



F U N D A Ç ã O
GETULIO VARGAS

EMAp

Escola de
Matemática Aplicada

Gramáticas computacionais no formalismo HPSG utilizando a *Grammar Matrix*

Leonel Figueiredo de Alencar (UFC e
Visitante EMAp/FGV)

Alexandre Rademaker (IBM Research Lab
e EMAp/FGV)



Estrutura de constituintes

- flat S
- NP VP
- NP VP com modificação adjetival



1º encontro 12 de abril de 2021

Gramática Livre de Contexto

- CFG1

$S \rightarrow D N V D N$

- CFG2

$S \rightarrow NP VP$

$NP \rightarrow D A^* N$

$VP \rightarrow V (NP) (VP)$

$CS \rightarrow C S$

$C \rightarrow \text{"that"}$



Teoria X-barra

$XP \rightarrow YP X'$

$X' \rightarrow X ZP$

YP especificador

X núcleo

ZP complemento

X' X-barra



Paralelismo entre categorias

- Sujeitos e determinantes
 - **We** created a monster.
 - **our** creation of a monster
- (Sag; Wasow; Bender, 2003, p. 64)

Gramática Universal

$XP \rightarrow YP, X'$

$X' \rightarrow X, ZP$

- Parâmetro da Ordem
 - Inglês, Português etc.: +
 - Japonês: -

(Mioto; Silva; Lopes, 2005, p. 35)

Exemplos

- the cat sleeps under the bridge
- Neko wa hashi no shita de nemuru
- cat bridge under sleeps
- The dog chases the cat.
- Inu wa neko o oikakemasu.
- cat dog chases



Construção de uma minigramática

- Gramática do inglês gerada pela **Grammar Matrix** disponível em:
<https://github.com/LR-POR/tutorial>

Algumas referências

- KLENK, Ursula. *Generative Syntax*. Tübingen: Narr, 2003.

Este livro apresenta a evolução da gramática gerativa a partir do trabalho de Chomsky nos anos de 1950, desembocando nos formalismos não transformacionais LFG, GPSG e HPSG.


Algumas referências

- BENDER, E. M. Grammar Engineering for Linguistic Hypothesis Testing. In: GAYLORD, N. et al. (Org.). *The Proceedings of the Texas Linguistics Society 10: Computational Linguistics for Less-Studied Languages*. Stanford: CSLI, 2008. p. 16-36.

Este artigo mostra como a implementação computacional pode servir para testar hipóteses linguísticas.



2º encontro 19 de abril de 2021



Grammar Matrix → Gramática
computacional (TDL + Lisp) ← HPSG
(teoria gramatical e formalismo para
descrição das estruturas gramaticais
de uma língua)

Gramática computacional

- Gramática de “papel e lápis”

“FRASE é um enunciado de sentido completo, a unidade mínima de comunicação.” (Cunha; Cintra, 1985, p. 116)
- Gramática computacional
 - CFG
$$S \rightarrow NP VP$$
 - HPSG: signos



Por que gramática computacional?

- Do ponto de vista da linguística: possibilidade de verificar automaticamente a consistência interna e a plausibilidade empírica de um modelo da linguagem humana ou de uma análise específica de uma língua particular



Por que gramática computacional?

- Do ponto de vista da ciência da computação:
 - motivação teórica: a gramática, como parte da linguagem (natural), constitui uma faculdade da mente humana e, portanto, integra o domínio da inteligência artificial
 - aplicações: extração de informações, resolução de perguntas, tradução automática do tipo FAHQT etc.



Q&A

Watson da IBM

FERRUCI, D. et al. Building Watson: An Overview of the DeepQA Project. *AI Magazine*, v. 31, n. 3, 2010.

MCCORD, M. C.; MURDOCK, J. W.; BOGURAEV, B. K. Deep parsing in Watson. *IBM Journal of Research and Development*, Armonk, v. 56, n. 3/4, p. 1-15, 2012.



Extração de informações

NOVICHKOVA, S; EGOROV, S.; DARASELIA, N. MedScan: a natural language processing engine for MEDLINE abstracts.
Bioinformatics, Oxford, v. 19, n. 13, p. 1699-1706, 2003.



Tradução automática

- Sistema KANT da Carnegie Mellon University (LFG)
- Sistema VERBMOBIL (HPSG)
- Sistema MOLTO (Grammatical Framework)

Engenharia da gramática

- **Abordagem baseada em conhecimento:** elaboração manual de uma gramática num formalismo computacionalmente implementado → *parser* → análise sintática automática (*parsing*)
- **Abordagem baseada em dados:** manual de anotação (gramática) → construção de um *treebank* → algoritmo de aprendizagem de máquina → modelo estatístico → *parsing*



Indicações de leitura

DUCHIER, D.; PARMENTIER, Y. High-level Methodologies for Grammar Engineering, Introduction to the Special Issue. *Journal of Language Modelling*, Warszawa, Poland, v. 3, n. 1, p. 5-19, 2015.

DOI:

<https://doi.org/10.15398/jlm.v3i1.117>

Indicações de leitura

- BENDER, E. M. Grammar Engineering for Linguistic Hypothesis Testing. In: GAYLORD, N. et al. (Org.). *The Proceedings of the Texas Linguistics Society 10: Computational Linguistics for Less-Studied Languages*. Stanford: CSLI, 2008. p. 16-36.



3º encontro

26 de abril de 2021


Recapitulação

- Versão 03 da minigramática do inglês (sen = *simple English*, arquivo [choices03.txt](#))
 - Fenômenos implementados:
 - Concordância sujeito-verbo na 3ª pessoa
 - Sentença simples com verbos intransitivos e verbos transitivos
 - Flexão nominal regular (plural)

Testfile output [out03.txt](#)

Redundâncias na codificação

- A atual hierarquia verbal mistura valência (propriedade dos radicais) com propriedades expressas pelas flexões verbais, como tempo, modo, pessoa, número etc.
- Para cada forma verbal, são necessárias 15 escolhas.
- Por exemplo, repete-se a especificação de -s em *chases* e *sleeps*
- Desse modo, a codificação do léxico verbal torna-se extremamente cansativa.

- 
- **Agrupamento de propriedades associadas em tipos**
 - Todo verbo no modo indicativo ou subjuntivo é finito.
 - A codificação da pessoa e número é independente da codificação do tempo e do modo (especialmente relevante para línguas com flexão verbal mais rica como latim ou português)

Classes posicionais

- Decomposição das formas flexionadas em classes posicionais

radical (stem)	vogal temática	tempo- modo	pessoa- número
cant	á	va	mos
part	í	a	mos

Implementação na Matriz Gramatical

- Cada classe posicional recebe um determinado *input* e comporta um determinado número de regras lexicais.
- A classe pode ser prefixal ou sufixal.
- Regras lexicais operam sobre o *input*, acrescentando algum material ortográfico (segmentos) e propriedades (*features* 'traços').
- Exemplo em [choices11d.txt](#)

Modelação prévia

- Desenhar fluxograma:
 - ClassePosicional 1 => ClassePosicional 2
=> ... => ClassePosicional n

Verbos auxiliares

- Dados
 - the dog **is** following the cat
 - the dog **has** followed the cat
 - the dog **will** follow the cat
- Cada auxiliar exige uma forma própria do verbo principal
- Implementação na Matriz Gramatical: **Word Order** e **Lexicon/Auxiliary** types no questionário

Alçamento (*raising*)

- Movimento do sujeito do verbo principal para sujeito do auxiliar

is (VP the dog following the cat) =>
the dog is (VP following the cat)

Cf. verbete *raising* na Wikipedia.

Auxiliar: com ou sem predicado?

- Os verbos *be*, *have* e *will*, como auxiliares do presente contínuo (progressivo), presente perfeito e do futuro, contribuem para a sentença apenas com especificações de tempo, modo, aspecto, pessoa e número (POULSON, 2011).
- Outros auxiliares, como os modais, possuem uma especificação semântica própria, devendo receber, na codificação, um predicado.

Implementação do auxiliar *be* progressivo

progr (aux1)

Auxiliary type 1:

Type name: progr

This auxiliary type contributes:

(X) No predicate.

() An independent predicate.

Verbos auxiliares modais

- Operador de modalidade \Box (ou M, do alemão *Möglichkeit* ‘possibilidade’):
 - the dog **can** bark
- Operador de modalidade \Diamond (ou N, do alemão *Notwendigkeit* ‘necessidade’):
 - the dog **must** sleep

Cf. lógica modal na [Encicl. Stanford de Fil.](#)

Implementação dos auxiliares modais

- Análoga à dos auxiliares *be*, *have* e *will*, exceto pela atribuição de um predicado aos modais

`possibility (aux4)`

Auxiliary type 4:

Type name: `possibility`

This auxiliary type contributes:

☐ No predicate.

☒ **An independent predicate.**

Hipergeração na versão choices11d

- *chase + ed = *chaseed* etc.

False positives

9 *the cat slept 1 21

14 *the dog chaseed the cat 1 34

[...]

Ver [results11d.txt](#)

Hipogeração na versão choices11c

- Formas verbais irregulares como *chased* e *slept* não são analisadas

False negatives

[...]

29 the dog has slept 0 12

31 the dog has chased the cat 0 20

Ver [results11d.txt](#)

Indicações de leitura

- ARONOFF, M. *Word formation in generative grammar*. Cambridge: MIT, 1976.
- DAVIS, A. R. *Linking by types in the hierarchical lexicon*. Stanford: CSLI, 2001.
- MONTEIRO, J. L. *Morfologia portuguesa*. 4. ed. Campinas: Pontes, 2017.

Indicações de leitura

- POULSON, L.
Meta-modeling of Tense and Aspect in a Cross-linguistic Grammar Engineering Platform.
UW Working Papers in Linguistics, v. 28, 2011.