








PROCESSAMENTO DE LINGUAGEM NATURAL

Prof. Dr. Carlos Alberto de Oliveira

SUMÁRIO

1. CONSIDERAÇÕES INICIAIS 
2. O QUE É LINGUAGEM NATURAL 
3. A VISÃO QUE TEMOS DA LÍNGUA -
IMPLICAÇÕES 
4. ESCOLHAS DE PROCESSAMENTO 
5. OPERACIONALIZAÇÃO
EM LÍNGUA ESCRITA 
6. COMENTÁRIOS E APLICAÇÕES 
7. BIBLIOGRAFIA 

FIM

OPERACIONALIZAÇÃO DE INTERFACES



- A Gramática Gerativo-transformacional



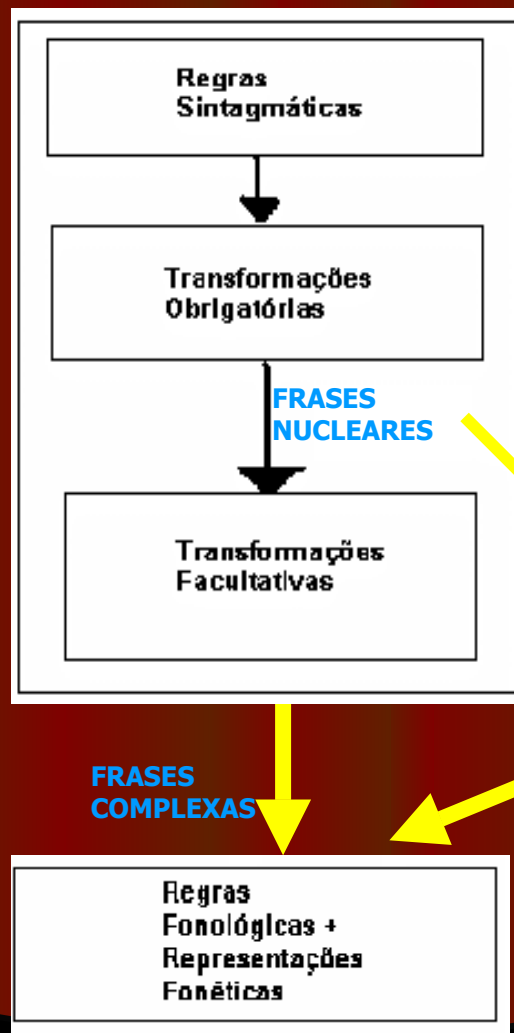
- Outras Gramáticas (dentre muitas)

- Gramática Formal
- Gramática de Estados Finitos
- Gramática de Cláusulas Definidas
- Gramática Probabilística
- Gramática Lexicalizada



A Gramática Gerativo-transformacional

NOAM CHOMSKY



Componente Sintático

Componente Fonológico

A Gramática Gerativo-transformacional



•Dois componentes sintáticos:

o sintagmático
o transformacional

O componente sintagmático:

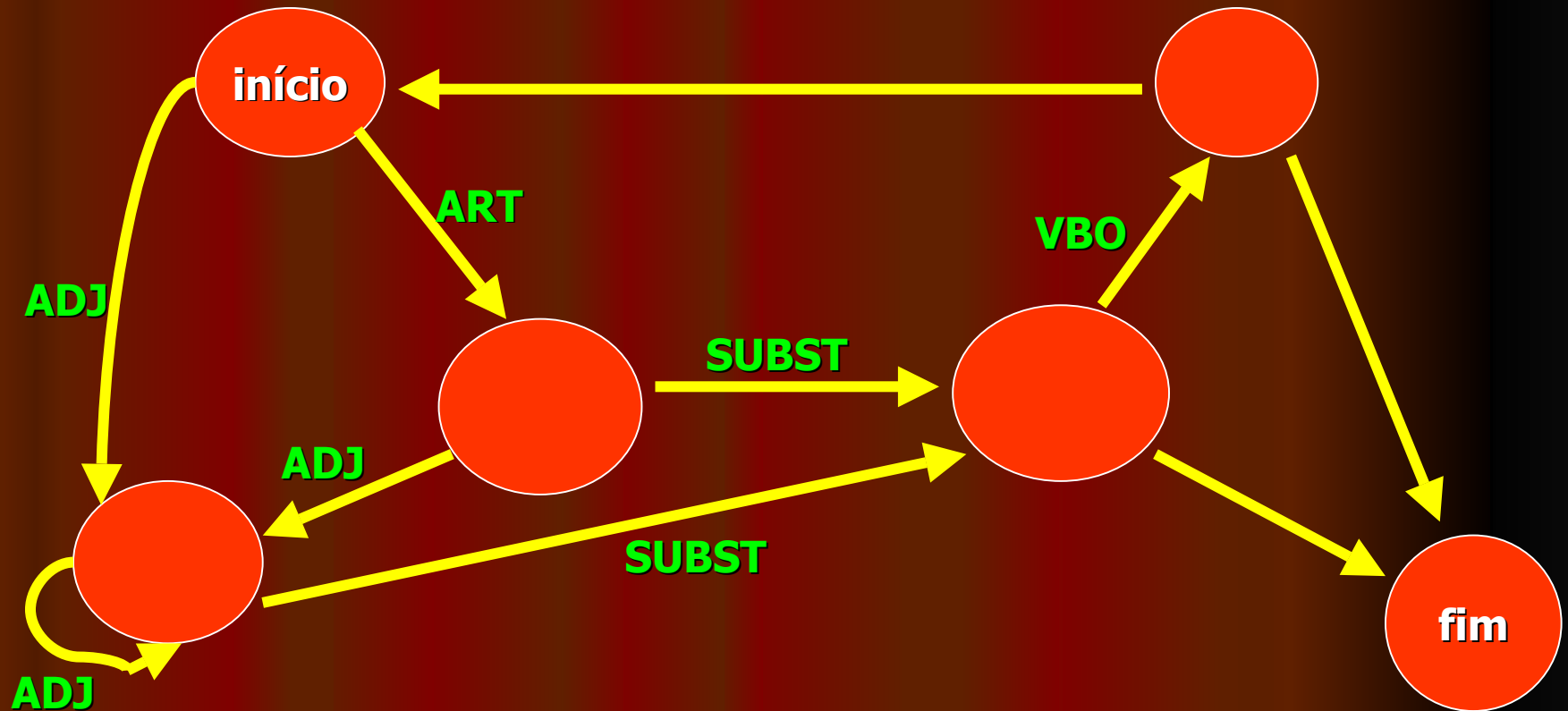


regras de reescrita
regras de subcategorização
o léxico

O componente transformacional



Gramática de estados finitos



Gramática de Cláusulas Definidas

A frase “me comi a fruta” está correta sintaticamente.
Mas, ...

Assim, subcategorizam-se as possibilidades:

$$SN_1 = \{ \text{eu} \}$$
$$SN_0 = \{ \text{me, mim, ...} \}$$

Ou seja, em lugar de introduzir novas regras, *aumentam-se as existentes*, subcategorizando-as.

Gramática de Cláusulas Definidas

*sentenca(S_0S) :-
 sintagma_nominal(S_0S_1), sintagma_verbal(S_1S).*

sintagma_nominal (S_0S) :- determinante(S_0S_1), nome(S_1S).

*sintagma_verbal (S_0S) :-
 verbo(S_0S_1), sintagma_nominal(S_1S).*

determinante([o / S], S).

nome([músico / S], S).

nome([violino / S], S).

verbo([toca / S], S).

Gramática de Cláusulas Definidas



sentenca --> sintagma_nominal(Numero, pessoa),
sintagma_verbal(Numero).

sintagma_nominal(Numero, Tipo) --> determinante(Numero),
nome(Numero, Tipo).

sintagma_verbal(Numero) --> verbo(Numero),
sintagma_nominal(Numero, coisa).

determinante(singular) --> [o].
determinante(plural) --> [os].

nome(singular, pessoa) --> [músico].
nome(plural, pessoa) --> [músicos].
nome(singular, coisa) --> [violino].

verbo(singular) --> [toca].
verbo(plural) --> [tocam].

Gramática Probabilística

Livre de Contexto

S → SN SV *[1,00]*

SN → Pron *[0,10]* /
Nome *[0,10]* /
Subst *[0,20]* /
Art Subst *[0,50]* /
SN SPrep *[0,10]*

SV → Verbo *[0,60]* /
SV SN *[0,20]* /
SV SPrep *[0,20]*

SPrep → Prep SN *[1,00]*

Subst → gato *[0,10]* / arroz *[0,15]*

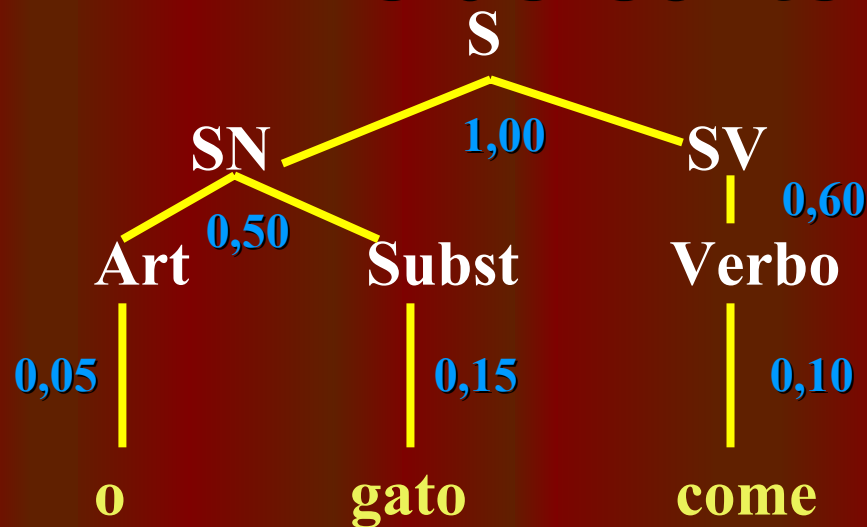
Verbo → vê *[0,15]* / come *[0,25]*

Art → o *[0,05]* / uma *[0,10]*

Prep → por *[0,30]* / em *[0,03]*

Pron → ...

Gramática Probabilística Livre de Contexto



A probabilidade da árvore como um todo é
 $1,0 \times 0,5 \times 0,05 \times 0,15 \times 0,6 \times 0,1 = 0,000225$

PROBLEMA: “*comer uma banana*” e “*comer uma bandana*”, depende apenas de “*banana*” ou “*bandana*” já que **não existe relacionamento entre “*comer*” e seu objeto.** Uma **GPLC lexicalizada** poderia resolver tal questão!

Gramática Livre de Contexto Lexicalizada

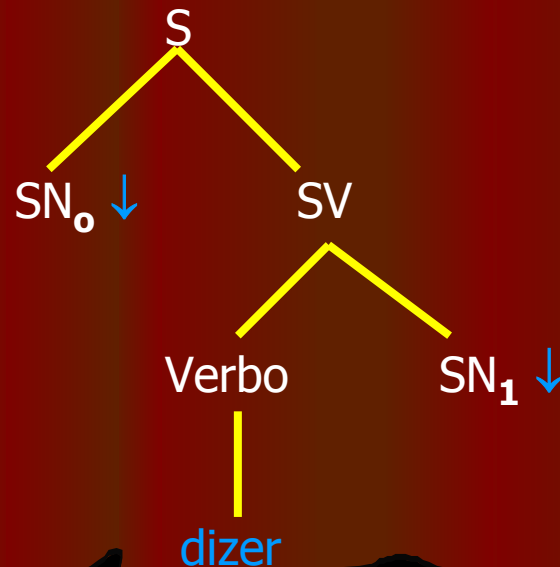
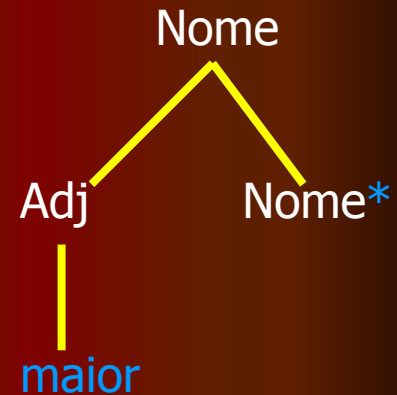
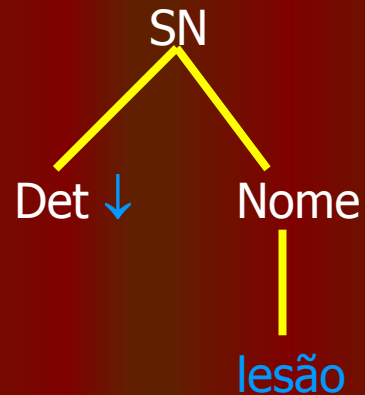
GLCL é uma 5-tupla (Σ, NT, I, A, S)
que gera *árvores* (e não *cadeias*)

- Σ é um conjunto de símbolos terminais
- NT é um conjunto de símbolos não-terminais
- I e A são conjuntos de árvores rotuladas por símbolos terminais e não-terminais
- S é um símbolo inicial não-terminal

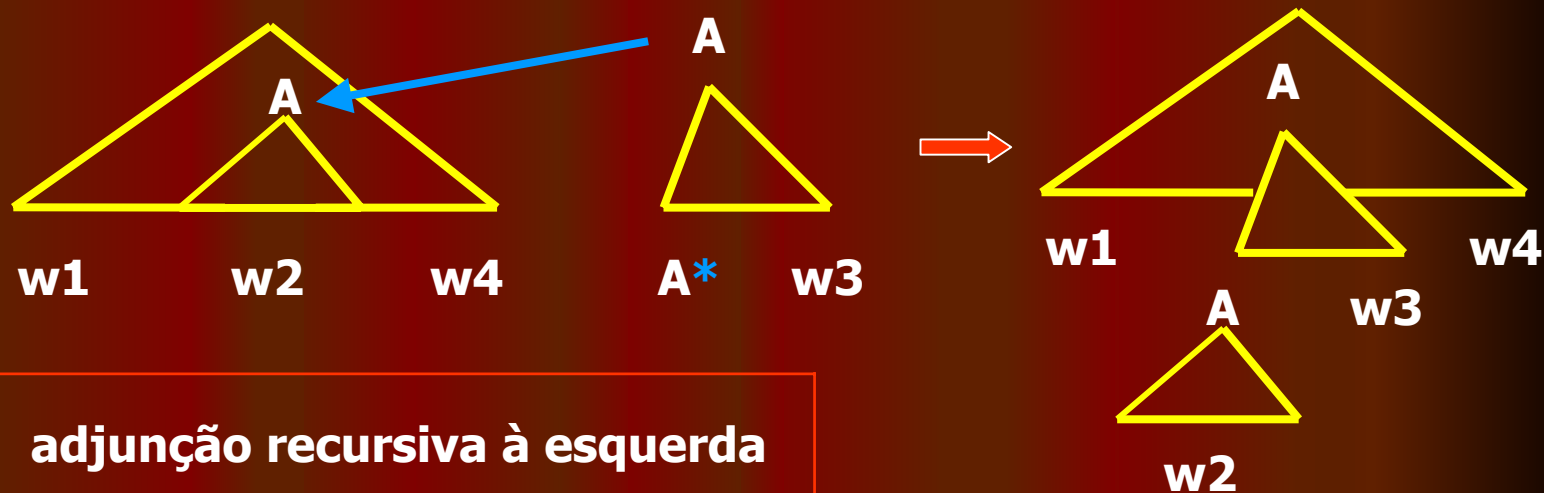
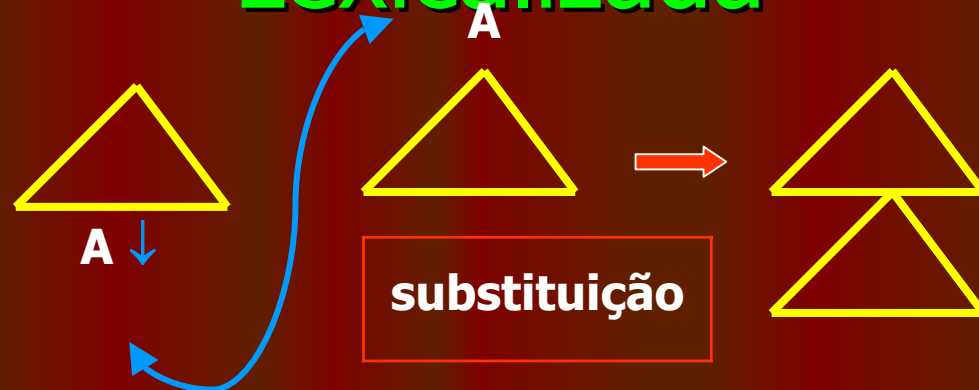
Gramática Livre de Contexto Lexicalizada

- Em cada árvore **inicial** de **I**:
 - os nós interiores são rotulados por símbolos não-terminais
 - os símbolos não terminais na fronteira são marcados por **substituição** (convenção: ↓)
- Em cada árvore **auxiliar** de **A**:
 - os nós interiores são rotulados por símbolos não-terminais
 - todos menos um dos símbolos não-terminais na fronteira são marcados por **substituição**
 - o não-terminal remanescente na fronteira da árvore é chamado de **base** (convenção: *)

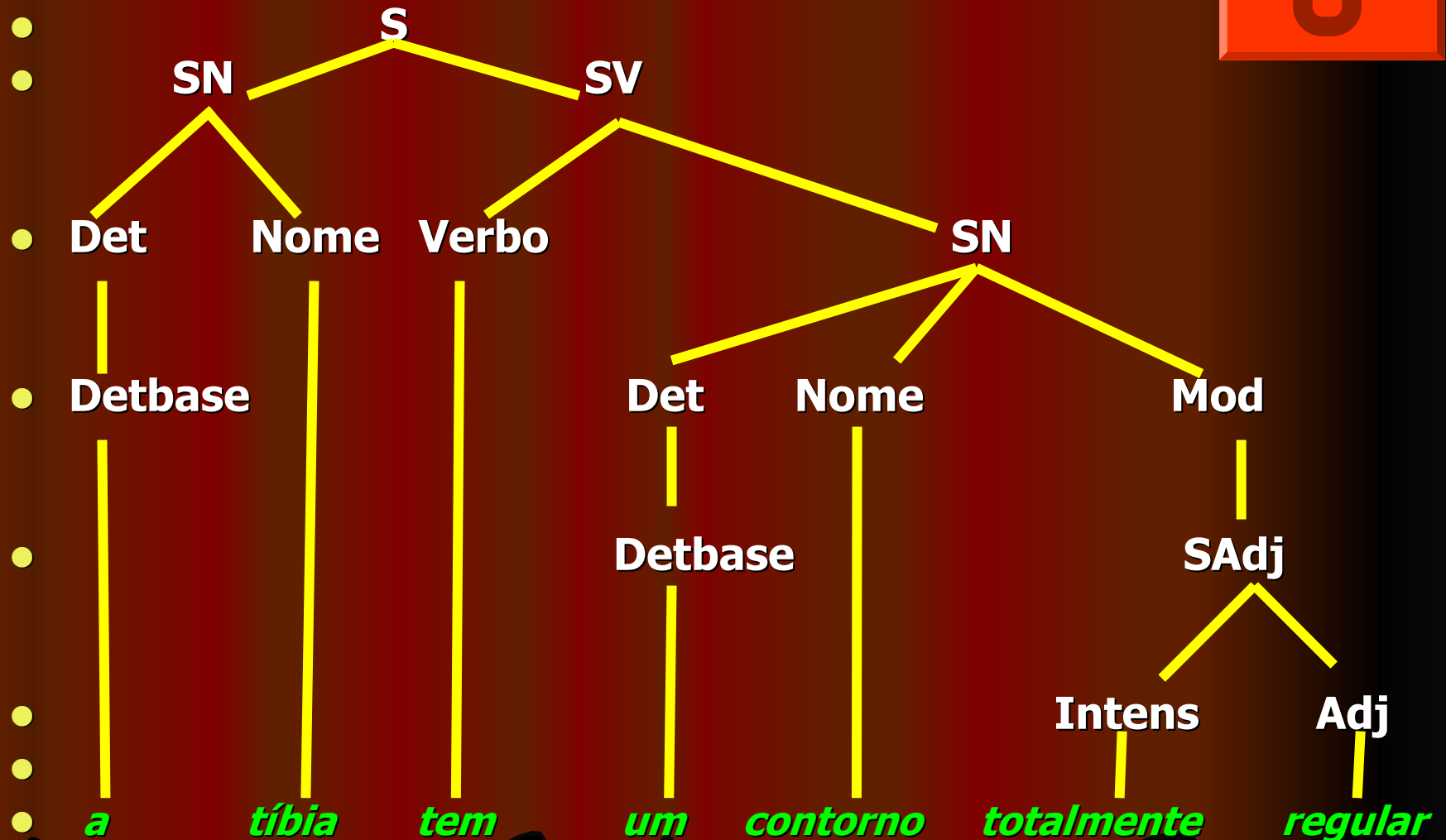
Gramática Livre de Contexto Lexicalizada



Gramática Livre de Contexto Lexicalizada



Gramática Livre de Contexto Lexicalizada



COMENTÁRIOS E APLICAÇÕES


Modelar o USUÁRIO como parte *essencial* do Sistema (não apenas como parte integrante ou accessória)

- colher o **saber/conhecimento lingüístico** do mesmo (*manualmente* ou *automatizadamente*)
- lembrar sempre que ele (**o usuário**) é que vai, um truísmo, usar o sistema
- lembrar de uma das “*leis de Murphy*” (extendida):

faça um sistema que só um 'idiota' não saberia utilizar que vai aparecer um 'idiota' que derrubará o seu sistema

COMENTÁRIOS E APLICAÇÕES

Lembrar, pelo exposto na tela anterior, que
a linguagem sempre será determinada pelo domínio 

- por exemplo: o **Léxico** de um domínio médico será sempre específico 
- a **Gramática** (regras de reescrita) também

COMENTÁRIOS E APLICAÇÕES



Optar, conforme a aplicação, por:

- tratar o **Léxico** (economia de memória; gasto em tempo computacional)
- ou incluir todas as **lexias** possíveis de ocorrerem (gasto de memória; economia em tempo computacional)

Lexema	Lexias
mudar	mudança, mudaremos, mudem, mudam, ...
pedra	pedras, pedrada, pedregulho, pedreira, ...

COMENTÁRIOS E APLICAÇÕES



- Interface Dialógica Em Linguagem Natural **AL** para sistemas especialistas - **IDEAL** 
- Linguagem de transformações 
- Interface para o INSIRIUS
- Ferramenta de auxílio à elaboração de currículos educacionais

COMENTÁRIOS E APLICAÇÕES



- **Alguma sugestão de pesquisa:**
 - *gramática por indução* (por exemplo, para avaliar autoria de textos enviados por internet)
 - *linguagem alto nível* para situações em que isto seja de extrema importância (ex.: robôs marcianos)

BIBLIOGRAFIA



- RUSSELL, Stuart; NORVIG, Peter.
Inteligência Artificial. Rio de Janeiro:
Elsevier, 2004
- BARBOSA, Cynthyan R. S. C. de. *Técnicas de parsing para gramática livre de contexto lexicalizada da Língua Portuguesa*. Tese de Doutorado. ITA, 2004

ESCOLHAS



geração



fala

escrita

interpretação



fala

escrita

interfaces

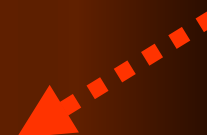
(um estilo de interação)



fala

escrita

E é sobre isto
que falaremos
HOJE!!



O QUE É LINGUAGEM NATURAL

MODALIDADE ORAL

MODALIDADE ESCRITA

sons (fonemas)



letras (grafemas)

fluxo contínuo



unidades discretas



fala, audição



digitação, visão

inflexão, entonação



pontuação

sintagmas fragmentados



sintagmas completos



....

....

O QUE É LINGUAGEM NATURAL

NÍVEIS DE ANÁLISE

MORFOLÓGICO

⇒ palavras (léxico)

SINTÁTICO

⇒ gramáticas (regras)

SEMÂNTICO

⇒ significado

PRAGMÁTICO

⇒ significação

AQUI RESIDE
A PRIMEIRA
ESCOLHA!!!

⇒ regras de uso em
condições reais(*componente
extralingüístico*)

O QUE É LINGUAGEM NATURAL

A REPRESENTAÇÃO DO CONHECIMENTO

Casos

Agentes

Dependência Conceitual

Redes Semânticas

Frames

Lógica

Regras Gramaticais



....

A VISÃO QUE TEMOS DA LÍNGUA E SUAS IMPLICAÇÕES

VISÃO ANALOGISTA

- Regras Gramaticais são precisas
- Conhecimento (logo, pode ser sistematizado)

OUTRAS DENOMINAÇÕES

- Racionalista
- Mentalista
- Naturalista

ESTUDIOSOS

- Chomsky

VISÃO ANOMALISTA

- Uso em condições reais são imprecisas
- Saber (logo, só pode ser aprendido)

OUTRAS DENOMINAÇÕES

- Subjetiva
- Realista
- Convencionalista

ESTUDIOSOS

- Piaget, Vygotsky

AQUI RESIDE
A SEGUNDA
ESCOLHA!!!

- e, de certo modo, a **IA**

A VISÃO QUE TEMOS DA LÍNGUA E SUAS IMPLICAÇÕES

UM CONJUNTO DE **PALAVRAS**
OU UM CONJUNTO DE **FRASES**

LOGO, UM **PRODUTO**, UM
SISTEMA, UMA ESTRUTURA

IMPLICA

- o uso de Lógica Clássica e de Gramáticas Formais
- considerar que um existe um emissor (que codifica) e um receptor (que decodifica)

UM **TEXTO**



LOGO, UM **PROCESSO** E VÁRIOS
SISTEMAS

IMPLICA

- o uso de Lógicas modais, paraconsistentes, e Teorias Textual e do Discurso
- considerar que o significado se estabelece NO ATO da interação

**AQUI RESIDE
A TERCEIRA
ESCOLHA!!!**

REGRAS DE REESCRITA



S → **SN SV**
SN → **Art Subst**
SV → **Verbo SN**

LÉXICO (subcategorização)

subst1= homem, aluna, transeunte

subst2= cachorro

subst3= perna

subst4= maçã

verbo1= comer, chamar

verbo2= morder

FRASES POSSÍVEIS

O homem chama o transeunte.

A aluna come a maçã.

(*) O transeunte come a perna.

O cachorro morde a perna.

(*) O cachorro chama a maçã.

(*) A perna come a maçã.

(*) A maçã chama a perna.

Gramática = (V ocabulário, Σ , P roduções, S)

$V \Rightarrow$ CONJUNTO FINITO, NÃO-VAZIO

$P \Rightarrow$ CONJUNTO FINITO DA FORMA:

$\alpha \rightarrow \beta$, onde $\alpha \in V^* N V^*$ e $\beta \in V^*$

N É O OPERADOR QUE INDICA QUE O SÍMBOLO ANTEREDEDENTE PODERÁ
OCCORRER QUALQUER NÚMERO DE VEZES

$N = V - \Sigma \Rightarrow$ ALFABETO NÃO-TERMINAL

$\Sigma \Rightarrow$ ALFABETO TERMINAL, SUBCONJUNTO NÃO-VAZIO DE V

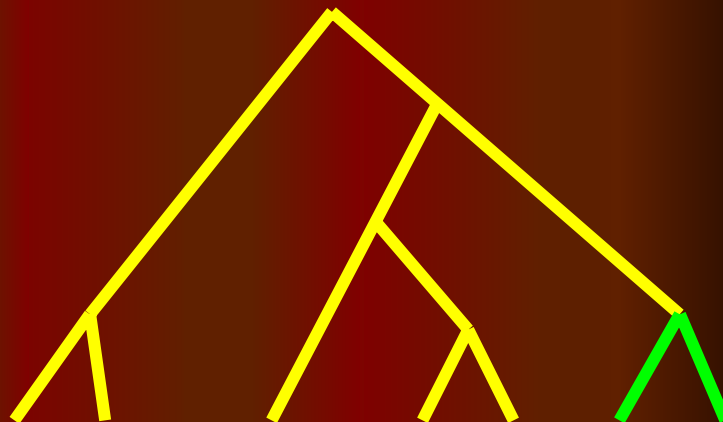
$S \Rightarrow$ ELEMENTO INICIAL DE Σ

A GRAMÁTICA FORMAL DE CHOMSKY

REPRESENTAÇÃO ARBÓREA



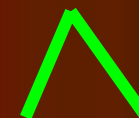
O HOMEM RECEBEU UMA FOTO



O HOMEM RECEBEU UMA FOTO DE FRANCA

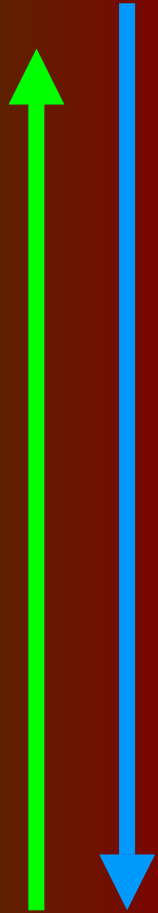


O HOMEM RECEBEU UMA FOTO



DE FRANCA

- Estrutura Profunda (cognição; pensamento)



- **o que se quer dizer**

- **aplicação de regras**

- **aplicação de transformações**

- Estrutura de Superfície (realização da fala)

As transformações

- De substituição

- **EP** *Carlos* almoçar



- **ES** *ele* almoçou

- *o professor* almoçou

de **permuta**, de **adição** e de **supressão**

- **EP** gato comer rato
- (*def* + gato) (*pass* + comer) (*indef* + rato)
- o gato comeu um rato (possível **ES**)
- *um rato* comeu *o gato*
- um rato (*foi comido*) o gato
- um rato foi comido *pelo* gato (possível **ES**)
- um rato foi comido *pelo gato*
- **ES** Um rato foi comido



LINGUAGEM

- Uma **linguagem formal** é definida como um conjunto (possivelmente infinito) de **cadeias**. Normalmente, trata-se a **linguagem natural (LN)** com fosse formal (embora não haja uma correspondência perfeita entre ambas).

GRAMÁTICA

- É um conjunto finito de **regras** que especificam uma linguagem.
- O **lingüista** é um cientista que busca definir uma **LN** como **ela é** (com gramáticas **descritivas**). O **gramático** busca prescrever a **LN** como **ela deve ser** (gramática **normativa** ou **prescritiva**).

DISCURSO / TEXTO (língua em condições reais de uso)

- **Deixis** (ex.: uma mensagem numa garrafa na praia)
 - Encontre-me **aqui hoje** com um sapato bem **grande**.

- **Referência**

→ **João** falou com **José**.

- ("João" e "José" **referem-se** a pessoas).

→ **Eles** disseram **isto**: não podemos mais compactuar com a ignorância.

- ("Eles" **referencia** "João" e "José", anaforicamente)
- ("isto" **referencia** "não podemos mais compactuar com a ignorância", cataforicamente)



“ ... a linguagem passa a ser encarada como forma de ação, **ação sobre o mundo**, dotada de **intencionalidade**, veiculadora de **ideologia**, caracterizando-se, portanto, pela **argumentatividade**. ”

Koch, I. G. V. *Argumentação e linguagem*. São Paulo: Cortez, 1984



**RELEVANDO QUE ISTO É APENAS UMA BRINCADEIRA,
ATENTEM PARA A DIFERENÇA (SEMPRE EM AMARELO)
ENTRE A PALAVRA FALADA E A PALAVRA ESCRITA.**

Sapassado, taveu no cuzinha tomano uma pincumel e cuzinhano um quidicarne cumastomate pra fazê uma macarronada cum galinhassada.

Quascaí dissusto quandovi um barúi vindidendufoño. Parecia tidiguerra!

A receita mandopô midipipoca denda galinha prassá.

O forno isquentô, o mistorô e o fiofó da galinhispludiu. Nossinhora!

Fiquei branco quineim lidileite. Foi um treim doidimais, sô!

Quascaí dendapia. Fiquei sem sabê doncovim, oncotô, poncovô.
Grazadeus ninguém simachucô!



FRAGMENTO DE DICIONÁRIO DE UM DOMÍNIO MÉDICO

dic_n(ewing, sing,masc)

dic_n(osteopenia, sing,fem)

dic_n(osteopenias, plu,fem)

dic_n(metáfise, sing,fem)

dic_n(paciente, sing,masc)

dic_adj(localizada, sing, fem)

dic_adj(localizado, sing, masc)

dic_adj(significativa, sing. fem)

dic_adj(significativo, sing, masc)

dic_ncom(partes moles,plu,fem)

dic_ncom(reação periosteais, plu,fem)



Domínio: Radiografia da tíbia

Cenário R1.1

Qual é a incidência?

Quais as principais estruturas anatômicas que foram vistas?

Compromete espaço reticular?

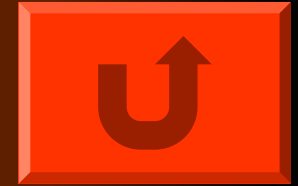
Descreva a integridade.



Cenário R1.2

Descreva a chapa.

Qual é o tipo de fratura?



- Banco de gramemas (dicionários fechados)
- Banco de regras gramaticais ('**blocos mínimos**' – hoje gramática lexicalizada)
- Representações de conhecimento (várias)
- Operacionalização via *blackboard* (hoje, **agentes**)
- Aquisição de conhecimento



>> Exibir o quadrado

-- Em quais das ações abaixo se enquadra EXIBIR?

1 - mostrar 2 - criar 3 – alterar

>> 1

-- Qual quadrado? 1- (10,45) 2 – (20, 27)



>> Mudar o quadrado em 10,45

-- Mudar em 1 – cor 2 – posição 3 – tamanho?

>> posição

-- Qual posição?

>> Mudar o **lugar** do quadrado **de 10,45 para 19,20**

>> Rotacionar o círculo