

# O tratamento da partícula “se” para fins de anotação de papéis semânticos

Magali Sanches Duran, Sandra Maria Aluísio

Núcleo Interinstitucional de Linguística Computacional  
ICMC – Universidade de São Paulo - São Carlos – SP – Brasil  
magali.duran@uol.com.br, sandra@icmc.usp.br

***Abstract.** This paper reports the reflections regarding the particle “se” arisen during the annotation of a Brazilian Portuguese corpus with semantic role labels. The particle “se” is multi-function in Portuguese and poses some difficulties to be classified, even syntactically. As the guidelines used for such semantic role labels annotation were conceived for English corpus, they obviously are not suitable to tackle all the functions of “se”. Project decisions present as a result a table with labels to be assigned to the particle “se” pronoun, which may benefit not only SRL, but also other studies in Natural Language Processing.*

***Resumo.** Este artigo relata as reflexões acerca da partícula “se” que surgiram durante a anotação de um corpus de português do Brasil com rótulos de papéis semânticos. A partícula “se” é multifuncional em português e apresenta certa dificuldade para ser classificada, mesmo sintaticamente. Como o guia usado para anotação de rótulos de papéis semânticos foi concebido para o inglês, obviamente não consegue contemplar todas as funções do “se”. As decisões de projeto tomadas resultaram em uma tabela com os rótulos a serem atribuídos à partícula “se” quando pronome, o que poderá beneficiar não somente estudos de rotulação de papéis semânticos, como também outros estudos em Processamento de Línguas Naturais.*

## 1. Introdução

A descrição do português pode ter como finalidade prover informações a falantes nativos, a aprendizes estrangeiros ou a sistemas computacionais. É nesse último caso que se enquadra o trabalho descrito neste artigo. Reportamos as decisões tomadas para anotar a partícula “se” para fins de rotulação de papéis semânticos em um corpus de português do Brasil. O projeto de anotação semântica que motivou o trabalho aqui descrito é o Propbank-Br (Duran, 2009) e utilizou o corpus Bosque, parte manualmente revisada do Floresta Sintá(c)tica (<http://www.linguateca.pt/floresta/>).

A partícula “se” é um dos casos que oferecem mais dificuldade para a anotação de papéis semânticos. Essa partícula tem duas classificações morfológicas: pronome e conjunção (além de poder ser substantivado como em “Existe um *se* nesta questão”). No corpus utilizado, que já estava sintaticamente anotado pelo parser Palavras (Bick, 2000),

observamos que a determinação automática da classe morfológica do “se” foi realizada com êxito. Na anotação sintática realizada pelo parser Palavras, o “se” conjunção não recebe nenhum rótulo individual (sua função é introduzir orações subordinadas) ao passo que o “se” pronominal recebe um rótulo de sintagma nominal (NP). No entanto, o parser não distingue se o “se” pronominal é recíproco, reflexivo, partícula integrante do verbo, partícula apassivadora, partícula expletiva ou índice de indeterminação do sujeito, como detalhado na Tabela 1:

Função	Exemplo	Testes
Pronome reflexivo	<i>Ele se feriu</i>	Aceita substituição do “se” por “a si mesmo”
Pronome recíproco	<i>Eles se encontraram</i>	Admite substituição do “se” por “um ao outro”
Partícula apassivadora	<i>Vendem-se casas</i>	Aceita transformação em passiva analítica: <i>Casas são vendidas</i> . Só ocorre com verbos transitivos diretos.
Índice de indeterminação do sujeito	<i>Concordou-se com tudo o que foi dito.</i>	Ocorre com verbos transitivos indiretos. Não aceita passiva analítica. Aceita substituição do “se” por “alguém” ou “ninguém”. Verbo na 3ª. pessoa do singular.
Partícula expletiva	<i>Acabou-se a festa.</i>	Aceita omissão do “se” sem que a frase se torne incorreta. Ocorre com verbos intransitivos
Partícula integrante do verbo	<i>Ele apaixonou-se assim que a viu.</i>	Não aceita substituição por “a si mesmo”

**Tabela 1. Funções do “se” pronominal e testes para distingui-las**

No nível semântico, somente o “se” pronominal é candidato a receber um papel na estrutura argumental do verbo, pois ele pode refletir o verdadeiro “ocupante” do papel. Ocorre, porém, que nem sempre o “se” pronominal reflete um argumento do verbo. Quando o “se” for índice de indeterminação do sujeito, partícula apassivadora, partícula integrante do verbo ou partícula expletiva, não deve receber anotação de papel semântico. A análise automática do parser, contudo, não fornece nenhuma informação que auxilie a distinguir essas funções. A fim de suprir essa lacuna, decidimos realizar uma anotação extra no corpus, com as funções do “se” pronominal, mesmo aquelas que não correspondam a papéis semânticos.

Na Seção 2 fornecemos uma breve descrição do que é a anotação de papéis semânticos, na Seção 3 relatamos nosso método de anotação, na Seção 4 fazemos a discussão do assunto e na Seção 5 apresentamos as considerações finais e apontamos possibilidades de trabalhos futuros.

## 2. Anotação de Papéis Semânticos

A anotação de papéis semânticos consiste em atribuir rótulos de papéis semânticos aos argumentos de um verbo. A teoria subjacente é a da semântica de frames, de Fillmore (1968).

Há pouco mais de uma década diversas iniciativas têm sido tomadas no sentido de anotar corpora com papéis semânticos, de modo a ter um modelo para a construção de anotadores automáticos de papéis semânticos, o que permitirá melhorar o processamento automático dos textos. As primeiras iniciativas foram para o inglês e são tomadas como modelo para projetos semelhantes em outras línguas. Até o momento não há corpora de português do Brasil disponíveis com rótulos de papéis semânticos, mas estão previstos

nos projetos Framecorp (Chishman et al. 2009), Framenet Brasil (Salomão, 2009) e Propbank-Br (Duran, 2009). Nos dois primeiros casos, o modelo seguido é o da Framenet (Baker et al. 1998) e no último, o do Propbank (Palmer et al. 2005). É natural, no entanto, que muitas das definições contidas nos manuais de anotação dos modelos do inglês não se apliquem ou não sejam suficientes para lidar com a língua portuguesa. Isso exige decisões que contribuirão para a elaboração de um guia de anotação de papéis semânticos para o português.

### 3. Materiais e Métodos

Estamos anotando a porção brasileira do Bosque, que possui 4213 sentenças. A ferramenta de anotação utilizada é a SALTO (Burchardt, 2006). Estamos anotando os papéis sobre a árvore sintática, ou seja, agregamos uma nova camada de anotação a um corpus já anotado sintaticamente pelo parser Palavras (Bick, 2000).

No caso do “se” anotado como pronome pelo parser Palavras, realizamos uma dupla anotação no Propbank-Br. Uma anotação detalhada foi atribuída às sentenças em que ele ocorre, visando trabalhos futuros (rótulos da primeira coluna da Tabela 2). Uma anotação com rótulo de papel semântico, seguindo as instruções do Propbank, foi atribuída “em cima” da anotação sintática do “se” (terceira coluna da Tabela 2).

Rótulo atribuído à sentença	Conteúdo	Rótulo atribuído ao “se”
SE-REF-OD	Pronome reflexivo como objeto direto	argm-rec
SE-REF-OI	Pronome reflexivo como objeto indireto	argm-rec
SE-REC	Pronome reflexivo recíproco	argm-rec
SE-PAS	Partícula apassivadora	nenhum
SE-IND	Índice de indeterminação do sujeito	nenhum
SE-EXP	Partícula expletiva	nenhum
SE-VPR	Partícula integrante do verbo	nenhum

Tabela 2. Rótulos de anotação do “se” pronominal no Propbank-Br

### 4. Discussão

Para a atribuição de papéis semânticos, estabelecemos as três regras. A primeira delas é que os pronomes reflexivos e recíprocos devem ser anotados com uma etiqueta especial, “argm-rec” (Figura 1), o que permitirá um pós-processamento para ligar esses pronomes ao sintagma nominal que eles refletem e que são os ocupantes de um papel semântico na estrutura argumental do verbo.

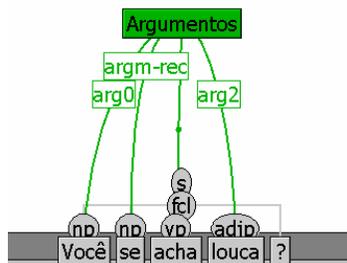


Figura 1. Anotação da partícula “se” pronominal com valor reflexivo no Propbank-Br

A segunda regra é que os pronomes na função de partícula apassivadora, de índice de indeterminação do sujeito e de partícula expletiva não devem receber anotação de papel semântico (Figura 2):

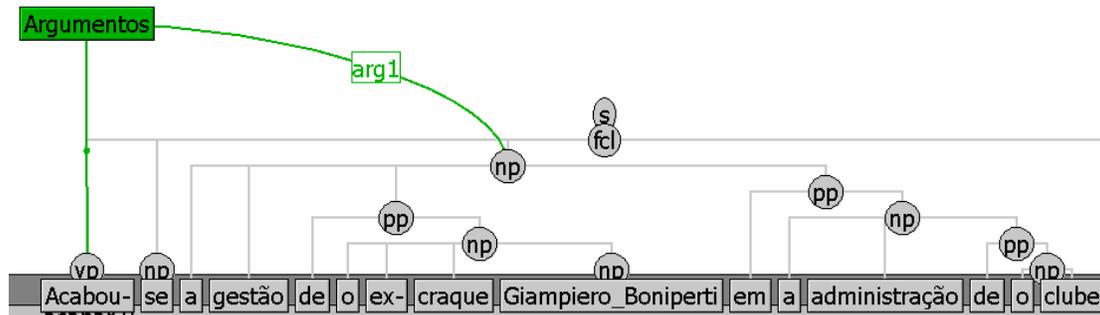


Figura 2. Partícula “se” pronominal com valor expletivo (não anotada no Propbank-Br)

A terceira regra é que a partícula “se” integrante do verbo deve ser “ligada” ao verbo por meio da anotação (Figura 3).

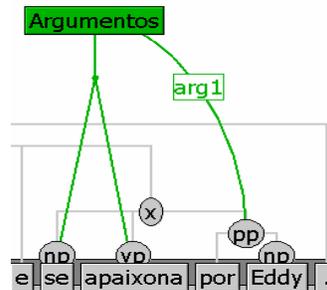


Figura 3. Tratamento da partícula “se” integrante do verbo no Propbank-Br

## 5. Considerações Finais

A anotação extra com as funções do “se” pronominal poderão beneficiar tanto a anotação de papéis semânticos quanto esforços de aperfeiçoamento dos parsers. As instâncias anotadas poderão ser usadas para levantar regras ou calcular correlação entre rótulos morfológicos, sintáticos e os parâmetros indicando as funções do “se”. Alguns recursos léxicos já se anunciam importantes para subsidiar a desambiguação automática das funções do “se”: uma relação dos verbos transitivos diretos, dos transitivos indiretos e dos intransitivos (importante para distinguir partícula apassivadora de índice de indeterminação do sujeito), uma relação de: verbos pronominais (esquecer-se, queixar-se, arrepende-se etc.); verbos que admitem uso reflexivo (pentear-se, ferir-se, cortar-se etc.) e verbos que admitem uso reflexivo recíproco (beijar, abraçar, encontrar, ver, amar etc.).

## 6. Agradecimentos

Agradecemos à Fapesp pelo financiamento do projeto Propbank-Br, dentro do qual realizamos o trabalho aqui relatado.

## **Referências Bibliográficas**

- Baker, C.F., Fillmore, C. J.; Lowe. J. B. (1998).The Berkeley FrameNet Project. In: Proceedings of Computational Linguistics 1998 Conference, University of Montréal, pp. 86-90.
- Bick, E. (2000) The Parsing System Palavras Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework. Aarhus, Denmark, Aarhus University Press.
- Burchardt, K. E. et alli. (2006) SALTO - A Versatile Multi-Level Annotation Tool. In: Proceedings of LREC-2006, Genoa, Italy.
- Chishman, R.; Bertoldi, A.; Padilha, J. G. (2009) Usando o FrameNet para a descrição semântica: um experimento de anotação de corpus. In: Jornada de Descrição do Português, 2009, São Carlos. VII STIL, 2009. v. 1.
- Duran, M. S. (2009) Propbank-Br: anotação de papéis semânticos em um corpus do português do Brasil. Projeto de Pós-Doc financiado pela FAPESP, Processo 2009/07394-9.
- Fillmore, C. The Case for Case (1968) In: Bach and Harms (Ed.): Universals in Linguistic Theory. New York: Holt, Rinehart, and Winston, 1-88.
- Palmer, M.; Gildea, D.; Kingsbury, P. (2005) The Proposition Bank: An Annotated Corpus of Semantic Roles. Computational Linguistics, 31:1., pp. 71-105, March, 2005.
- Salomão, M. M. M. (2009) FrameNet Brasil: um trabalho em progresso. Calidoscópico Vol. 7, n. 3, p. 171-182, set/dez 2009.