

ANOTAÇÃO SEMÂNTICO-DISCURSIVA DE VERBOS PARA O PROCESSAMENTO DE LINGUAGEM NATURAL: TRAÇOS ASPECTUAIS E MODAIS

Flávia Regina de Santana Evangelista
flystantana@hotmail.com
<http://lattes.cnpq.br/7903756471121467>

Rebeca Rodrigues de Santana
rebeca.rodrigues.17@gmail.com
<http://lattes.cnpq.br/7843680671771834>

Raquel Meister Ko Freitag
rkofreitag@uol.com.br
<http://lattes.cnpq.br/2582841591375626>

RESUMO

Atualmente, na área do processamento da linguagem, embora a anotação morfossintática tenha taxa de produtividade alta, ainda há categorias gramaticais, como os verbos, que necessitam de uma anotação linguística mais detalhada, dada a complexidade de valores semântico-discursivos envolvidos. Dando continuidade ao processo de anotação de verbos do português (FREITAG; CUNHA; EVANGELISTA; SANTANA, 2013), neste trabalho apresentamos os procedimentos de elaboração de uma árvore de decisão para a anotação das categorias de tempo, aspecto e modalidade verbais. Tal árvore possibilitará a anotação de corpus com etiquetas semânticas para verbos, contribuindo com o aprimoramento do processamento automático do português, visto que dotar uma ferramenta eletrônica de descrições linguísticas é capacitá-la a reconhecer a língua, de forma otimizada, apesar de suas especificidades.

Palavras-chave: Verbos; etiquetagem; processamento automático.

INTRODUÇÃO

A etiquetagem de *corpus* é uma atividade essencial para a automatização da linguagem humana. É a partir das descrições presentes nas *tags* (etiquetas) de um *corpus* de treino que se geram estatísticas e probabilidades para o aprendizado de *softwares* que reproduzem essa linguagem natural automaticamente. Atualmente, muitos linguistas estão envolvidos na área do processamento da linguagem e muito já se sabe sobre a anotação morfossintática, porém ainda há categorias, como os verbos, que necessitam de uma descrição linguística mais detalhada, devido ao nível de sua complexidade, o que dificulta o processamento, pois suas minúcias de significação não são captadas. Decorrente disso, as margens de erro em analisadores ainda são altas, e se devem,

principalmente, pela ausência descritiva e aplicada de traços que vão além do nível morfossintático, como os semântico-discursivos.

A anotação de traços semântico-discursivos de verbos é ainda uma abordagem pouco explorada, principalmente no que diz respeito aos traços aspectuais, que tem distinções tão utilizadas no português, sobretudo como recurso para tornar o enunciado mais expressivo (COSTA, 1990) e aos traços modais, pois as formas verbais além de apresentarem uma função dêitica de localização do evento, apresentam também uma função modal de indicação de realidade (CORÔA, 2005).

Dando continuidade ao projeto “Anotação semântico-discursiva de verbos para o processamento de línguas naturais” (FREITAG, 2012), que visa desenvolver etiquetas para verbos do português brasileiro contemplando os valores de tempo, aspecto e modalidade, após a definição do corpus mínimo para a anotação dos verbos (FREITAG; CUNHA; EVANGELISTA; SANTANA, 2013), neste trabalho apresentamos os procedimentos de elaboração de uma árvore de decisões com estas categorias semânticas dos verbos, focando os traços de aspecto e modalidade.

1 AS CATEGORIAS VERBAIS DE ASPECTO E MODALIDADE

Todas as línguas do mundo codificam, de alguma forma, as noções de tempo, aspecto e modalidade (BYBEE; PERKINGS; PAGLIUCA, 1994).

A categoria aspecto, para Comrie (apud COAN et al, 2006, p. 1365), “está relacionada aos diferentes modos de perceber a constituição temporal interna de uma situação”. Ou seja, a categoria aspecto está relacionada à percepção da duratividade ou não-duratividade da ação verbal.

Os valores aspectuais estão pautados nos estados de coisas permansivo, que não muda no decorrer do tempo, sendo incontestável, homogêneo e contínuo, e operativo, que muda no intervalo da estrutura temporal, podendo ser contado (COAN et al, 2006).

O estado operativo compreende situações qualitativas e quantitativas. As situações qualitativas são perfectivas e imperfectivas. “Uma situação perfectiva tem existência

quando do seu desfecho (COAN et al, 2006, p. 1369)". Já "uma situação imperfectiva tem sua existência quando inicia, independente de um ponto final no intervalo da estrutura temporal (COAN et al, 2006, p. 1369)". O aspecto indica duração, não duração, repetição e resultado e, quando indica duração, representa "os graus de desenvolvimento dessa duração, ou seja, se ela está no começo, no meio ou no fim" (CASTILHO; ELIAS, 2012, p. 161).

Os autores definem três representações do aspecto: o que dura (aspecto imperfectivo); o que começa e acaba (aspecto perfectivo); e o que se repete (aspecto iterativo).

Ainda segundo Castilho e Elias, tempo e aspecto apresentam ligação; as designações: "perfeito" e "imperfeito" do pretérito remetem aos aspectos perfectivo e imperfectivo, respectivamente. O pretérito perfeito, então, associa-se a ações passadas finalizadas e o pretérito imperfeito a ações passadas em andamento.

Outra categoria que contemplamos no desenvolvimento das etiquetas semântico-discursivas é a modalidade, que conforme Freitag (2011, p.1125), "costuma ser definida como a gramaticalização das atitudes do falante em relação ao conteúdo proposicional", ou seja, é caracterizada por manter a relação entre a atitude do falante e o conteúdo da sua declaração em um enunciado, sobre isso Freitag ainda afirma que "a noção de realidade/factuality remete à existência factual em algum tempo real (verdadeiro), ou a não-existência em um tempo real (falso) ou, ainda, à existência potencial em um tempo (possível)".

A modalidade também é composta por propriedades pragmático-discursivas, podendo ser dividida em epistêmica ou deôntica. A primeira corresponde aos valores de verdade, possibilidade, certeza, crença, evidência; enquanto a segunda corresponde aos valores de desejo, preferência, intenção, habilidade, obrigação, permissão, necessidade e manipulação, como afirma Coan et. alli (2006)

As diferentes atitudes dos falantes em relação ao conteúdo proposicional do enunciado podem levar a mudanças nos usos dos tempos verbais. Lyons (apud Coan et alli, 2006 p.170) "observa que o passado se combina com o modo de tal maneira que

introduz um matiz por vezes mais “incerto” ou mais “distante” ou mais “polido”, pontuando que a categoria tempo se converte em uma expressão secundária nesses casos”, ou seja, a partir disto podemos perceber que em determinados contextos a função modal exercida pelo verbo pode sobressair-se em relação a sua função dêitica de localização do evento.

Isso fica claro quando Coan (2003) nos mostra que quando se é empregado o mais-que-perfeito nem sempre é indicado um passado em relação a outro passado: pode ser atribuído à forma o valor de distância modal marcando menos assertividade em relação ao falante e o conteúdo da proposição, ou seja, a distância temporal na direção do passado é pressionada a expressar distância modal, o que reforça a atenção dada a esta categoria em nosso trabalho.

O detalhamento das informações acerca do aspecto e da modalidade verbal a partir do desenvolvimento de etiquetas para as categorias verbais contribuirá com o aprimoramento do processamento automático do português, possibilitando captar suas minúcias significativas, pois a anotação de textos com tais descrições linguísticas podem otimizar o desempenho de analisadores automáticos.

2 ELABORAÇÃO DA ÁRVORE DE DECISÕES

Para a definição das etiquetas semântico-discursivas, utilizamos uma ferramenta computacional de codificação linguística de texto, o Systemic Coder (<http://www.wagsoft.com/Coder/index.html>), desenvolvida por Mick O'Donnell. Tal ferramenta é constituída por um sistema no qual podemos segmentar os textos e gerar estatísticas descritivas do corpus a partir do resultado da segmentação. De acordo com a necessidade do programa, para iniciarmos a nossa anotação de dados padronizados da amostra “Textos Opinativos e Narrativos” (ARAUJO; PEIXOTO; FREITAG, 2012), pertencentes ao banco de dados Falares Sergipanos (FREITAG, 2013), elaboramos uma árvore de decisão contemplando as categorias semântico-discursivas dos verbos: o tempo, o aspecto e a modalidade (figura 1). A árvore de decisão funciona como uma rede sistêmica, na qual vamos acrescentando informações referentes à codificação das categorias verbais nos textos.

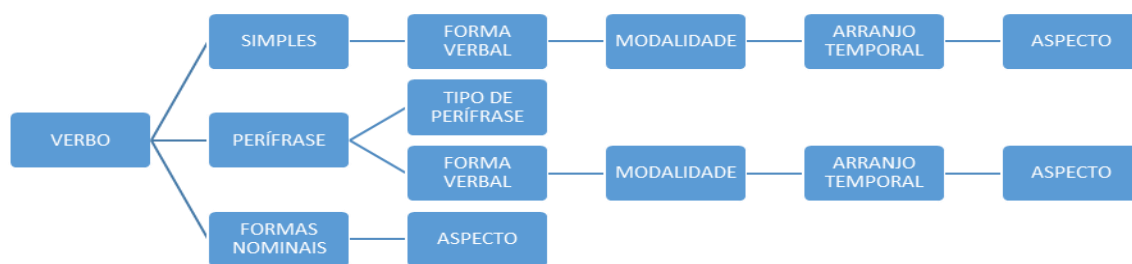


Figura 1: Ilustração da árvore de decisão para verbos

A primeira informação atribuída aos verbos etiquetados no modelo de árvore que elaboramos é a de que tipo de verbo se trata: simples, perifrástico ou um verbo nominal (figura 1), se tratando de verbo simples ou de perífrase, a árvore traz as opções de **formas verbais**, que são todas as do português, tabela 1.

Formas verbais
Presente do indicativo
Pretérito perfeito
Pretérito mais-que-perfeito
Futuro do presente
Futuro do pretérito
Presente do subjuntivo
Pretérito imperfeito do subjuntivo
Futuro do subjuntivo
Imperativo afirmativo
Imperativo negativo

Tabela 1: Formas verbais do português

Nas perífrases, analisamos o verbo auxiliar, atribuindo-lhe uma etiqueta de **forma verbal**, e o verbo principal recebe etiqueta em **tipo de perífrase**, que pode ser de infinitivo, particípio ou gerúndio, contemplando deste modo, as formas compostas. A figura 2 traz um recorte das opções de etiquetas para marcar a forma simples de pretérito perfeito da árvore de decisões.

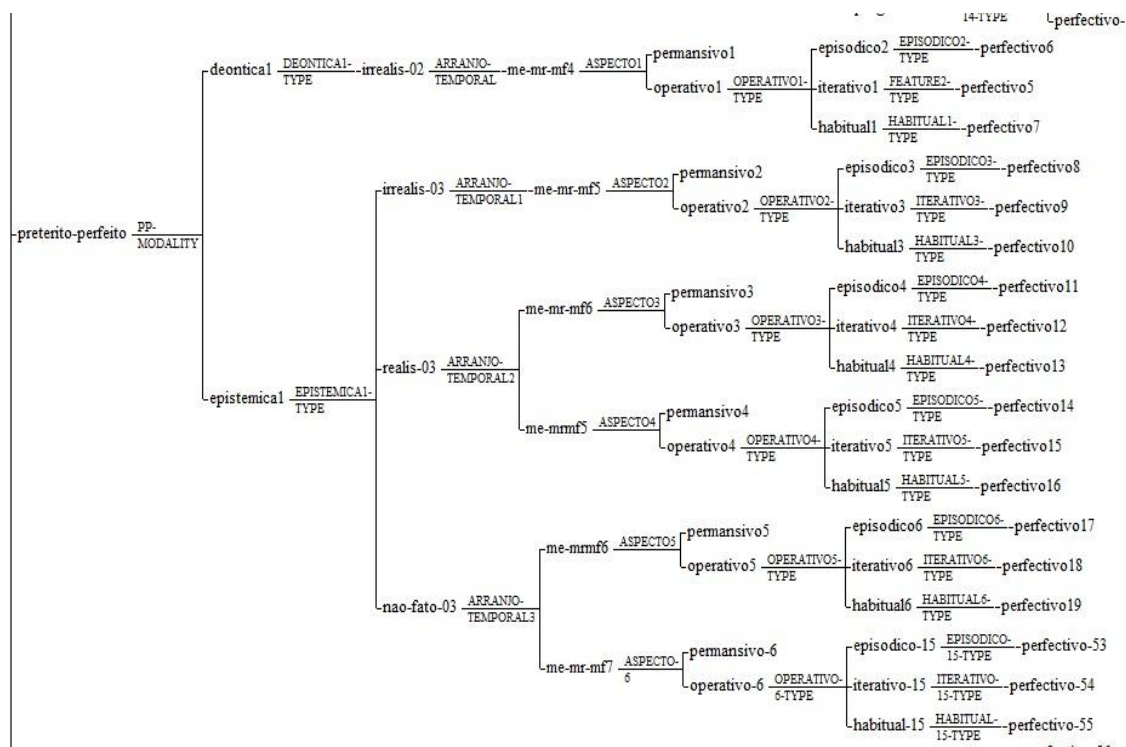


Figura 2: Recorte da árvore de decisões

Também foi necessário adicionar à árvore a etiqueta **formas nominais** para dar conta, por exemplo, de casos em que um verbo no infinitivo em sua forma finita vem após um que necessita, como complemento, de um objeto direto, o que podemos observar em (1); ou em casos de infinitivo flexionado.

(1) Tento *refletir* sempre sobre a situação da nossa querida cidade... (se ita fse 01)".

Em relação ao tempo das formas verbais, adicionamos etiquetas de **arranjo temporal**, que, conforme Corôa (2005), é formado por três momentos relevantes para a distinção dos *tempora*: o momento do evento (ME), momento em que se dá o evento; momento da fala (MF), momento em se faz a anúncio sobre o evento, e o momento da referência (MR): “o tempo da referência; o sistema temporal fixo com respeito ao qual se definem simultaneidade e anterioridade; é a perspectiva do tempo relevante que o falante transmite ao ouvinte, para a contemplação do ME” (CORÔA, 2005, p. 41).

No que se refere à modalidade, a árvore apresenta as possibilidades necessárias para classificar os verbos quanto às categorias deôntica ou epistêmica, em que a primeira se encontra no campo da obrigação, da proibição e permissão; enquanto a segunda está relacionada ao conhecimento e crença, e sua origem é usualmente o falante que a partir de evidências chega à conclusão se a proposição é verdadeira ou não (COSTA, 2009). Como exemplo, vejamos os fragmentos (2) e (3) em que as modalidades deôntica e epistêmica estão presentes nos verbos:

(2) As autoridades competentes e a administração pública *devem fazer* sua parte (se ita mse 18)

(3) *Acho* um absurdo e uma falta de respeito com a classe discente o sistema de cotas (se ita fse 17)

No fragmento (2), podemos perceber o valor de obrigação expresso pelo verbo *dever*, neste contexto, recaindo sobre o alvo deôntico, as autoridades competentes e a administração pública. No exemplo (3), o informante estabelece o seu ponto de vista em relação ao sistema de cotas, comprometendo-se com o enunciado de acordo com o ele acredita ser contrassenso.

Outro aspecto referente à modalidade existente na árvore é a equivalência das modalidades epistêmicas da tradição aristotélica no contexto comunicativo, com os tipos lógicos redefinidos, tabela 2, em que tanto a modalidade epistêmica quanto a deôntica compartilham da asserção irrealis.

Verdade factual > asserção realis	Proposição fortemente assertada como sendo verdadeira.
Verdade possível > asserção irrealis	Proposição assertada como possível, provável ou incerta.
Não-verdade > asserção negativa	Proposição assertada como sendo falsa

Tabela 2: Correlação realis/irrealis (cf. FREITAG; ARAUJO; BARRETO, 2013)

Com estas informações presentes na árvore de decisões, podemos etiquetar os verbos referente às suas modalidades nos enunciados, como o do exemplo que se segue:

(4) “A cidade dos ceboleiros *creceu* e as próprias autoridades já se deram conta disto” (se ita mse 18)

O verbo “*creceu*”, neste contexto, expressa um acontecimento em um tempo real, já ocorrido, pois, para o informante a cidade *creceu* de fato, ou seja, a sua proposição é assertada como verdadeira, sendo assim asserção *realis*.

Já com relação ao aspecto a árvore de decisão, primeiramente, apresenta as opções: **permansivo**, **operativo** e **resultativo**, imagem 3. Utilizamos a primeira para marcar verbos que expressam acontecimentos que não mudam no passar do tempo, como em (5). A segunda dispõe das seguintes ramificações de valores aspectuais quantitativos: **episódico**, **habitual**, **iterativo** e **contínuo**. O aspecto episódico é utilizado para marcar verbos com acontecimentos que não tiveram sua duratividade expressa, como em (6). Segundo Freitag (2010), o aspecto habitual denota uma situação sistematicamente repetida em diferentes ocasiões, presente, passado, ou ambos, e o aspecto iterativo marca uma situação que é repetida em ocasião específica, (7) e (8) representam, respectivamente, acontecimentos habitual e iterativo. Já o aspecto contínuo compreende verbos com situações em andamento, como em (9).

(5) Eu *aprendi* a lição... (fnp44)

(6) O ônibus *quebrou*... (se ita fle 11)

(7) Na construção da Escola Técnica de Itabaiana *sempre tem* um empecilho... (se ita mse 20)

(8) Deveremos fazer duas inscrições... (mop2)

(9) Eu estava falando com minha mãe... (fnt65)

Para a etiquetagem do aspecto contamos ainda com o resultativo que foca “o resultado da ação, e não nela própria (FREITAG, 2007). Após apresentar como opções os aspectos de critério quantitativo, a árvore traz as opções para a classificação quanto ao aspecto qualitativo, perfectivo e imperfectivo. O primeiro representa situações finalizadas e o segundo situações que dão ideia de andamento, (5) e (9) são situações perfectiva e imperfectiva, respectivamente.

CONCLUSÃO

A anotação de um *corpus* com etiquetas semântico-discursivas para verbos é de grande relevância para o processamento automático do português, pois dotar uma ferramenta eletrônica de descrições linguísticas é capacitá-la a reconhecer a língua de forma otimizada, apesar de suas especificidades. Tal atividade possibilita um aprimoramento no desempenho de analisadores automáticos, contribuindo para a captura das minúcias linguísticas das categorias verbais.

Nossa pesquisa vem colaborar no desenvolvimento e crescimento de trabalhos na área de PLN, especificamente no processamento de verbos do português, pois o desenvolvimento de trabalhos neste ramo contribui para a elevação da qualidade da produção científica e tecnológica que visa automatização da linguagem humana.

REFERÊNCIAS BIBLIOGRÁFICAS

ARAUJO, A. S.; PEIXOTO, J. C.; FREITAG, R. M. **Banco de Dados de Escrita - Textos Narrativos e Opinativos**. In: JORNADA DE PESQUISA CIENTÍFICA DO GEMPS/CNPq, 2, 2012, Aracaju, 2013. p.1-10.

BYBEE, J.; PERKINGS, R.; PAGLIUCA, W. **The evolution of grammar: tense, aspect, and modality in the language of the world**. Chicago: The University of Chicago Press, 1994.

CASTILHO, Ataliba T. de; ELIAS, Vanda Maria. **Categorias semânticas do verbo**. In: **Pequena gramática do português brasileiro**. São Paulo: Contexto, 2012.

COAN, M; BACK, A.; REIS; FREITAG, R. M. K. **As categorias tempo, aspecto, modalidade e referência: pressupostos teóricos para uma análise semântico-discursiva**. Estudos Linguísticos (São Paulo), v. XXXV, p. 1463 – 1472, 2006.

COAN, M. **Tempo, aspecto, modalidade e referência na significação dos pretéritos mais-que-perfeito e perfeito: correlações entre função(ões) e forma(s) em tempo real e aparente**. 2003. 232 f. Tese (Doutorado em Linguística) - Universidade Federal de Santa Catarina, Florianópolis. 2003.

COSTA, S. B. B. **O aspecto em português**. São Paulo: Contexto, 1990.

CORÔA, M. L. M. S. **O tempo nos verbos do português**. São Paulo: Parábola, 2005.

COSTA, S. Entre o deôntico e o epistêmico: o caráter camaleônico do verbo modal 'poder'. **Letra Magna**, p. 1-22, 2009.

FREITAG, R. M. K. O domínio funcional tempo-aspecto-modalidade na expressão do passado imperfeito no português falado no Brasil. **Revista do GEL**, v. 7, p. 139-170, 2010.

FREITAG, R. M. K. . A expressão do passado iminencial em português: formas e contextos de uso. In: **VII Congresso Internacional da Abralín**, 2011, Curitiba. Anais do VII Congresso Internacional da Abralín. Curitiba: UFPR, 2011. v. 1. p. 3654-3662.

FREITAG, R. M. K. **Anotação semântica de verbos para o processamento de línguas naturais**. Projeto de Pesquisa. Universidade Federal de Sergipe, São Cristóvão/SE, 2012.

FREITAG, R. M. K.; ARAUJO, A. S.; BARRETO, E. A. Emergência e regularização de usos em categorias verbais do português: gradações de modalidade nos valores condicional, iminencial e habitual no domínio do passado imperfeito. **Revista do GELNE**, v. 15, p. 99-122, 2013.

FREITAG, R. M. K. ; CUNHA, E. C. ; EVANGELISTA, F. S. R. ; SANTANA, R. R. Etiquetagem de verbos para o processamento automático do português brasileiro: procedimentos de constituição de corpus. In: **X Evidosol/VII Ciltec-online** – Anais do Encontro Virtual de Documentação em Software Livre e Congresso Internacional de Linguagem e Tecnologia Online. Belo Horizonte: Textolivre.org, 2013. v. 2. p. 1-5.

FREITAG, R. M. K. . Banco de dados falares sergipanos. **Working Papers em Linguística (Online)**, v. 14, p. 156-164, 2013.

SOBRE AS AUTORAS:

Flávia Regina de Santana Evangelista

Graduanda em Letras Vernáculas pela Universidade Federal de Sergipe, pesquisadora bolsista PIBITI/ UFS vinculada ao projeto Anotação semântica de verbos: corpus de treino da web, desenvolvendo o plano de trabalho Anotação semântica de verbos: textos jornalísticos e integrante do Grupo de Estudos em Linguagem, Interação e Sociedade - GELINS.

Rebeca Rodrigues de Santana

Graduanda do curso de Letras-Português pela Universidade Federal de Sergipe, pesquisadora de Iniciação Tecnológica no plano de trabalho Anotação semântica de verbos: textos acadêmicos e integrante do grupo de estudos em Linguagem Interação e Sociedade - GELINS.

Raquel Meister Ko Freitag

Professora do Departamento de Letras Vernáculas, do Programa de Pós-Graduação em Letras e do Programa de Pós-Graduação em Educação da Universidade Federal de Sergipe. Graduada em Letras, mestre e doutora em Linguística pela Universidade Federal de Santa Catarina, área de concentração Sociolinguística. Investigação de fenômenos de variação e mudança linguística com ênfase na dimensão estilística, considerando contatos, em comunidades de práticas, especialmente as marcadas por relações de gênero. Ênfase em questões metodológicas relacionadas à coleta e constituição de amostras sociolinguísticas (gênese da entrevista sociolinguística; comunidades de fala e comunidades de práticas; modelagem e dimensionamento de amostras) e na dimensão estilística da variação. Pesquisas sobre policy cycle de avaliação em larga escala em língua portuguesa (Provinha Brasil, Prova Brasil, ENEM).