilitar a visualização das demais operações
 - Left-Arc(R): é criada uma relação de dependência Rel entre a palavra no topo da pilha e a abaixo dela (removendo-se da pilha o dependente)
 - Right-Arc(R): é criada uma relação de dependência R entre a palavra abaixo do topo da pilha e a palavra do topo (removendo-se da pilha o dependente)

○ Esquema de funcionamento do método

- Pilha (*stack*) começa com ROOT
- Sentença a ser analisada no buffer
- A é o conjunto de relações de dependências identificadas
- Transições indicadas pelo oráculo

Transition	Stack	Buffer	A
	[ROOT]	[He has good control .]	\emptyset
SHIFT	[ROOT He]	[has good control .]	
SHIFT	[ROOT He has]	[good control .]	
LEFT-ARC (nsubj)	[ROOT has]	[good control .]	A \cup nsubj(has,He)
SHIFT	[ROOT has good]	[control .]	
SHIFT	[ROOT has good control]	[.]	
LEFT-ARC (amod)	[ROOT has control]	[.]	A \cup amod(control,good)
RIGHT-ARC (dobj)	[ROOT has]	[.]	A \cup dobj(has,control)
...
RIGHT-ARC (root)	[ROOT]	[]	A \cup root(ROOT,has)



ORÁCULO

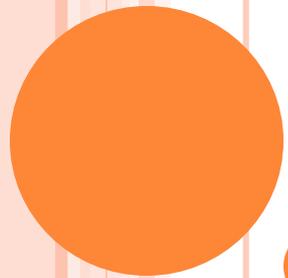
- No coração do processo
 - Responsável por orientar a análise
 - Em geral, é um classificador que, a partir de dados da configuração do sistema, indica a melhor operação
- Aprendizado de máquina
 - Atributos usuais (com **regressão logística** e **SVM**, normalmente) são as palavras no topo da pilha e na frente do buffer, suas etiquetas morfossintáticas, as relações já previstas e combinações desses e outros atributos.
 - Problemas dessas abordagens: engenharia de atributos (que atributos usar, que combinações de atributos fazer), dados esparsos, muito tempo para cômputo dos atributos

ORÁCULO

- No coração do processo
 - Responsável por orientar a análise
 - Em geral, é um classificador que, a partir de dados da configuração do sistema, indica a melhor operação
- Aprendizado de máquina
 - **Redes neurais** e os melhores resultados
 - Atributos usuais: *embeddings* e seu poder representacional (menos esparsidade, generalização) para palavras, etiquetas morfosintáticas e relações de dependência da configuração atual
 - Aprendizado automático de relevância dos atributos e suas combinações
 - Maior velocidade de processamento em relação ao cômputo dos atributos das abordagens anteriores

QUESTÕES

- Sistema baseado em transições e “projetividade” (ou seja, o não cruzamento de arcos na árvore de dependências)
 - Mas projetividade nem sempre é o ideal
- Busca gulosa e a velocidade de análise
 - Versus busca em um espaço maior de soluções (mas com maior tempo de processamento necessário)
 - Relembrando: qual o problema de se usar busca gulosa?



HUMANOS E PARSING

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Estruturas e palavras mais previsíveis (prováveis) são lidas mais rapidamente por humanos
 - Como se mede isso?

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Estruturas e palavras mais previsíveis (prováveis) são lidas mais rapidamente por humanos
 - Medidas empíricas: por exemplo, entropia vs. rastreamento do movimento dos olhos

PROCESSAMENTO HUMANO & PROBABILIDADE

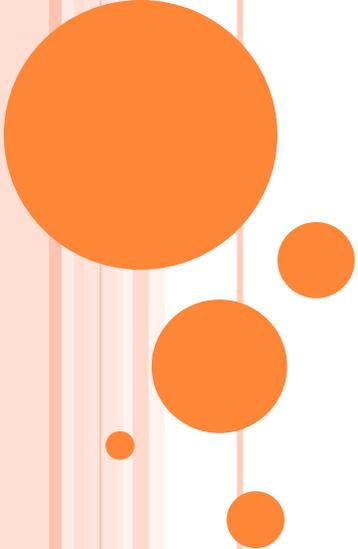
- Experimentos com humanos: **probabilidades na mente!**
 - Humanos desambiguam análises, preferindo análises mais prováveis
 - Sentenças ***garden-path***: temporariamente ambíguas
 - *The students forgot the solution was in the back of the book.*
 - *The horse raced past the barn fell.*
 - *The complex houses married and single students and their families.*

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Humanos desambiguam análises, preferindo análises mais prováveis
 - Sentenças *garden-path*: temporariamente ambíguas
 - *Maria beijou João e o irmão dele arregalou os olhos de espanto.*

PROCESSAMENTO HUMANO & PROBABILIDADE

- Experimentos com humanos: **probabilidades na mente!**
 - Humanos desambiguam análises, preferindo análises mais prováveis
 - Sentenças *garden-path*: há casos mais complexos!
 - *Um navio brasileiro entrava na baía um navio japonês.*



ATRIBUTOS E UNIFICAÇÃO

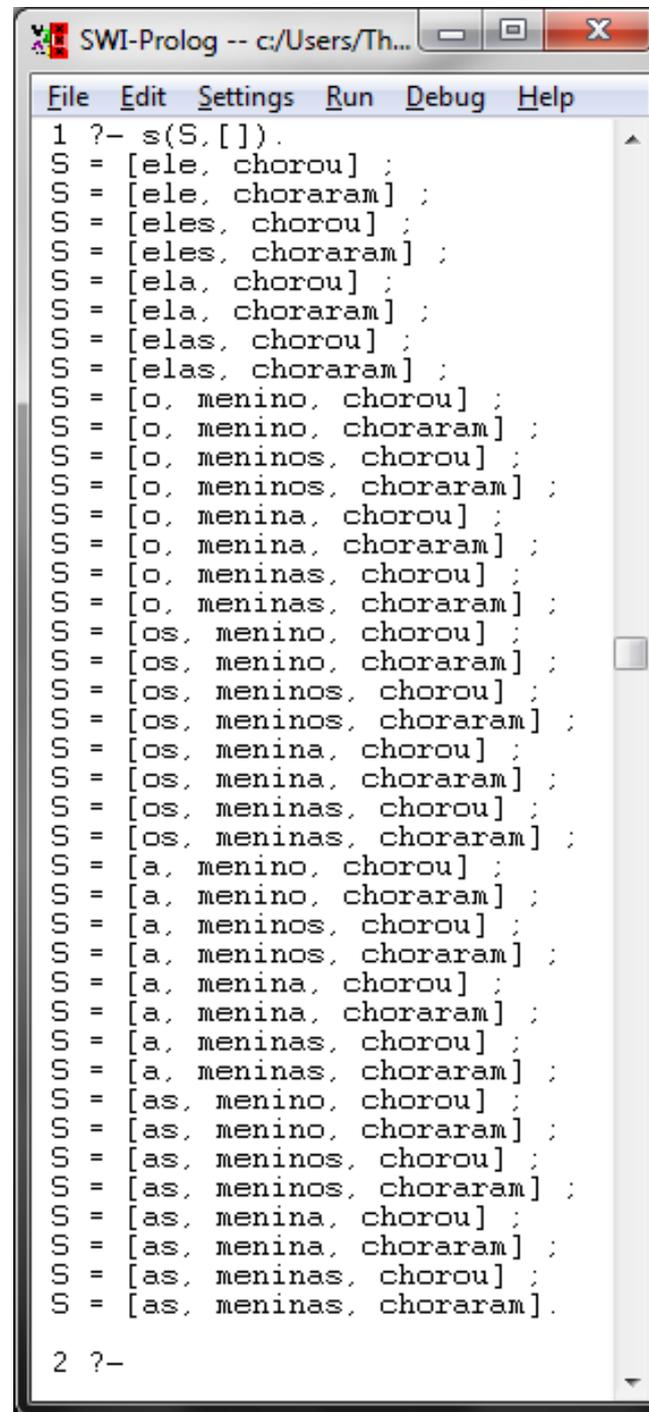
GRAMATICALIDADE

○ Exemplo simples

- $S \rightarrow SN SV$
- $SN \rightarrow \text{pronome} \mid \text{art subst}$
- $SV \rightarrow \text{verbo}$
- $\text{pronome} \rightarrow \text{ele} \mid \text{eles} \mid \text{ela} \mid \text{elas}$
- $\text{art} \rightarrow \text{o} \mid \text{os} \mid \text{a} \mid \text{as}$
- $\text{subst} \rightarrow \text{menino} \mid \text{meninos} \mid \text{menina} \mid \text{meninas}$
- $\text{verbo} \rightarrow \text{chorou} \mid \text{choraram}$

EXEMPLO

- O que aconteceu?
- Por que aconteceu?
- Como resolver?



```
SWI-Prolog -- c:/Users/Th...
File Edit Settings Run Debug Help
1 ?- s(S, []).
S = [ele, chorou] ;
S = [ele, choraram] ;
S = [eles, chorou] ;
S = [eles, choraram] ;
S = [ela, chorou] ;
S = [ela, choraram] ;
S = [elas, chorou] ;
S = [elas, choraram] ;
S = [o, menino, chorou] ;
S = [o, menino, choraram] ;
S = [o, meninos, chorou] ;
S = [o, meninos, choraram] ;
S = [o, menina, chorou] ;
S = [o, menina, choraram] ;
S = [o, meninas, chorou] ;
S = [o, meninas, choraram] ;
S = [os, menino, chorou] ;
S = [os, menino, choraram] ;
S = [os, meninos, chorou] ;
S = [os, meninos, choraram] ;
S = [os, menina, chorou] ;
S = [os, menina, choraram] ;
S = [os, meninas, chorou] ;
S = [os, meninas, choraram] ;
S = [a, menino, chorou] ;
S = [a, menino, choraram] ;
S = [a, meninos, chorou] ;
S = [a, meninos, choraram] ;
S = [a, menina, chorou] ;
S = [a, menina, choraram] ;
S = [a, meninas, chorou] ;
S = [a, meninas, choraram] ;
S = [as, menino, chorou] ;
S = [as, menino, choraram] ;
S = [as, meninos, chorou] ;
S = [as, meninos, choraram] ;
S = [as, menina, chorou] ;
S = [as, menina, choraram] ;
S = [as, meninas, chorou] ;
S = [as, meninas, choraram] .

2 ?-
```

GRAMATICALIDADE

- Sentença **correta/bem formada**

- **S → SN SV não basta**

- Normalmente é necessário que haja

- Concordância de número e gênero dentro do SN

- Concordância de número e pessoa entre SN e SV

- Concordância entre argumentos esperados pelo verbo e argumentos realizados

- Etc.

- **Restrições!**

GRAMATICALIDADE

- Questão: como resolver isso?

GRAMATICALIDADE

- Questão: como resolver isso?

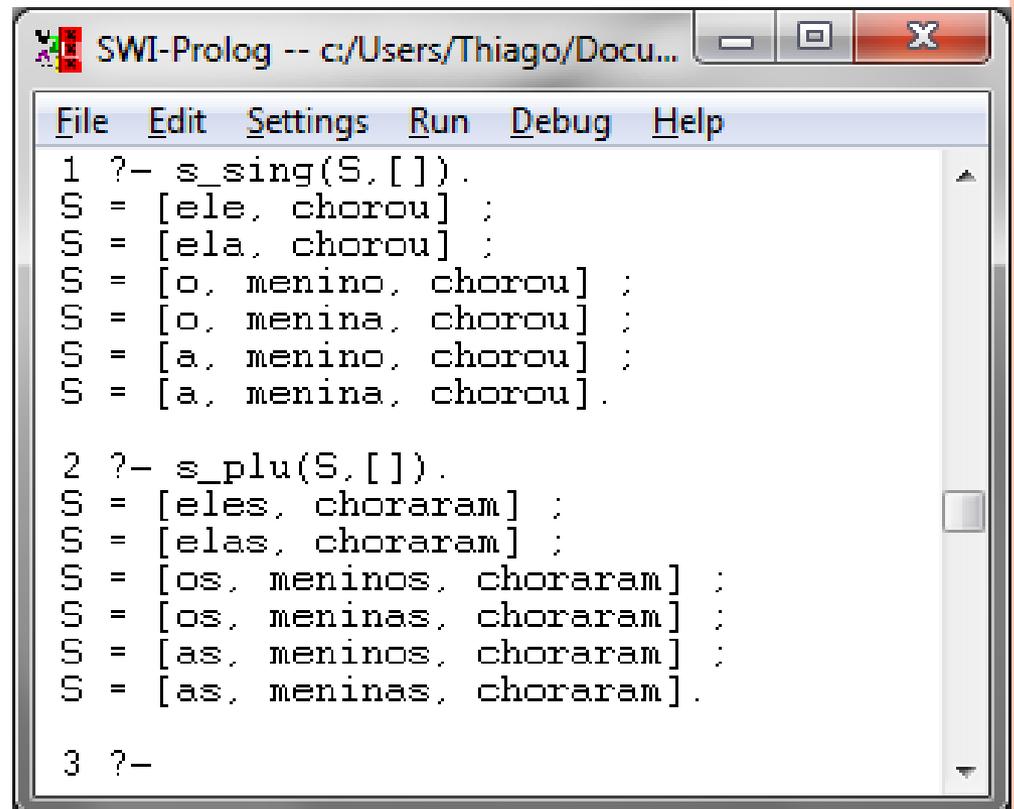
- 1ª opção: duplicar regras

- Exemplo para **número**, somente

- $S_{\text{sing}} \rightarrow SN_{\text{sing}} SV_{\text{sing}}$
- $S_{\text{plu}} \rightarrow SN_{\text{plu}} SV_{\text{plu}}$
- $SN_{\text{sing}} \rightarrow \text{pronome}_{\text{sing}} \mid \text{art}_{\text{sing}} \text{subst}_{\text{sing}}$
- $SN_{\text{plu}} \rightarrow \text{pronome}_{\text{plu}} \mid \text{art}_{\text{plu}} \text{subst}_{\text{plu}}$
- $SV_{\text{sing}} \rightarrow \text{verbo}_{\text{sing}}$
- $SV_{\text{plu}} \rightarrow \text{verbo}_{\text{plu}}$
- $\text{pronome}_{\text{sing}} \rightarrow \text{ele} \mid \text{ela}$
- $\text{pronome}_{\text{plu}} \rightarrow \text{eles} \mid \text{elas}$
- ...

EXEMPLO

- Número resolvido
- E gênero?



```
SWI-Prolog -- c:/Users/Thiago/Docu...
File Edit Settings Run Debug Help
1 ?- s_sing(S, []).
S = [ele, chorou] ;
S = [ela, chorou] ;
S = [o, menino, chorou] ;
S = [o, menina, chorou] ;
S = [a, menino, chorou] ;
S = [a, menina, chorou].

2 ?- s_plu(S, []).
S = [eles, choraram] ;
S = [elas, choraram] ;
S = [os, meninos, choraram] ;
S = [os, meninas, choraram] ;
S = [as, meninos, choraram] ;
S = [as, meninas, choraram].

3 ?-
```

GRAMATICALIDADE

- **Questão: como resolver isso?**
 - 1ª opção: duplicar regras
 - **Desvantagens sérias**
 - Explosão do número de regras
 - Perda de generalidade da gramática
 - Legibilidade da gramática prejudicada

GRAMATICALIDADE

- **Questão: como resolver isso?**
- **2ª opção: atributos/propriedades associados aos constituintes**
 - Exemplo
 - $S \rightarrow SN SV$
 - Atributo “número” para SN e SV
 - $SN.número = SV.número$
 - **Vantagens: elegância, generalidade, legibilidade**

ESTRUTURA DE ATRIBUTOS

○ *Feature structure*

- Conjunto de atributos e valores
 - Valores atômicos ou subestruturas de atributos
- Matriz atributo-valor

$$\begin{bmatrix} \text{atributo}_1 & \text{valor}_1 \\ \text{atributo}_2 & \text{valor}_2 \\ \dots & \dots \\ \text{atributo}_N & \text{valor}_N \end{bmatrix}$$

ESTRUTURA DE ATRIBUTOS

○ *Feature structure*

- Exemplos

[Número singular]

[Número singular
Pessoa 3a]

[Categoria SN
Número singular
Pessoa 3a]

[Categoria SN
Concordância [Número singular
Pessoa 3a]]

ESTRUTURA DE ATRIBUTOS

○ Operações

- União de estruturas compatíveis
- Rejeição de estruturas incompatíveis

○ Unificação!

○ Exemplos

$[\text{Número singular}] \cup [\text{Número singular}] = [\text{Número singular}]$

$[\text{Número singular}] \cup [\text{Número plural}] \rightarrow \text{FALHA}$

ESTRUTURA DE ATRIBUTOS

○ Portanto

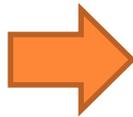
- Utilidade das estruturas de atributos
 - Representar informação sobre algum objeto linguístico
 - Impor restrições sobre comportamento do objeto linguístico
- Unificação
 - União de informações de estruturas de atributos compatíveis
 - Produz uma nova estrutura de atributos, que pode ser mais específica ou idêntica às originais

ATRIBUTOS E GRAMÁTICAS

- Uso das estruturas de atributos
 - Forma de associar restrições às regras gramaticais
 - Extensão das regras gramaticais
 - Em vez de constituintes, temos constituintes+atributos/propriedades

Originalmente

$S \rightarrow SN SV$



Extensão para concordância de número entre SN e SV

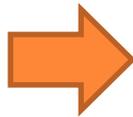
$S \rightarrow SN SV$
 $\langle SN \text{ num} \rangle = \langle SV \text{ num} \rangle$

ATRIBUTOS E GRAMÁTICAS

- Uso das estruturas de atributos
 - Forma de associar restrições às regras gramaticais
 - Extensão das regras gramaticais
 - Em vez de constituintes, temos constituintes+atributos/propriedades

Originalmente

$S \rightarrow SN SV$



Extensão para concordância de número entre SN e SV

$S \rightarrow SN SV$
 $\langle SN \text{ num} \rangle = \langle SV \text{ num} \rangle$

Significado

S se SN e SV

Significado

S se SN e SV e se SN.num=SV.num

EXERCÍCIO

- Adicionar atributos de gênero e número à gramática abaixo

S → SN SV

S → SV

SN → pronome

SN → substantivo

SN → artigo substantivo

SV → verbo

SV → verbo SN

SV → verbo SN SP

SP → preposição SN

pronome → eu | ele | ela | ...

artigo → o | a | os | as

substantivo → carro | carros | pessoa | ...

verbo → corre | correm | morre | morrem
| ...

preposição → de | para | ...

INÍCIO DA RESPOSTA

- S → SN SV
SN.num = SV.num
- SN → artigo substantivo
artigo.num = substantivo.num
artigo.gen = substantivo.gen
SN.num = substantivo.num
- SV → verbo
SV.num = verbo.num
- artigo → o
artigo.num = singular
artigo.gen = masculino
- substantivo → copo
substantivo.num = singular
substantivo.gen = masculino
- verbo → quebrou
verbo.num = singular

...

Exercício

Fazer derivação sintática da sentença “o copo quebrou”

O que acontece com “o copo quebraram”?

ATRIBUTOS E GRAMÁTICAS

○ DCG

- É possível representar atributos na forma de argumentos dos elementos
- Exemplo

`s --> sn(Num,Gen), sv(Num).`

`sn(Num,Gen) --> art(Num,Gen), subst(Num,Gen).`

...

`art(singular,masculino) --> [o].`

`art(plural,masculino) --> [os].`

...

ATRIBUTOS E GRAMÁTICAS

○ DCG

- É possível representar atributos na forma de argumentos dos elementos

- Exemplo

$s \text{ --> } sn(\text{Num}, \text{Gen}), sv(\text{Num}).$

$sn(\text{Num}, \text{Gen}) \text{ --> } art(\text{Num}, \text{Gen}), subst(\text{Num}, \text{Gen}).$

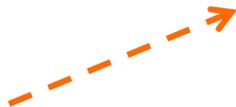
...

$art(\text{singular}, \text{masculino}) \text{ --> } [o].$

$art(\text{plural}, \text{masculino}) \text{ --> } [os].$

...

$S \rightarrow SN SV$
 $\langle SN \text{ Num} \rangle = \langle SV \text{ Num} \rangle$



ATRIBUTOS E GRAMÁTICAS

- Atributos podem ser computados
 - **Após** geração/reconhecimento das sentenças, cortando algumas análises
 - **Durante** geração/reconhecimento, podando possibilidades