

## **PARSER: UM ANALISADOR SINTÁTICO E SEMÂNTICO PARA SENTENÇAS DO PORTUGUÊS**

João Luís Garcia ROSA  
(PUCCAMP)

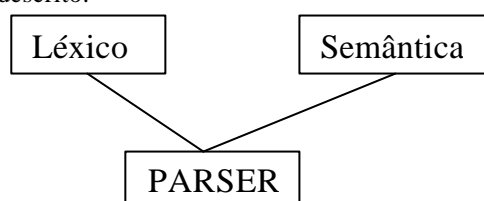
*ABSTRACT: This work aims at the analysis and implementation of a semantic parser for oral portuguese sentences. Its approach is based in word thematic role networks. A subset of Portuguese grammar is implemented in Prolog, a programming language based on predicate logic. The system includes morphological analysis, syntactic analysis, and semantic analysis that generates a logical form of the input sentence.*

**KEY WORDS:** parser, semantic analysis, logical form, natural language processing.

### **0 - INTRODUÇÃO**

O objetivo deste projeto é propor a análise e implementação de um “parser semântico” para sentenças da língua portuguesa oral, utilizando uma abordagem baseada nas redes de papéis temáticos das palavras.

Está em desenvolvimento no Instituto de Estudos da Linguagem da UNICAMP, um projeto de um  *sintetizador de fala*  em conjunto com a Faculdade de Engenharia Elétrica da UNICAMP. A idéia deste projeto é gerar a linguagem oral, a partir de um texto escrito. É necessário um  *parser*  que transformará o texto escrito, num determinado padrão, com marcas hierárquicas, que alimentará um outro módulo, que se encarregará do restante da análise (análise fonológica, etc.). Como parte do meu projeto de tese de Doutorado, devo participar deste  *parser* , que deverá estabelecer fronteiras sintáticas (relações entre as palavras) e informações relativas às relevâncias das palavras, etc. O  *parser*  pode ser assim descrito:



O  *léxico*  deve conter o vocabulário usado pelo  *parser* . Deve conter elementos de análise lingüística, estrutura e inter-relacionamento das palavras. A  *semântica*  deve incluir a

forma lógica, etc., mecanismos para permitir a inferência lógica. Em relação ao *parser* propriamente dito, várias abordagens podem ser consideradas.

## 1 - IMPLEMENTAÇÃO

A implementação computacional de um fragmento de gramática foi feita através da linguagem de programação lógica Prolog (do francês, *Programmation en Logique*), onde os comandos da linguagem são transposições de fórmulas da lógica de predicados de primeira ordem. Justamente por este fato preferiu-se implementar uma gramática livre de contexto, baseada em eventos e papéis temáticos. Este analisador semântico fornece várias formas lógicas possíveis de sentenças declarativas da língua portuguesa, analisando o determinante usado e fornecendo o quantificador para o mesmo. A gramática inclui a conjunção de sentenças, possibilitando a utilização de anáforas (pronomes pessoais) na segunda sentença da conjunção. Inclui também sintagmas preposicionais, através das chamadas *com-SN*, isto é, um sintagma nominal iniciando com a palavra *com*. Isto permite com que o analisador forneça duas leituras para sentenças como

*O homem abateu o menino com a maleta*

obtendo-se uma forma lógica onde *maleta* é o instrumento de *abateu* e também a forma onde *maleta* é objeto de posse de *menino*.

O programa inclui também, uma análise morfológica, que classifica cada verbo regular, em todos os tempos verbais e cada substantivo, adjetivo, etc. em gênero e número. Alguns verbos irregulares foram incluídos, como *ser*, *estar*, etc. Deve-se considerar, entretanto, que se trata apenas de uma primeira versão de um *parser*, que certamente servirá de base para o projeto maior. Várias considerações lingüísticas terão de ser revistas.

O sistema inclui diversos advérbios, segundo a classificação de Ilari et alli (1990), entre os quais os predicativos qualitativos e intensificadores, os sentenciais (modalizadores e aspectualizadores) e os não-predicativos (negação). Como a análise fornece a forma lógica baseada em eventos, optou-se por declarar como *adjunto* o advérbio que acompanha o verbo, o nome ou o adjetivo.

Na linguagem Prolog, não é trivial, talvez nem mesmo possível, a implementação de predicação de segunda ordem. Optou-se por usar a notação baseada em eventos, onde o verbo é o elemento principal da sentença que atribui papéis temáticos aos outros elementos que o acompanham.

A implementação consiste de um léxico pequeno, onde se define apenas as formas singulares das palavras (a análise morfológica se encarregará de descobrir o número). Consiste também das terminações dos verbos regulares, para todos os tempos do modo indicativo. Foram incluídos também alguns verbos irregulares. O programa dispõe de uma tela de informação sobre o léxico armazenado. A análise se baseia numa gramática

sintagmática, a Gramática de Cláusulas Definidas, de Pereira e Warren (1980). Caso a sentença seja agramatical, ou caso o sistema não possua previsão de análise, o *parser* rejeita a sentença.

*RESUMO: Este trabalho propõe a análise e implementação de um parser semântico para sentenças da língua portuguesa oral, utilizando uma abordagem baseada nas redes de papéis temáticos das palavras. Foi implementado um fragmento de gramática em Prolog, uma linguagem de programação baseada em lógica. O sistema inclui uma análise morfológica, uma análise sintática e uma análise semântica que gera uma forma lógica para a sentença de entrada.*

**PALAVRAS-CHAVE:** parser, análise semântica, forma lógica, processamento de linguagem natural

## ANEXO

O programa , quando executado, fornece a seguinte janela de entrada:

```
*****
*   PARSE   *
*****

Análise Morfológica
Análise Sintática
Geração de Forma Lógica

*****
* João Luís Garcia Rosa *
*****

Deseja informações sobre o léxico (i) ou entrar com uma frase? (f):
```

Caso se entre com um “i” (informações), o sistema fornece o seguinte:

## PARSER BASEADO EM EVENTOS E PAPÉIS TEMÁTICOS

Sentença declarativa, cláusula relativa, sintagma preposicional e anáfora.

Léxico disponível:

Determinantes:	um, uma, algum, alguma, o, a, todo, toda
Substantivos:	homem, menino, menina, mulher, bola, maleta, faca, gente.
Nomes Próprios:	joão, maria
Adjetivos:	bonito, bonita, posta.
Advérbios:	bem, muito, mais, demais, meio, não, felizmente, normalmente, realmente, humanamente.
Verbos Transitivos:	abater, matar.
Verbos Intransitivos:	viver, partir.
Verbos de Cópula:	ser, estar.
Pronomes Pessoais:	
Outras:	que, e, com.

Quer terminar? (s/n)

Se o usuário digitar “n”, o sistema volta a pedir a entrada de uma frase (“f”). As entradas do usuário estão em itálico, para diferenciar das saídas do sistema.

Deseja informações sobre o léxico (i) ou entrar com uma frase? (f): *f*

Digite a frase para a geração da(s) forma(s) lógica(s) (entre “aspas”):  
*“O menino bonito matou as mulheres”.*

Forma(s) Lógica(s):

existe(x): (menino(x) & bonito(x)) & qualquer(y): mulher(y) -> existe(e,pret\_perf):  
matar(e) & agente(e,x) & tema(e,y)

Quer terminar? (s/n)

Caso se digite uma sentença ambígua, o sistema dá as leituras possíveis (uma de cada vez). Suponha a entrada de uma sentença, na qual *maleta* pode ser instrumento da ação de *abater* ou posse de *menina*:

Deseja informações sobre o léxico (i) ou entrar com uma frase? (f): *f*

Digite a frase para a geração da(s) forma(s) lógica(s) (entre “aspas”):  
“*O menino abateu a menina com a maleta*”.

Forma(s) Lógica(s):

existe(x): (menino(x) & existe (z): maleta(z) & existe(y): menina(y) & existe (e,pret\_perf): abater(e) & agente(e,x) & tema(e,y) & instr(e,z)

Quer terminar? (s/n) *n*

existe(x): (menino(x) & existe (z): maleta(z) & existe(y): menina(y) & existe (e,pret\_perf): abater(e) & agente(e,x) & tema(e,y) & possui(y,z)

Quer terminar? (s/n) *n*

Deseja informações sobre o léxico (i) ou entrar com uma frase? (f):

Uma outra sentença ambígua, na qual a palavra *faca* também pode ser interpretada como instrumento ou posse:

Deseja informações sobre o léxico (i) ou entrar com uma frase? (f): *f*

Digite a frase para a geração da(s) forma(s) lógica(s) (entre “aspas”):  
“*João matou a mulher bonita com a faca*”.

Forma(s) Lógica(s):

existe(x): *joão(x)* & existe(z): *faca(z)* & existe(y): (*mulher(y)* & *bonita(y)*) &  
existe(e,pret\_perf): *matar(e)* & agente(e,x) & tema(e,y) & instr(e,z)

Quer terminar? (s/n) *n*

existe(x): *joão(x)* & existe(z): *faca(z)* & existe(y): (*mulher(y)* & *bonita(y)*) &  
existe(e,pret\_perf): *matar(e)* & agente(e,x) & tema(e,y) & possui(y,z)

Quer terminar? (s/n) *n*

Deseja informações sobre o léxico (i) ou entrar com uma frase? (f):

---

#### REFERÊNCIAS BIBLIOGRÁFICAS

ILARI, R. et alii (1990). Considerações sobre a Posição dos Advérbios. In *Gramática do Português Falado - Volume I: A Ordem*. Editora da Unicamp/Fapesp.

PEREIRA, F. C. N. & WARREN, D. H. D. (1980). Definite Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison with Augmented Transition Networks. *Artificial Intelligence* **13**, 231-278.