

Semântica: Uma Tentativa de Tornar as Estruturas com Predicado Secundário Acessíveis aos Sistemas de PLN

Gabriela Betania Hinrichs Conteratto¹, Rove Luiza de Oliveira Chishman²

1 Pontifícia Universidade Católica do Rio Grande do Sul. Programa de Pós-Graduação em Linguística Aplicada – Porto Alegre, RS – Brasil.

2- Universidade do Vale do Rio dos Sinos. Programa de Pós-Graduação em Linguística Aplicada – São Leopoldo, RS - Brasil.

ghinrichs@uol.com.br, rove.chishman@gmail.com

Abstract. *The main goal of this paper is a descriptive explanatory study of the secondary predication of the linguistic phenomenon, considering its utility for the improvement of the computational systems that need to process natural language. To accomplish this task, the main motivation was the limitation of the parser PALAVRAS in analyzing the structure with secondary predicate.*

Resumo. *A meta principal deste trabalho é a realização de um estudo descritivo explanatório do fenômeno linguístico da predicação secundária, tendo em vista a sua utilidade para o aperfeiçoamento de sistemas computacionais que necessitam processar a linguagem natural. Para a realização de tal meta, teve-se como motivação principal as limitações do parser PALAVRAS em analisar as estruturas com predicado secundário.*

1. Introdução

A meta central deste trabalho é apresentar um estudo descritivo explanatório das estruturas com predicado secundário descritivo voltado para o sujeito, tendo em vista a sua utilidade para o aperfeiçoamento de sistemas computacionais que necessitam processar a linguagem natural. Para a realização de tal meta, teve-se como motivação principal as limitações do *parser* PALAVRAS em analisar as estruturas desse tipo.

Vale lembrar que o presente artigo é um recorte de uma pesquisa mais ampla, que tratou dos três tipos de estruturas com predicado secundário: descritivo voltado para o sujeito, descritivo voltado para o objeto e resultativo. Contudo, as estruturas com predicado secundário descritivo voltado para o sujeito abrangem um número maior de casos na Língua Portuguesa do Brasil, conforme verificado através do estudo de corpus naquela pesquisa, e por isso se optou por tratar aqui apenas dessas construções. Por estruturas com predicado secundário descritivo voltado para o sujeito, está se entendendo as estruturas como em (1).

- (1) a- O tricolor entrou em campo **assustado**.
- b- Ele fumou um cigarro **pensativo**.

O artigo está organizado como segue: a seção 2 apresenta as limitações do *parser* PALAVRAS em lidar com as estruturas com predicado secundário descritivo

voltado para o sujeito; a seção 3 descreve a semântica dessas estruturas a partir de um estudo de *corpus*. Os resultados obtidos serão apresentados na seção 4. Por último, realizam-se as considerações finais.

2. Limitações do *parser* PALAVRAS

O analisador automático PALAVRAS foi desenvolvido por Eckhard Bick no contexto de um projeto de doutoramento (1994-2000), na Universidade de Århus (Dinamarca), com a finalidade de analisar estruturas em Português. O sistema do analisador PALAVRAS, em termos de representação formal, utiliza a Gramática de Dependência e baseia-se no formalismo Gramática de Restrições (de Helsinki) para lidar com ambigüidades morfológicas e sintáticas.

A análise sintática é realizada com a ajuda de etiquetas morfológicas e sintáticas aplicadas ao texto de entrada pelo analisador e etiquetador morfológico. A cada palavra do texto em análise é associada uma descrição sintática que inclui a classe morfológica (indica se é um substantivo, um advérbio, etc.) e uma forma sintática, representada essa por marcadores de dependência, que indicam o núcleo da unidade. A etiqueta usada pelo *parser* para anotar o predicativo do sujeito é a @SC. As etiquetas @PRED e @FCs são usadas pelo analisador quando o predicativo do sujeito está distante do seu hospedeiro.

Vale lembrar que o símbolo @ é utilizado para introduzir as etiquetas de função sintática, e os marcadores (< , >) indicam a direção do núcleo sintático de que os constituintes são dependentes, com exceção do verbo principal, que não exibe marcadores dependenciais. Para diagnosticar quais as limitações do PALAVRAS em analisar estruturas com predicado descritivo voltado para o sujeito, foram realizados alguns testes, ilustrados abaixo. Na figura (1), tem-se a etiqueta “PROP” indicando que o sujeito é constituído de um nome próprio (masculino/singular). O verbo principal da estrutura — que está na terceira pessoa do singular do pretérito perfeito do modo indicativo — é indicado pelo símbolo <fmc>, e o predicado descritivo secundário possui a etiqueta @SC.

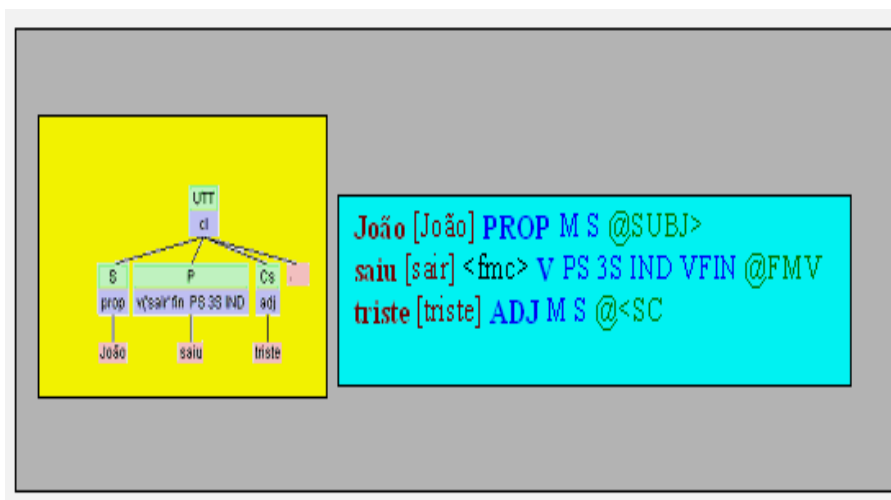


Figura 1. Teste (1) – predicado secundário

A fim de se verificar como o PALAVRAS analisa sentenças com constituintes descontínuos, acrescentou-se o complemento circunstancial, *de casa*, no VP da estrutura anteriormente examinada e solicitou-se uma nova análise. Como se pode ver na figura (2), o resultado foi inadequado.

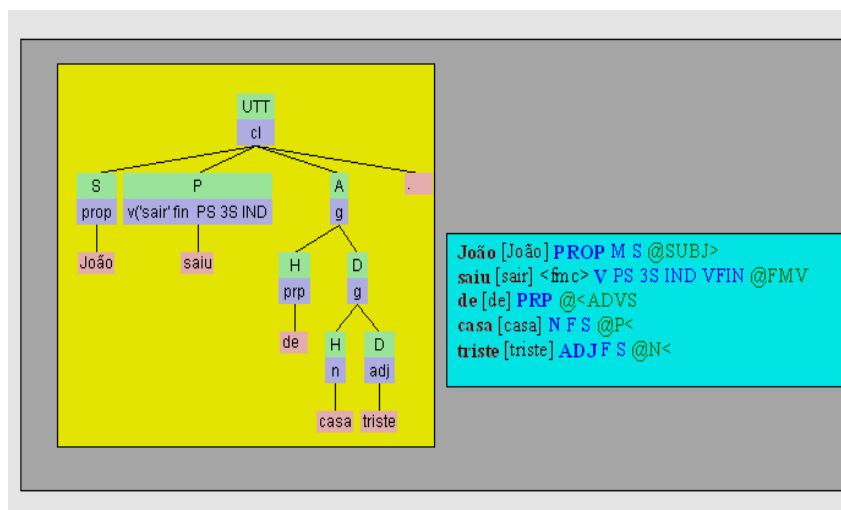


Figura 2. Teste (2) – predicado secundário

Observa-se que o analisador não conseguiu relacionar *triste* com o sujeito, ligando-o equivocadamente com o núcleo nominal mais próximo, *casa*. Além disso, ele classifica o adjetivo *triste* como um simples modificador, em vez de reconhecê-lo como um predicativo do sujeito. Isso mostra que, quanto mais longe o predicado secundário estiver do seu hospedeiro, maior será a dificuldade para localizá-lo e reconhecê-lo. A fim de comprovar isso, realizou-se mais um teste. Solicitou-se a análise de estruturas do tipo como: *João fumou um cigarro nervoso*. Novamente, como se pode observar na figura (3), o resultado não foi adequado.

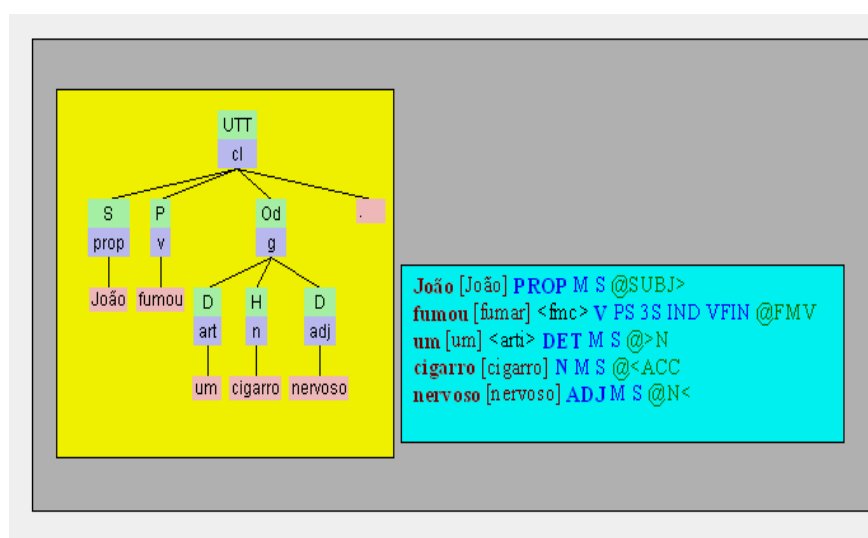


Figura 3. Teste (3) - predicado secundário

Nessa ocorrência, o *parser* realizou a mesma operação do exemplo anterior: ligou o adjetivo ao núcleo nome mais próximo e o classificou como seu modificador. É importante salientar que processar a linguagem natural envolve tanto questões de arquitetura de sistemas mais ligadas à área da Computação quanto questões de linguagem ligadas à descrição lingüística. Aqui, concentram-se os esforços em mostrar que um dos caminhos necessários para tornar essas estruturas com predicação secundária acessíveis à máquina é realizar um estudo na interface sintático-semântica.

3. A semântica das estruturas com predicado secundário

A semântica está ganhando especial atenção no âmbito do PLN, pois tem lançado luzes sobre algumas questões problemáticas. Acredita-se que, no caso específico das construções com predicação secundária, um estudo voltado para suas particularidades semânticas pode auxiliar na resolução dos problemas encontrados no seu processamento. Essa hipótese se fortalece com a afirmação de Franchi (2003) de que a predicação secundária é a verdadeira *predicação semântica*, pois, para se compreenderem as estruturas com predicados secundários, é preciso recorrer a um estudo semântico.

Aqui, partindo do pressuposto de que a informação semântica dessas estruturas consiste em um certo número de predicados e argumentos, cujas combinações produzem diferentes sentidos, assume-se que somente uma semântica enriquecida será capaz de capturar as informações relevantes para a interpretação dessas sentenças. De acordo com Franchi (2003), em estruturas com predicado secundário, um argumento recebe, composicionalmente, um papel temático que resulta de uma dupla predicação. Por exemplo, em (2), ambos os predicados, *sair* e *furiosa*, atribuem diferentes papéis temáticos a Érol: agente e experienciador.

(2) Érol saiu do quarto **furiosa**.

O autor chama a atenção para o fato de que essa atribuição de papel temático do predicado secundário a um argumento do predicado primário ocorre a partir de uma maior compatibilidade entre as propriedades semânticas já acarretadas na predicação primária desse ou daquele argumento. Assim, é possível compreender algumas restrições das construções com predicado secundário a partir da atribuição de papéis temáticos, uma vez que o que tem estatuto teórico, nessa abordagem, são as propriedades semânticas acarretadas pela relação dos predicadores e seus argumentos. Dessa forma, em uma estrutura do tipo como em (3), a compatibilidade das propriedades semânticas entre o predicado secundário e seu hospedeiro é que evidenciaria essa relação.

(3) Ele fumava um cigarro **pensativo**.

O predicado secundário *pensativo*, na estrutura demonstrada em (3), expressa um estado psicológico de um ser animado, de forma que, necessariamente, mantém relação com um argumento que possua a característica de ser animado, que nesse caso é o SN (sujeito). Isso fica melhor esboçado na figura (4).

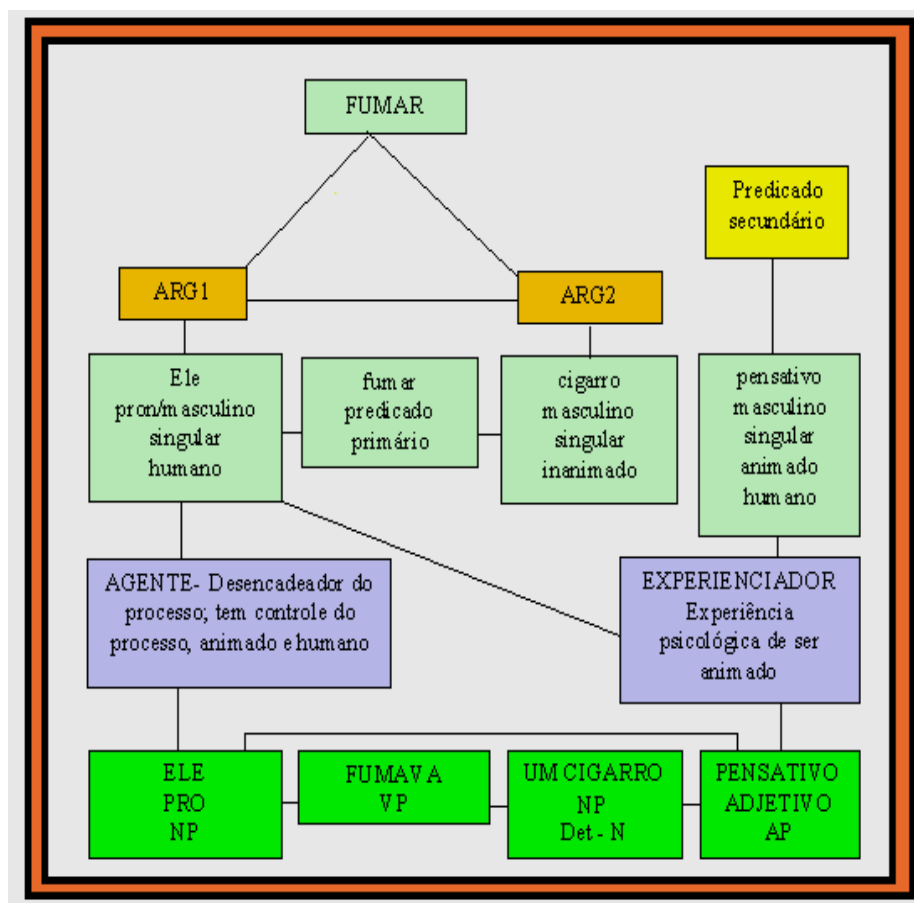


Figura 4. Características sintático-semânticas

Na segunda linha, tem-se a relação do predicado primário e seus argumentos: *fumar* (x,y), ou ainda, a configuração do predicator *fumar* envolve um argumento externo e um argumento interno. Em seguida, acrescentam-se aos argumentos alguns traços semânticos a fim de restringir ligações equivocadas entre os argumentos e seus possíveis predicadores. Na terceira camada, mostra-se que o argumento externo recebe composicionalmente dois papéis temáticos provenientes de duas predicções: *Ele* recebe o papel temático de agente do predicado primário e de experienciador do predicado secundário.

É importante mencionar que o estudo aqui apresentado foi realizado em um *corpus* constituído por quatrocentos e trinta e duas estruturas com predicado secundário descritivo voltado para o sujeito extraídas do Corpus do NILC (Núcleo Interinstitucional de Linguística Computacional). De acordo com esse corpus, 98% das estruturas analisadas com constituintes descontínuos apresentam compatibilidade semântica das propriedades semânticas somente entre o AP e o NP sujeito, como exemplificado em (4).

- (4) a- Imaginava que Capitu saísse da janela **assustada**.
- b- O jogador deixou o campo **assustado**.
- c- Itamar desceu do carro **irritado**.
- d- O monge cerrou os punhos **irritado**.

Assim, pode-se dizer que os dados coletados corroboram a intuição de que um estudo semântico pode tornar as estruturas com predicado secundário acessíveis aos sistemas de PLN, pois as propriedades semânticas dos predicadores funcionam como restrições seletivas.

4. Implementação

A fim de ilustrar uma implementação das informações lingüísticas da estrutura anteriormente discutida *Ele fumava um cigarro pensativo*, realiza-se um exercício na linguagem de programação Prolog, mostrando-se a importância de se ter um léxico informativo e organizado. Para implementar o léxico, utiliza-se o formato de arquivo texto, sendo as entradas especificadas de acordo com a sintaxe do Prolog. É importante mencionar que as regras sintagmáticas descrevem o modo como as categorias sintagmáticas são formadas pela combinação de outras categorias sintagmáticas ou lexicais. As regras lexicais são responsáveis por introduzir as categorias lexicais, que são as correspondentes às unidades básicas.

Na gramática, tem-se um conjunto de regras que permite fornecer as possibilidades de associação das palavras entre si, segundo as suas categorias lexicais (aspectos sintáticos), e estabelecer as relações semânticas entre os componentes da frase, resolvendo alguns problemas — como o de constituintes descontínuos —, pois os componentes semânticos dos predicadores devem funcionar como restrições seletivas, ou melhor, como traços componenciais que selecionam os papéis argumentais compatíveis com as características combinatórias da predicação. Na tabela (1), têm-se as regras sintáticas e as regras terminais (léxico) das estruturas em pauta.

Tabela 1. Prolog — Regras sintáticas

Ele fumava um cigarro pensativo. REGRAS SINTÁTICAS S-->np, vp. np-->pro. vp-->v, np, ap. np-->det, n. ap-->adj LÉXICO - REGRAS TERMINAIS n-->[ele]. v-->[fumava]. det-->[um]. n-->[cigarro]. adj-->[pensativo].

Como pôde ser visto em (1), a construção de regras gramaticais é bastante simples, conforme a notação usual do Prolog. Entretanto, o fato é que as características sintáticas não são suficientes para interpretar estruturas com predicado secundário, como já se tentou mostrar no decorrer desta investigação. Acredita-se que uma solução seria implementar algumas particularidades semânticas já discutidas dessas construções — como papéis temáticos (Pt) e propriedades/traços semânticas (Ps) — que funcionariam como uma espécie de "filtro". Esse procedimento se aproxima da idéia de Marrafa (1993) com o recurso de *unificação de traços*. Para isso, usa-se o DCG (*Definite Clause Grammar*), uma ferramenta específica de modelagem lingüística incorporada às diversas versões do Prolog, que permite a representação de informações

gramaticais, sintáticas e semânticas e a inserção de regras de concordância, como está ilustrado em (2).

Tabela 2: Prolog — recurso — (DCG)

REGRAS SINTÁTICAS
s-->np(Pt,Ps),vp(Pr1,Ps1).
np(Pt,Ps)--> pro(Pt,Ps)
vp(Pr1,Ps1)-->v(Pr1,Ps1),np(Pt,Ps3),ap(Pr2,Ps).
np(Pt,Ps3)--> det,np(Pt,Ps3).
ap(Pr2,Ps)-->adj(Pr2,Ps).
LÉXICO ENRIQUECIDO - REGRAS TERMINAIS
pro(agente_humano,experienciador)-->[ele].
v(Prepri,ac)-->[fumava].
n(loc,ina)-->[cigarro].
det--> [um].
adj(prsen,humano)→[pensativo].

De acordo com esse recurso, a realização da análise de uma sentença passa a ser a prova de que ela obedece às regras de formação sintático-semânticas impostas por condições ou axiomas gramaticais. As regras sintáticas são definidas segundo uma hierarquia, na qual regras mais genéricas (que recebem como entrada estruturas sintáticas completas, ou mesmo partes de estruturas sintáticas passíveis de decomposição) são decompostas em regras mais específicas, até que a decomposição não seja mais possível ou necessária (quando uma regra manipula diretamente uma ou mais palavras da sentença). As regras mais gerais são as regras intermediárias e as regras mais específicas, que acessam diretamente o léxico, são as regras terminais. Quando esses dados são carregados num interpretador Prolog, tem-se o processamento da estrutura na figura (5).

1. ?['F:\gabriela\amostral.pl']
2. Yes
3. [trace] 2 ?-s([ele,fumava,um,cigarro,pensativo],[]).
4. Call: (7) s([ele, fumava, um, cigarro, pensativo], []) ? creep
5. Call: (8) np(_L203, _L204, _L205, [ele, fumava, um, cigarro, pensativo], _L187) ? creep
6. Call: (9) pro(_L203, _L204, _L205, [ele, fumava, um, cigarro, pensativo], _L187) ? creep
7. Exit: (9) pro(agente, humano, experienciador, [ele, fumava, um, cigarro, pensativo], [fumava, um, cigarro, pensativo]) ? creep
8. Exit: (8) np(agente, humano, experienciador, [ele, fumava, um, cigarro, pensativo], [fumava, um, cigarro, pensativo]) ? creep
9. Call: (8) vp(_L203, _L204, [fumava, um, cigarro, pensativo], []) ? creep
10. Call: (9) v(_L203, _L204, [fumava, um, cigarro, pensativo], _L207) ? creep
11. Exit: (9) v(_L203, acom, [fumava, um, cigarro, pensativo], [um, cigarro, pensativo]) ? creep
12. Call: (9) np(_L224, _L225, [um, cigarro, pensativo], _L208) ? creep
13. Call: (10) det([um, cigarro, pensativo], _L228) ? creep
14. Exit: (10) det([um, cigarro, pensativo], [cigarro, pensativo]) ? creep
15. Call: (10) n(_L224, _L225, [cigarro, pensativo], _L208) ? creep
16. Exit: (10) n(afetado, inanimado, [cigarro, pensativo], [pensativo]) ? creep
17. Exit: (9) np(afetado, inanimado, [um, cigarro, pensativo], [pensativo]) ? creep

18. Call: (9) ap(_L224, _L225, [pensativo], []) ? creep
 19. Call: (10) adj(_L224, _L225, [pensativo], []) ? creep
 20. Exit: (10) adj(pr2, humano, [pensativo], []) ? creep
 21. Exit: (9) ap(pr2, humano, [pensativo], []) ? creep
 22. Exit: (8) vp(_L203, acom, [fumava, um, cigarro, pensativo], []) ? creep
 23. Exit: (7) s([ele, fumava, um, cigarro, pensativo], []) ? creep

Figura 5: Histórico do Prolog

A fim de que a linguagem de programação possa verificar se as regras têm as condições necessárias para ser considerada uma sentença, lança-se na primeira linha o endereço do arquivo texto — *['F:\gabriela\amostral.pl']* —. Logo na segunda linha, verifica-se que a resposta é positiva. Aplica-se, então, o comando *trace* para visualizar o histórico da análise. Para melhor o fazer, propõe-se, também, um esquema na figura (6).

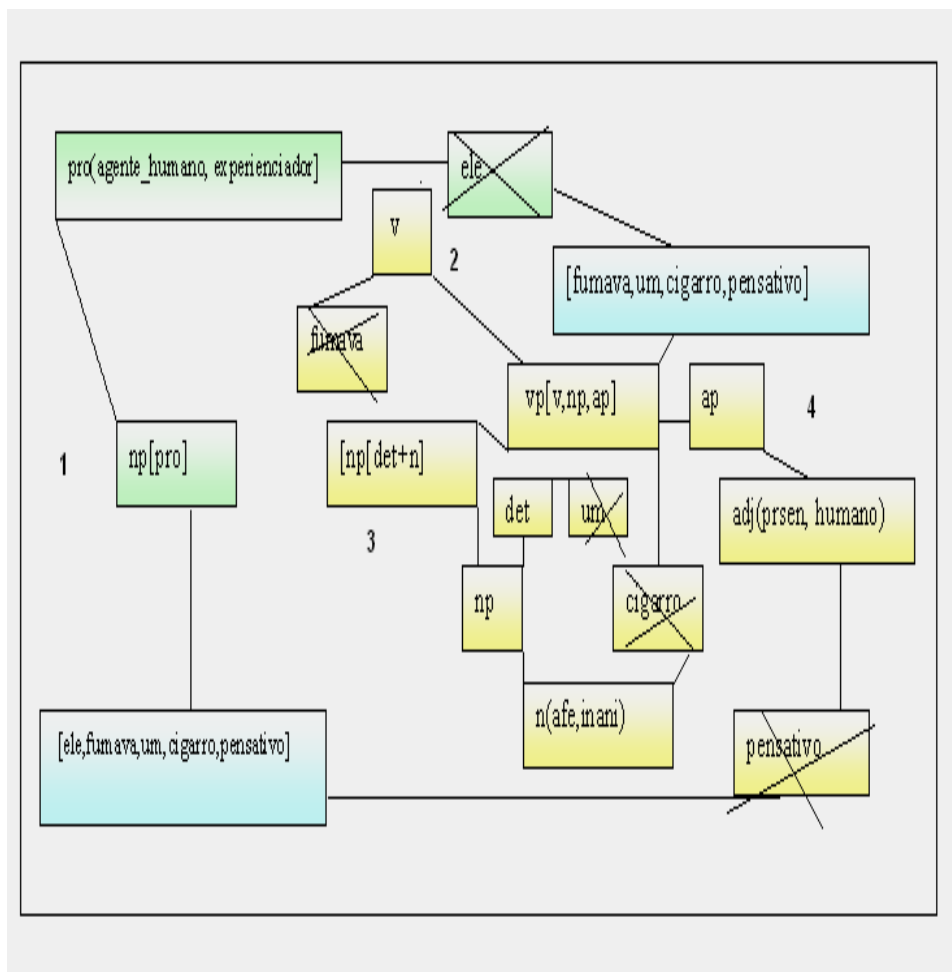


Figura 6: Esquema do histórico

A numeração indica os passos feitos para localizar cada constituinte. Cabe salientar que esses dois exercícios foram realizados para demonstrar o modo como a semântica pode contribuir para que os sistemas de PLN sejam mais eficientes. Ou

melhor, tenta-se evidenciar que os componentes semânticos dos predicadores irão funcionar como restrições de seleção. Ou, em outras palavras, busca-se ilustrar a maneira como a semântica dessas estruturas possibilita a resolução de certos entraves na área de PLN. Isso pôde ser verificado com a implementação de algumas das propriedades semânticas discutidas anteriormente no *parser* PALAVRAS, pois, como ilustrado em (7), a análise de estruturas como *Ele fumava um cigarro pensativo* foi realizada com sucesso após pequenas adequações.

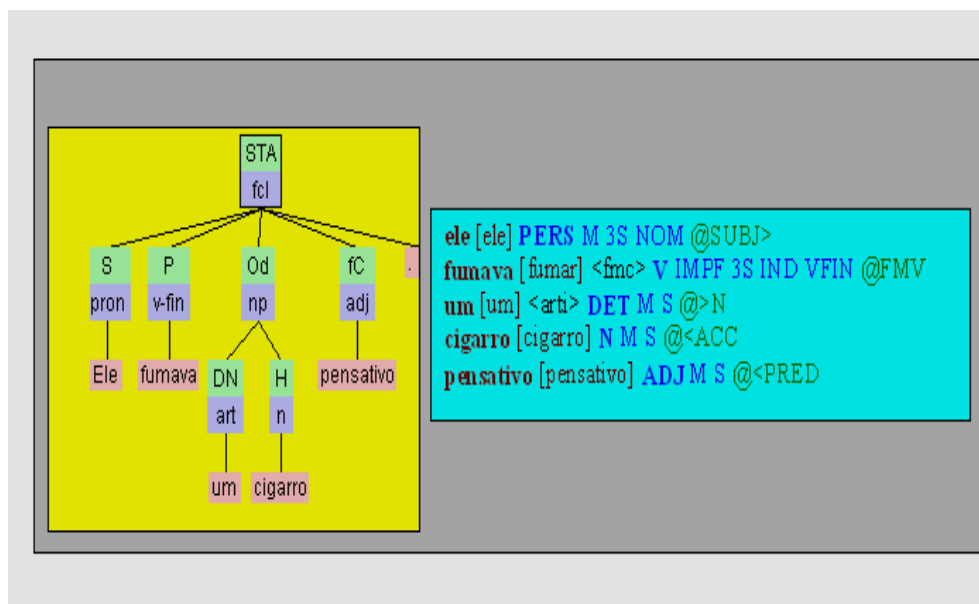


Figura 7. Teste (4) — predicado secundário

Os resultados mostram que os estudos semânticos podem ajudar a tornar as estruturas com predicado secundário acessíveis aos sistemas computacionais que necessitam processar a linguagem natural.

5. Considerações finais

Este trabalho apresenta um recorte de uma pesquisa mais ampla acerca das construções com predicado secundário. De qualquer forma, espera-se que se tenha conseguido evidenciar que um dos fatores que implica diretamente na eficiência de um sistema computacional é se ter uma descrição lingüística suficientemente informativa e organizada nele. Para isso, necessita-se de abordagens teóricas que visem não apenas a configurar os fenômenos da linguagem e sua resolução, mas também à eficiência necessária à sua inclusão em aplicações computacionais.

Referências

- BICK, Eckhard. (2001) *VISL – VISUAL INTERACTIVE SYNTAX LEARNING*. Odense, Dinamarca, University of Southern Denmark, Institute of Language and Communication, Disponível em: < <http://beta.visl.sdu.dk>>.
- CONTERATTO, Gabriela Betania. Hinrichs. (2005) *Predicação Secundária: Uma Contribuição da Lingüística ao Processamento Computacional da Linguagem*.

São Leopoldo, UNISINOS. Dissertação de Mestrado, Centro de Ciências da Comunicação, Universidade do Vale do Rio dos Sinos.

FRANCHI, C. (1987) Criatividade e gramática. *Trabalhos em Lingüística Aplicada*, Campinas, v. 9, p. 5-45.

FRANCHI, C. (2003) Teoria da adjunção: predicação e relações "temáticas". *Revista de Estudos da Linguagem*, Belo Horizonte, v.11, n.2, p.1-191.

MARRAFA, Palmira. (1993) *Predicação Secundária e Predicados Complexos em Português*. Lisboa: Universidade de Lisboa, Tese (Doutorado em Lingüística Portuguesa).