
**GUIA DE ANOTAÇÃO DE PAPÉIS SEMÂNTICOS SEGUINDO O MODELO
PROPBANK NO CÓRPUS PORTTINARI-BASE**

**MAGALI SANCHES DURAN
CLÁUDIA FREITAS**

Nº 449

RELATÓRIOS TÉCNICOS



**São Carlos – SP
Nov./2024**

POeTiSA

POrtuguese processing – Towards Syntactic Analysis and parsing

Guia de Anotação de Papéis Semânticos seguindo o modelo PropBank no Córpus Porttinari-base

Magali Sanches Duran

Cláudia Freitas

Novembro de 2024

Este trabalho foi realizado no âmbito do Centro de Inteligência Artificial da Universidade de São Paulo (C4AI - <http://c4ai.inova.usp.br/>), com o apoio da Fundação de Amparo à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM. Este projeto também foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei N. 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Residência em TIC 13, DOU 01245.010222/2022-44.

Sumário

1. Introdução	4
2. A tarefa de anotação de Papéis Semânticos no modelo Propbank	5
3. Anotação de ArgNs	7
2.1 Quando atribuir Arg0	8
2.2. Quando atribuir Arg1	9
3. Anotação de Modificadores (ArgMs)	11
3.1 ArgM-tmp: Tempo	13
3.2 ArgM-loc: Locativo	14
3.3. ArgM-mnr: Modo ou Maneira	15
3.2. ArgM-cau: Causa	16
3.3. ArgM-prp: Propósito ou Finalidade	17
3.4. ArgM-neg: Negação	18
3.5. ArgM-ext: Quantidade	19
3.6. ArgM-dir: Direção	20
3.7 ArgM-rec: Recíprocos	21
3.8. ArgM-prd: Marcadores de predicação secundária	22
3.9. ArgM-dis: Marcadores Discursivos	23
3.10. ArgM-adv: Adverbiais	24
3.11. ArgM-com: Companhia	26
3.12 Auxiliares	27
3.15 ArgM-cond: Condicionais	31
3.16 ArgM-comp: Comparações	32
3.17 ArgM-src: Fonte da informação	32
3.18 ArgM-conseq: Consequência	33
3.19 Casos ambíguos entre dois ou mais ArgMs	34
3.19.1 aqui	34
3.19.2 cada vez mais	34
3.19.3 conforme	35
3.19.4 depois	35
3.19.5 diante de	35
3.19.6 geralmente, em geral	36
3.18.7 jamais, nunca	36
3.19.8 tradicionalmente	36

3.20 Casos Especiais	37
3.20.1 Modificadores em frases complexas	37
3.20.2 Atribuindo rótulos de papéis semânticos em interrogações	38
3.20.3 Anotação de Agentes da passiva	39
3.20.4 Anotando a partícula “se”	39
3.20.5 Utilização não padrão do verbo “tratar”	41
3.20.6 Verbos que integram locuções conjuntivas ou adverbiais	41
4. Anotando sobre Árvores Sintáticas	41
4.1 Relação entre elementos sintáticos e papéis semânticos	41
4.2 O argumento pertence ou não à estrutura argumental do verbo-alvo da anotação?	42
5. Considerações Finais	42
Referências	44
Apêndice 1: Advérbios terminados em -mente anotados como ArgM-tmp	45
Apêndice 2: Advérbios terminados em -mente anotados como ArgM-mnr	46
Apêndice 3: Advérbios terminados em -mente anotados como ArgM-dis	47
Apêndice 4: Advérbios terminados em -mente anotados como ArgM-adv	48

Guia de Anotação de Papéis Semânticos no Corpus Porttinari-Base

1. Introdução

Este Guia reflete as diretrizes seguidas pelo projeto de anotação de papéis semânticos (Freitas, 2024) no corpus Porttinari-base (Duran et al. 2023).

O conteúdo aqui apresentado tomou por base o Manual de Anotação de Papéis Semânticos, cuja primeira versão foi publicada sob a forma de um Relatório Técnico do NILC (Duran, 2010) a fim de guiar a anotação do corpus PropBank-Br (Duran & Aluísio, 2012) e cuja segunda versão, estendida, foi publicada em 2014, no projeto Prosa¹, a fim de guiar a anotação conjunta do corpus PropBank-Br.v2 (Duran et al., 2014). Este guia incorpora toda a experiência acumulada ao longo de 10 anos e de três projetos de anotação de papéis semânticos em português. Tanto as duas versões do Manual, quanto o presente guia foram baseados nas *Guidelines*² do PropBank do inglês (Babko-Malaya, 2005; Bonial et al., 2015).

O projeto que motivou a presente versão foi desenvolvido no âmbito do projeto POeTiSA (POrtuguese processing - Towards Syntactic Analysis and parsing), que faz parte da iniciativa de Processamento de Línguas Naturais (NLP2 - Natural Language Processing for Portuguese) do Centro de Inteligência Artificial (C4AI - Center for Artificial Intelligence) da Universidade de São Paulo, financiado pela IBM e pela FAPESP (projeto nr. 2019/07665-4).

O C4AI faz parte do Programa de Centros de Pesquisa em Engenharia da FAPESP e tem o compromisso de desenvolver trabalhos que explorem o que há de mais moderno em Inteligência Artificial, combinando pesquisa de base com aplicações. Em especial, o POeTiSA é um projeto de longo prazo que visa aumentar os recursos baseados em sintaxe e desenvolver ferramentas e aplicações relacionados à língua portuguesa do Brasil, buscando alcançar resultados de ponta nesta área. Nos últimos quatro anos o projeto vem produzindo, entre outros recursos, um corpus multigênero grande e abrangente, chamado Porttinari (Pardo et al. 2021), anotado segundo o modelo Universal Dependencies³ (UD) (de Marneffe et al. 2021), do qual o Porttinari-base é a primeira parte concluída. Destacam-se o Porttagger⁴ (Silva et al. 2023) e o PortParser⁵ (Lopes & Pardo, 2024), treinados com base no Porttinari-base, que

¹ <https://sites.google.com/view/prosa-nilc>

² A primeira versão do guia de anotação do PropBank do inglês está disponível em <http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf>. A versão mais recente (2015) está disponível em <https://verbs.colorado.edu/propbank/EPB-Annotation-Guidelines.pdf>

³ <https://universaldependencies.org/>

⁴ <https://huggingface.co/spaces/Emanuel/porttagger>

⁵ <http://portparser.icmc.usp.br:8082/>

alcançaram acima de 99% e 95% de precisão, para *Part-of-Speech* e relações de dependência sintática, respectivamente.

Seguindo a tendência em PLN de reunir em um mesmo corpus diversos tipos de anotação, propôs-se, no último ano, adicionar ao corpus Porttinari-base a anotação de papéis semânticos, usando o conjunto de papéis do PropBank (Palmer, Gildea & Kingsbury, 2005), que já foi usado por outras línguas e vem sendo adotado pelos corpus anotados com UD. Essa nova camada de anotação atribui, ao verbo, um identificador de sentido contido no repositório Verbo-Brasil⁶, bem como etiquetas de papéis semânticos aos argumentos do verbo. O propósito final é constituir um corpus de treinamento para um classificador automático de papéis semânticos.

A anotação dos papéis semânticos foi feita nos arquivos CoNLL-U do Porttinari-base. O CoNLL-U é um modelo de arquivo adotado pela UD, para cada sentença de um corpus, constituído originalmente de 10 colunas de informações. Com a anotação de papéis semânticos, o CoNLL-U passa a ter 12 colunas, gerando o que vem sendo chamado de CoNLL-U Plus⁷. A anotação de papéis semânticos no Porttinari-base está descrita em Freitas & Pardo (2024) e, com mais detalhes, em relatório técnico específico (Freitas, 2024).

2. A tarefa de anotação de Papéis Semânticos no modelo Propbank

No PropBank, os papéis semânticos podem ser divididos em dois grandes blocos: os papéis semânticos numerados (ArgNs) e os papéis semânticos modificadores (ArgMs).

Os ArgNs (Arg0, Arg1, Arg2, Arg3, Arg4, Arg5) são os papéis previstos pela estrutura argumental dos verbos e podem ser encontrados no repositório Verbo-Brasil⁸, que lista os sentidos de cada verbo num arquivo (*framefile*) e, para cada sentido, apresenta um conjunto de papéis semânticos previstos (*roleset*). Raros são os verbos que preveem mais de três argumentos numerados.

No Verbo-Brasil, cada sentido “traduz” os papéis numerados em palavras, tornando mais simples para o anotador identificar cada um deles nas instâncias de anotação. Por exemplo, o conjunto de papéis semânticos do primeiro sentido do verbo “abrir”, ou seja, “abrir.01”, prevê quatro ArgNs: o Arg0 é o “abridor” ou “Agente”, o Arg1 é “coisa abrindo” ou “Tema”, o Arg2 é “Instrumento” e o Arg3 é “Beneficiário”. O Verbo-Brasil traz exemplos anotados para ilustrar a atribuição dos papéis semânticos.

⁶ <http://143.107.183.175:21380/verbobrasil/>

⁷ <https://universaldependencies.org/ext-format.html>

⁸ No projeto PropBank inglês, o recurso lexical que apoia a anotação também é chamado de Propbank. Em português, o recurso lexical que apoia a anotação de propbanks chama-se Verbo-Brasil. O repositório de verbos do PropBank inglês pode ser consultado em <https://verbs.colorado.edu/propbank/framesets-english-aliases/>

Embora previstos, os ArgNs não precisam ocorrer todos ao mesmo tempo. Por exemplo, na sentença: “O treino do Brasil foi aberto ao público”, ocorrem apenas o Arg1 “O treino do Brasil” e o Arg3 “ao público”.

Já os ArgMs não são previstos pela estrutura argumental dos verbos, podem ocorrer com vários verbos, e podem ser suprimidos sem que o sentido do verbo fique incompleto. Correspondem, no nível sintático, aos adjuntos adverbiais e podem ser expressos por advérbios, locuções adverbiais ou orações subordinadas adverbiais.

O conjunto de papéis semânticos utilizados na anotação de papéis semânticos no *Porttinari-base* é o mesmo usado pelo PropBank do inglês, com o acréscimo de alguns ArgMs, como será exposto mais adiante.

Assim como no PropBank inglês, este guia utiliza o recurso dos ArgM para anotar outros fenômenos que afetam a anotação semântica, mas que não são propriamente modificadores. Por exemplo, o ArgM-Dis é usado para anotar marcadores discursivos, ou seja, elementos cujo papel é fazer a coesão entre orações e que não pertencem à estrutura argumental dos verbos em si.

É importante saber que existe uma zona de intersecção entre ArgNs e ArMs. Por exemplo, o papel semântico Locativo (lugar) é um típico ArgM (ArgM-Loc). Para alguns verbos, porém, o Locativo é um argumento previsto e deve receber um rótulo de ArgN (argumento numerado). É o caso do verbo “morar” (alguém mora em algum lugar) e do verbo “colocar” (alguém coloca alguma coisa em algum lugar). Isso está explícito nos conjuntos de papéis semânticos desses verbos no Verbo-Brasil.

Este guia é importante sobretudo para esclarecer dúvidas a respeito dos ArgMs, pois, não sendo previstos pela semântica dos verbos, não são contemplados no Verbo-Brasil, exceto eventualmente nos exemplos de cada sentido dos verbos.

Foram adotadas, sempre que possível, as mesmas diretrizes do PropBank de língua inglesa. Algumas diferenças, contudo, são inevitáveis devido às diferenças entre as línguas e entre a anotação sintática produzida pelo *parser* utilizado em cada um dos projetos em suas respectivas línguas.

A anotação é feita no nível de sentenças e, dentro das sentenças, cada verbo pleno tem sua estrutura argumental anotada. Nos exemplos do guia que apresentam a anotação, o verbo cuja estrutura está sendo anotada é marcado como REL. Sendo assim, o número de estruturas argumentais anotadas de uma sentença será igual ao número de verbos plenos contidos na sentença (verbos auxiliares e de cópula não possuem estrutura argumental).

Um mesmo constituinte sintático poderá fazer parte da estrutura de mais de um verbo, como mostra o exemplo a seguir, que apresenta dois verbos e, portanto, terá duas estruturas argumentais anotadas:

*Pedro fez a pergunta por não ter **entendido** direito a explicação.*

Arg0: Pedro

REL: fez

Arg1: a pergunta

ArgM-cau: por não ter entendido direito a explicação

Arg0: Pedro

ArgM-neg: não

ArgM-tml: ter

REL: entendido

ArgM-mnr: direito

Arg1: a explicação

Nesse exemplo, o Arg0 de “fez” e de “entendido” é “Pedro”. Já a oração adverbial causal é um modificador ArgM-cau na estrutura argumental de “fez”, e seu núcleo (“entendido”), é centro de outra estrutura argumental analisada na sentença.

A anotação de papéis semânticos de que trata esse guia e os manuais que o precederam é feita sobre *córpus* já analisados sintaticamente por um parser, ou seja, sobre *árvores sintáticas*. Nos Projetos Propbank-Br e Propbank-br V2, tratava-se de um parser de constituintes e no Porttinari-base trata-se de um parser de dependências. Os exemplos deste guia mostram o constituinte que corresponde a cada papel semântico, porém, na *árvore de dependências*, apenas o nó principal do constituinte receberá a etiqueta de papel semântico (pois os demais *tokens* do constituinte estarão ligados a ele por relações de dependência).

3. Anotação de ArgNs

Para que realizem seu sentido, os verbos “pedem” alguns argumentos. Esses argumentos correspondem, na sintaxe, ao sujeito e aos objetos direto e indireto, salvo raras exceções. Não existe, porém, uma regra simples para atribuição dos papéis semânticos a partir dos papéis sintáticos.

Para alguns verbos, é impossível fornecer um único conjunto de papéis semânticos que atenda a todos seus sentidos. Por exemplo, o verbo “partir” nos exemplos abaixo:

Ele partiu o bolo antes da hora.

Ele partiu ontem.

Nesses exemplos, os dois sentidos do verbo “partir” pedem diferentes argumentos. Em casos assim, o *framefile* do verbo no Verbo-Brasil distingue dois ou mais sentidos e cada um desses sentidos é chamado de *frameset*. Cada *frameset* tem seu *roleset*, ou seja, seu conjunto específico de etiquetas para anotar os argumentos com papéis semânticos.

Partir.01 - cortar

Arg0: aquele que corta

Arg1: coisa cortada

Partir.02 - ir embora

Arg0: aquele que vai embora

É importante, portanto, verificar se o verbo a ser anotado tem mais de um sentido no Verbo-Brasil e, se tiver, decidir qual deles é o mais próximo do sentido do verbo no contexto de anotação.

Assim, de acordo com o exemplo, um argumento anotado com Arg0 do verbo partir.01 poderá ser interpretado semanticamente como “aquele que corta” (Agente do ato de partir algo) e um argumento anotado com Arg0 do verbo partir.02 poderá ser interpretado semanticamente como o “aquele que vai embora” (Agente do ato de ir embora).

2.1 Quando atribuir Argo

O Arg0 é atribuído a argumentos que desempenham o papel de Agentes, Causas ou Experienciadores, mas nunca a Pacientes ou Temas (que sofrem a ação do verbo ou são objeto da ação do verbo).

Pedro pescou um peixe.

Arg0: Pedro

REL: pescou

Arg1: um peixe

Pedro pescou ontem.

Arg0: Pedro

REL: pescou

ArgM-tmp: ontem

Os carros dos filhos são frequentemente pagos por seus pais.

Arg1: os carros dos filhos

Argm-tmp: frequentemente

REL: pagos

Arg0: por seus pais

O cinema abre às duas horas.

Arg1: o cinema

REL: abre

ArgM-tmp: às duas horas.

O gerente abre a agência às nove em ponto.

Arg0: o gerente

REL: abre

Arg1: a agência

ArgM-tmp: às nove em ponto

O Arg0 tem como propriedades:

- ter envolvimento emocional com o evento ou estado
- ser causa ou causador de um evento ou mudança de estado em outro participante
- denotar movimento em relação à posição de outro participante.
- praticar ações que exigem volição (vontade)

Na voz ativa, todos Arg0 são sujeitos sintáticos (mas nem todo sujeito sintático é Arg0, como será visto a seguir). Na voz passiva, o Arg0 é o Agente da passiva, normalmente introduzido pela preposição “por”.

O vaso foi derrubado pelo vento.

Arg1: o vaso

ArgM-pas: foi

REL: derrubado

Arg0: pelo vento

Essa hipótese foi levantada pelos estudantes.

Arg1: essa hipótese

ArgM-pas: foi

REL: levantada

Arg0: pelos estudantes

2.2. Quando atribuir Arg1

O Arg1 é atribuído a argumentos que desempenham o papel de Pacientes, Temas e Tópicos, isto é, os argumentos que sofrem a mudança de estado ou que são afetados pela ação do verbo. Eles podem ocorrer sintaticamente como sujeito ou como objeto.

Mesmo que ocorra alternância sintática (mudança de posição dos constituintes sintáticos), o papel semântico não é alterado. Nos exemplos a seguir, *a vidraça* e *a comida* são sempre Arg1.

Eu quebrei a vidraça.

A vidraça quebrou.

A vidraça foi quebrada.

Eu queimei a comida.

A comida queimou.

A comida foi queimada.

O Arg1 tem como propriedades:

- não apresentar volição;
- sofrer mudança de estado;
- ser afetado pela ação de outro participante;
- ser estacionário em relação ao movimento de outro participante.

Alguns verbos admitem tanto o Arg0 quanto o Arg1 na posição de sujeito. Isso sempre ocorre quando os verbos estão na voz passiva, pois o Paciente toma o lugar do Agente:

VOZ ATIVA

O soldado prende o homem.

Arg0: O soldado

REL: prende

Arg1: o homem

VOZ PASSIVA

O homem foi preso pelo soldado.

Arg1: o homem

ArgM-pas: foi

REL: preso

Arg0: pelo soldado

Na voz ativa, apenas alguns verbos permitem tanto Arg0 quanto Arg1 na posição de sujeito. São exemplos desses verbos: abrir, fechar, queimar e quebrar.

COM ARG0 SUJEITO

Os lojistas fecharam as portas às 18 horas.

Arg0: os lojistas

REL: fecharam

Arg1: as portas

ArgM-tmp: às 18 horas

COM ARG1 SUJEITO

As portas das lojas fecharam às 18 horas.

Arg1: as portas

REL: fecharam

ArgM-tmp: às 18 horas

Outros verbos admitem a alternância do Arg0 e do Arg1 na posição de sujeito da voz ativa, mas para isso precisam se pronominalizar, marcando o movimento, como ilustrado pelos exemplos a seguir.

COM ARG0 SUJEITO

O professor magoou o aluno com a advertência que fez.

Arg0: O professor

REL: magoou

Arg1: o aluno

ArgM-mnr: com a advertência que fez

COM ARG1 SUJEITO

O aluno magoou-se com a advertência feita pelo professor.

Arg1: o aluno

REL: magoou

ArgM-nse: se

ArgM-cau: com a advertência que feita pelo professor

Percebe-se que a pronominalização não tem função reflexiva, pois “magoar-se” não significa “magoar a si mesmo”, mas “ficar magoado”. Nesses casos o pronome (se, me, nos...) deve ser anotado como ArgM-nse (etiqueta criada no Propbank-Br), pois não tem função argumental e é usado apenas para marcar o movimento do Arg1 para a posição do sujeito. Note-se que no primeiro exemplo “com a advertência” é um Instrumento (ArgM-mnr) de “magoar” e no segundo exemplo “com a advertência” é a Causa (ArgM-cau) de “magoar-se”.

3. Anotação de Modificadores (ArgMs)

Como já foi dito na Introdução, os ArgMs são aqueles que não estão previstos na estrutura argumental dos sentidos dos verbos no Verbo-Brasil. A maioria deles não apresenta desafios para os anotadores de papéis semânticos e correspondem a adjuntos adverbiais na anotação sintática. Ao contrário dos ArgNs, um ArgM pode ocorrer mais de uma vez em uma mesma estrutura argumental. Assim, desde que cada ArgM constitua uma unidade independente, deve ser anotado separadamente.

Os projetos de anotação dos propbanks de português brasileiro citados neste guia utilizam as etiquetas de ArgM apresentadas na Tabela 1. Com exceção de ArgM-nse, todas as etiquetas da Tabela 1 são idênticas às do PropBank inglês.

Tabela 1. Etiquetas de Argumentos Modificadores (ArgM)

ArgM-	PropBank	Papel semântico no Porttinari-base
ArgM-tmp	Time	Tempo
ArgM-cau	Cause	Causa
ArgM-loc	Locative	Local
ArgM-neg	Negation	Negação
ArgM-mnr	Manner	Modo
ArgM-prp	Purpose	Propósito ou Finalidade
ArgM-ext	Extent	Quantidade
ArgM-com	Commitative	Companhia
ArgM-dir	Directional	Direção
ArgM-adv	Adverbial	Advérbios que modificam a oração inteira
ArgM-prd	Secondary Predication	Predicação Secundária
ArgM-rec	Reciprocals	Reciprocidade
ArgM-dis	Discourse	Marcador Discursivo
ArgM-nse	Marcador de alternância	Pronome reflexivo não argumental

Além disso, ao longo da anotação de papéis semânticos no corpus Porttinari-base, foram criadas mais três etiquetas: ArgM-src, ArgM-conseq, e ArgM-cond (Tabela 2), todas elas especificações de relações semânticas que antes estavam agrupadas na etiqueta genérica ArgM-adv.

Tabela 2. Etiquetas de Argumentos Modificadores que são especificações de ArgM-adv, usados apenas na anotação do Porttinari-base PropBank.

ArgM-	PropBank	Papel semântico no Porttinari-base
ArgM-src	-	FONTE da informação veiculada (<i>source</i>)
ArgM-conseq	ArgM-adv	CONSEQUÊNCIA
ArgM-cond	ArgM-adv	CONDIÇÃO

A seguir, cada um desses ArgM serão explicados e exemplificados em uma subseção. A Subseção 3.19 é dedicada a comentar unidades lexicais que apresentam ambiguidade entre os ArgMs e a Subseção 3.20 comenta casos especiais.

A fim de auxiliar a anotação, relações de advérbios terminados em -mente e suas respectivas classificações mais frequentes no corpus Porttinari-base são apresentadas sob forma de apêndices ao final deste guia.

3.1 ArgM-tmp: Tempo

Modificadores de tempo informam quando uma ação acontece e podem ser expressos por uma única palavra (“ontem”, “futuramente”, “frequentemente”), por um sintagma preposicionado (“em novembro”, “de 5 a 9 de julho”, “desde 1933”, “durante a exibição do filme”, “na semana passada”, “cinco vezes por semana”) ou por uma oração completa (“ao acordar”, “quando veio ao Brasil”, “enquanto dormia”, “sempre que tem vontade”, “até não querer mais”, “assim que cheguei”, “tão logo cheguei”, “uma vez acordado”).

Os **ArgM-tmp** respondem às perguntas **Quando? Por quanto tempo? Desde quando? Até quando? Com que frequência? Em quanto tempo? Para quanto tempo?**

A única preposição exclusiva de tempo é “durante”. A preposição mais comum no ArgM-tmp é “em”, mas não é exclusiva de tempo, pois também é muito usada para modificadores de lugar e de modo.

Estão incluídos nessa categoria:

- Adjuntos adverbiais de tempo: “ontem”, “antigamente”, “depois”
- Adjuntos adverbiais de frequência: “sempre”, “raramente”, “às vezes”
- Adjuntos adverbiais de duração: “por um ano”, “durante duas semanas”, “eternamente”, “para sempre”
- Adjuntos adverbiais de ordem: “primeiramente”, “primeiro... segundo...”, “na primeira vez”
- Adjuntos adverbiais de repetição: “de novo”, “novamente”, “uma vez”

A ideia de frequência envolve, simultaneamente, tempo (ArgM-tmp) e quantidade (ArgM-ext), pois frequência se refere a uma quantidade de repetições no tempo. Desde a versão anterior do guia de anotação do PropBank, a ideia de frequência (que responde à pergunta “com que frequência?”) está associada apenas ao tempo - e é importante lembrar que, por princípio, o PropBank não permite a anotação de mais de um papel para um mesmo elemento. Assim, em construções do tipo “várias vezes”, “muitas vezes”, etc., prevaleceu a anotação temporal, e todas foram anotadas como ArgM-tmp no Porttinari-base.

Chegou a pensar em mudar várias vezes: azul, rosa, amarelo [...]

Julia Louis-Dreyfus ganhou pela sexta vez consecutiva o troféu.

Em críticas internas , por diversas vezes exemplifiquei tratamentos diferenciados [...]

A marca já foi batida dez vezes na prova.

O trailer, lançado em seu canal no YouTube, foi visto 900 mil vezes em um mês.

Com o cenário econômico mais benigno, a gente cada vez mais enxerga apetite maior a risco [...]

Casos que, dependendo do contexto, podem parecer ambíguos para o anotador são discutidos na Seção 3.19.

O Apêndice 1 traz os advérbios terminados em -mente anotados como ArgM-tmp no Porttinari-base.

3.2 ArgM-loc: Locativo

Os locativos indicam o lugar onde uma ação acontece. Os locativos podem ser constituídos por uma única palavra (“aqui”, “ali”, “lá”, “dentro”, “fora”) ou por um sintagma preposicionado (“desde o km 51 da rodovia”, “na fazenda”, “dentro da caixa”, “ao lado da irmã”, “na página 2”).

Os **ArgM-loc** respondem às perguntas **Onde? De onde? Para onde? Por onde? Desde onde? Até onde?**

A noção de Locativo não se restringe a lugares físicos, mas estende-se também a lugares abstratos. Ex:

Em seu discurso, o presidente anunciou a reforma da previdência.

ArM-loc: em seu discurso

Arg0: o presidente

REL: anunciou

Arg1: a reforma da previdência

Alguns verbos preveem um locativo como ArgN e, portanto, exigem atenção, pois os ArgNs têm precedência sobre os ArgMs. Exemplos disso são os verbos da classe de *por* (*alocar, colocar, dispor*, etc.).

Os locativos também podem denotar eventos ou situações de comunicação, como em “disse em uma entrevista” e “lamentou em uma conversa”. Argumentos desse tipo foram anotados como ArgM-loc desde que não houvesse nenhuma outra menção que direcionasse o entendimento para Tempo.

Algumas unidades lexicais podem apresentar ambiguidade entre ArgM-tmp e outros ArgMs. Esses casos são discutidos na Seção 3.19, juntamente com outros casos de ambiguidade entre ArgMs.

3.3. ArgM-mnr: Modo ou Maneira

Esses modificadores especificam como uma ação é realizada. Essa etiqueta inclui os *instrumentos, matérias primas, modo de vestir* e outros.

Os **ArgM-mnr** respondem às perguntas: **Como? De que jeito? De que modo? De que maneira?**

Podem ser expressos por uma única palavra (‘suavemente’, ‘loucamente’, ‘intuitivamente’), por um sintagma preposicionado (“de repente”, “em detrimento de algo”, “às cegas”, “ao vivo” “pelo avesso”, “aos poucos”, “em tom de confiança”, “sem cessar”, “de bengala”, “com roupa de festa) ou por uma oração (ele chegou “gritando”, “portando uma arma” “vestido de preto”), além de algumas expressões não introduzidas por preposição (“on-line”, “um a um”, “um por um”). Essa é a categoria de modificadores que está mais fortemente associada a um léxico de expressões fixas.

Também são anotados como ArgM-mnr os casos de expressões temporais introduzidas por “como”: “como sempre”, “como antes”, “como antigamente”, por falta de uma etiqueta mais precisa.

Como sempre , vamos ouvir os feedbacks de usuários e motoristas parceiros (ArgM-mnr)

Parodiando vulgarmente Marx - aliás , como sempre fazem alguns petistas (ArgM-mnr)

Praticamente nada mudou em a cidade , que continua abandonada como antes. (ArgM-mnr)

A forma mais comum de expressar modo por meio de uma oração é utilizar uma oração reduzida de gerúndio.

Os empreendedores enfrentaram o problema construindo edifícios vendidos em a planta [...]

Além disso, orações proporcionais, como aquelas introduzidas por *à medida que*, *à proporção que*, também são anotadas como ArgM-mnr.

À medida que empresas trabalham para mostrar aos empregados que oferecem igualdade de oportunidades internas, o interesse feminino começa a aumentar , ainda que de forma discreta.

Também são anotadas como ArgM-mnr as **orações conformativas**, desde que estas respondam às perguntas “Como?”, “De que maneira?”

Esses crimes serão punidos conforme prescrito na lei. (ArgM-mnr)

A preposição “conforme” pode introduzir diferentes papéis semânticos e, como isso pode constituir uma dificuldade para o anotador, é discutida no item 3.19 juntamente com outros itens lexicais ambíguos.

Por fim, nem sempre a ideia de modo/maneira será anotada como ArgM-mnr, por isso é importante consultar sempre o Verbo Brasil. No *frameset* variar.01, por exemplo, a ideia de modo/maneira corresponde ao Arg2, conforme mostra o exemplo a seguir.

Nos eventos , o valor varia conforme o pacote .

ArgM-loc: Nos eventos

Arg1: o valor

REL: varia

Arg2: conforme o pacote

O Apêndice 2 lista todos os advérbios terminados em -mente anotados como ArgM-mnr no corpus Portinari-base.

3.2. ArgM-cau: Causa

Esse tipo de ArgM indica a causa de uma ação. São exemplos típicos os sintagmas e orações introduzidos por “porque” “em razão de” “por causa de”, “em função de”, “devido a”, “em virtude de”, “graças a”, “já que” e “uma vez que”.

Os **ArgM-cau** respondem à pergunta: **Por quê?**

Esse papel semântico nunca é expresso por uma única palavra, mas sim por adjuntos adverbiais sob forma de sintagmas preposicionados e por orações adverbiais desenvolvidas ou reduzidas de gerúndio ou participio.

Em razão da epidemia de gripe suína, as aulas foram suspensas.

ArgM-cau: em razão da epidemia de gripe suína

Arg1: as aulas

ArgM-pas: foram

REL suspensas

A família pretende manter a busca por recursos , já que Jonatas vai precisar tomar o remédio por toda a vida .

Arg0: A família

REL: pretende

Arg1: manter a busca por recursos

ArgM-cau: já que Jonatas vai precisar tomar o remédio por toda a vida .

Uma vez que Cabral descobriu o Brasil, quem segue o Cabral descobrirá alguma coisa.

ArgM-cau: Uma vez que Cabral descobriu o Brasil

Arg0: Quem segue o Cabral

REL: descobrirá

Arg1: alguma coisa

Advertido por uma funcionária, gritou palavrões e disse que iria dar um soco em ela .

Não sendo biodegradáveis, quando descartados permanecem no meio ambiente por centenas de anos.

Tangidos por regras cada vez mais rígidas de emissão de fumaça [...] montadoras estão olhando para os motores elétricos como a mais viável das soluções.

3.3. ArgM-prp: Propósito ou Finalidade

Esse tipo de argumento nunca é expresso por uma única palavra, mas sim por sintagmas preposicionados ou por orações. As orações são sua forma mais comum e são introduzidas por preposições e locuções prepositivas como: “para”, “a fim de”, “com o objetivo de”, “com o intuito de”, “com o propósito de”, “em busca de”, “em troca de” normalmente seguidas de infinitivo.

Os **ArgM-prp** respondem às perguntas: **Para quê? Com que finalidade?**

Espero voltar aqui para novas experiências.

Ele comprou um guarda-chuva resistente para enfrentar tempestades.

Nem todo mundo trabalha com o único propósito de ganhar dinheiro.

Arqueólogos seguem cavando, em busca de alguma ossada.

Estantes eram devassadas em busca de volumes que servissem para incriminar os respectivos leitores.

Em troca de não deportar os sonhadores, como são chamados os filhos de esses imigrantes sem documentos, o presidente pede (...).

Ela é a cantora Natasha, que vende a alma ao vampiro Vlad (Ney) em troca de sucesso na carreira.

Ele busca empresas que aceitem patrocinar o gasto em troca de nomear o local.

Arg0: que [empresas]

REL: aceitem

Arg1: patrocinar o gasto

ArgM-cau: em troca de nomear o local

Atenção: cuidado para não confundir este modificador (ArgM-prp, causa/*purpose*) com o modificador de predicação secundária (ArgM-prd, da Seção 3.8), uma vez que ambos têm grafias parecidas.

3.4. ArgM-neg: Negação

Essa etiqueta é usada para anotar elementos como “não”, “nunca”, “não mais”, “jamais”, “sequer” e outros marcadores de orações negativas. A negação é uma noção importante na anotação do PropBank, então todos os marcadores que indicam negação devem ser anotados com ArgM-neg. Por exemplo, ao encontrar elementos como “jamais” e “nunca”, que poderiam ser anotado como ArgM-tmp ou ArgM-neg, a etiqueta ArgM-neg deve ser usada.

Há uma pequena parcela, no entanto, que é catastrofista ao ponto de afirmar que Lula sequer conseguirá tomar posse.

Arg0: Lula

ArgM-neg: sequer

REL: conseguirá

Arg1: tomar posse

O advérbio “lá” na construção “sei lá” também é considerado negação, pois assume-se que é equivalente a “não sei”.

Ou sei lá qual é, hoje em dia , o termo " politicamente correto " para " cucaracha ".

REL: sei

ArgM-neg: lá

Arg1: qual é o termo " politicamente correto " para " cucaracha ".

ArgM-tmp: hoje em dia

3.5. ArgM-ext: Quantidade

Modificadores de quantidade **ArgM-ext** indicam “quanto uma ação provocou de mudança”.

Os **ArgM-ext** respondem à pergunta: **Quanto?**

Os ArgM-ext são expressos normalmente por:

- Adjuntos numéricos: “andar 20 quilômetros”, “enxergar 100 metros à frente”;
- Quantificadores: “muito”, “pouco”, “suficientemente”, “exageradamente”, “em demasia”, “em excesso”, “demais”, “substancialmente”, “excessivamente”, “ligeiramente”, “profundamente” (conforme o contexto)
- Comparativos: “mais do que”, “menos do que

[...], o holandês da Red Bull levou seu carro ligeiramente para a esquerda [...]

Por isso que tomei o cuidado de não mexer excessivamente .

[...] , seu nível de bem-estar cai substancialmente .

Não posso desestruturar a equipe em demasia .

Amava demais aquilo tudo.

[...] , ele depende profundamente do desenvolvimento individual .

Os advérbios terminados em -mente que são anotados como ArgM-ext são parafraseáveis por “muito” ou “pouco”.

Na realidade , eles [OAS] deram ao PMDB uma contribuição de R\$ 5 milhões , o que irritou profundamente o Eduardo Cunha , que nunca nutriu por mim nenhuma simpatia .

Arg0: o que

REL: irritou

ArgM-ext: profundamente

Arg1: o Eduardo Cunha, que nunca nutriu por mim nenhuma simpatia

Muitos verbos, por indicarem um movimento em uma escala, têm os quantitativos previstos na sua estrutura argumental. Por exemplo, “engordar” prevê um ArgN para marcar quantos quilos foram engordados. Nesse caso, eles devem ser anotados com os respectivos ArgNs descritos no Verbo-Brasil.

Se um verbo está acompanhado de um quantitativo e não tem um quantitativo previsto em sua estrutura argumental descrita no Verbo-Brasil, deverá ter o argumento anotado como ArgM-ext. Ex:

Eu sonhei muito durante a viagem de volta.

Arg0: Eu

REL: sonhei

ArgM-ext: muito

ArgM-tmp: durante a viagem de volta

Eu achei camisetas a R\$20,00

Arg0: Eu

REL: achei

Arg1: camisetas

ArgM-ext: a R\$20,00.

A ideia de *quantidade* também aparece em uma série de construções que envolvem *frequência*, como em “ele caiu no choro várias vezes/duas vezes” etc. No entanto, por definição, elementos de frequência estão anotados como ArgM-tmp no Propbank. A expressão “cada vez mais”, contudo, foi anotada algumas vezes como ArgM-ext no Portinari-base, pois a ideia de intensificação parece sobrepor a de frequência. Esse caso de ambiguidade é discutido no item 3.19.2.

3.6. ArgM-dir: Direção

Modificadores de direção mostram a direção do movimento em determinado caminho. Por exemplo: “de cima”, “para cima”, “de baixo”, “para baixo”, “da esquerda”, “para a esquerda”, “da direita”, “para a direita”, “de trás”, “para trás”, “da frente”, “para frente”, “em frente”, “de lado”, “para o lado”, “do sul”, “para o sul, etc.

Os **ArgM-dir** respondem às perguntas: **Em que direção? De que direção? Para qual direção?**

Esse papel semântico pode se confundir com o de local. Por isso, é importante destacar que esse papel só ocorre com verbos de movimento. Portanto, em orações como “Veja abaixo” o papel de “abaixo” é ArgM-loc e não ArgM-dir. Se não houver claramente uma direção que responda a uma das perguntas de direção, a etiqueta de locativo (ArgM-loc) deve ser utilizada.

Ele guinou à direita para evitar uma batida.

ArgO: Ele

REL: guinou

ArgM-dir: à direita

ArgM-prp: para evitar uma batida

A estatal reajustou para cima, e isso faz a receita dela crescer.

Ao contrário do que ocorre no inglês, essa etiqueta não é muito utilizada no português, pois os verbos de movimento do português embutem a direção. Por exemplo, a combinação "vir de volta", literalmente "*come back*", quase não é usada, porque preferimos o verbo "voltar" (destacando que "vir de volta" é diferente de "vir da praia", em que "praia" é a origem, o ponto de início (Arg3 do verbo “vir”)). O mesmo vale para a combinação "ir para baixo", tradução literal de "*go down*", e de uso muito menos frequente em português do que "descer".

3.7 ArgM-rec: Recíprocos

Esse modificador inclui palavras e expressões que indicam reciprocidade, quando um pronome reflexivo recíproco já estiver preenchendo um ArgN, ou seja, são expressões redundantes no contexto: “entre si”, “entre nós”, “um ao outro”, “um com o outro”. No exemplo a seguir, o pronome “se” já está preenchendo o Arg1 de “ajudar” e a expressão “umas às outras” apenas reforça a leitura de reciprocidade que deve ser dada ao pronome.

As pessoas se ajudam umas às outras quando fazem parte de um grupo.

Arg0: As pessoas

Arg1: se

REL: ajudam

ArgM-rec: umas às outras

ArgM-tmp: quando fazem parte de um grupo

Porém, quando não redundantes, essas expressões poderão ocupar normalmente a posição de um argumento numerado:

Irmãos fazem bem um ao outro.

Arg0: Irmãos

REL: fazem

Arg1: bem

Arg2: um ao outro

As marcas brigam entre si com base no preço.

Arg0: As marcas

REL: brigam

Arg1: entre si

ArgM-mnr: com base no preço.

3.8. ArgM-prd: Marcadores de predicação secundária

Esses marcadores são usados para indicar que um argumento de um predicado poderia carregar sua própria estrutura argumental, ou seja, o marcador de predicação secundária é um argumento do argumento.

Os exemplos típicos são os descritivos, normalmente introduzidos por “como”, e orações adjetivas que estiverem sintaticamente separadas do NP a que se referem. Nos exemplos a seguir os ArgM-prd estão sublinhados:

Lenços umedecidos foram fornecidos como cortesia no avião.

Exibida como sempre, ela desfilou seu vestido novo nas galerias do teatro.

O presidente, apoiado pelo Congresso, tomou medidas duras.

Filha de mascates sírios, nasceu em Catalão (GO).

Vizinhos e conhecidos lembram dele como reservado e de poucos amigos.

Fácil de beber, tem apenas 10 IBUs (unidade de amargor que passa de cem).

Com 11 andares, a galeria térrea liga as ruas Barão de Itapetininga e Dom José de Barros.

O " bolsão " da Uber fica ao lado do terminal 1 , com capacidade para 230 carros.

O dragão colocou-se diante da mulher que estava para dar à luz , a fim de lhe⁹ devorar o filho.

O povo de Rosário é indiferente a ele porque o vê pouco em carne e osso.

(...) tendo sido graduado como oficial de inteligência na ESNI (Escola de o Serviço Nacional de Informações).

3.9. ArgM-dis: Marcadores Discursivos

Essa etiqueta de ArgM serve para anotar os marcadores discursivos. Esses elementos são usados para fazer a coesão textual, ou seja, estruturar a lógica do discurso, relacionando as orações e as sentenças de um texto. Não constituem modificadores típicos, mas são anotados no escopo do PropBank. São elementos que conectam uma sentença à sentença anterior.

Exemplos dessa categoria são:

“também”, “além disso”, “no entanto”, “mas”, “conforme exposto anteriormente”, “por outro lado”, “por exemplo”, “da mesma forma”, “contudo”, “entretanto”, “todavia”, “porém”, “Em relação a xxx”, “Com relação a xxx”; “Quanto a xxx”; “No que se refere a xxx”, “No que toca a xxx”, “A nível de xxx”, “Dessa forma”, “Em vez disso”, “Nesse contexto”, “Em resumo” “Assim” “Dito isso”, entre outros.

Em relação às delações premiadas , Dodge deve fazer um pente-fino nas negociações (...).

ArgM-dis: Em relação às delações premiadas

Arg0: Dodge

MOD: deve

Arg1: fazer um pente-fino nas negociações (...)

Observe que, as conjunções **Mas**, **Ou** e **E**, só são etiquetadas como **ArgM-dis** no início das orações. Nos demais casos, quando conectam elementos no interior da oração sendo anotada, ficam como **ArgM-adv**, por falta de etiqueta mais adequada.

⁹ Nesse caso, o “lhe” está sendo usado no lugar de um genitivo: “a fim de devorar o filho **dela**”

Estamos todos bem, porém apreensivos.

=Todos nós estamos bem, porém [todos nós estamos] apreensivos.

Rel: estamos

Arg0: todos

Arg1: bem

Também são etiquetados como **ArgM-dis** os constituintes sintaticamente analisados como **vocativos**.

Maria, pode servir o jantar.

ArgM-dis: Maria

ArgM-mod: pode

REL: servir

Arg1: o jantar

O Apêndice 3 lista advérbios terminados em -mente e anotados como ArgM-dis no Porttinari-base.

3.10. ArgM-adv: Adverbiais

Os ArgM-adv são usados para anotar elementos que modificam toda a estrutura do evento ou que não se enquadram nas demais categorias de modificadores. Esses modificadores podem ser uma palavra, uma expressão ou uma oração.

Entre os que modificam toda a oração em que o verbo está inserido, estão:

“provavelmente”, “possivelmente”, “somente”, “felizmente”, “infelizmente”, “francamente”, “realmente”, “no fundo”, “na verdade”, “na realidade”, “em tese”, “teoricamente”, “em grande parte”, “ao que tudo indica”, “levando em conta xxx”, “levando em consideração xxx”, “em outras palavras”, “finalmente”, “por fim”, “exceto”, “fora isso”, “ainda que”, “a não ser que”, “por favor”, entre outros.

Já entre os que não se enquadram nas demais categorias de modificadores estão:

“até”, “inclusive”, “só”, “apenas”, “ao menos”, “em vez de”, “ao invés de”, “aliás”, “amiúde”, “ademais”, “decerto”, “tampouco”, “afinal”, “de fato”, “curiosamente”, “basicamente”, “simplesmente”, “especialmente”, “certamente”, “principalmente”, “difícilmente”, “surpreendentemente”, “necessariamente”, “difícilmente”, “justamente”, “diferentemente”, “praticamente”, “definitivamente”, “acima de tudo”, e as orações adverbiais concessivas introduzidas por: “embora”, “mesmo que”, “apesar de”.

A mesma anotação como ArgM-adv se aplica à expressão “ainda mais”, com sentido equivalente a “principalmente”:

Considero este fato essencial, ainda mais se tratando do presidente da República.

ArgM-adv: ainda mais

Arg0: se

REL: tratando

Arg1: do presidente da República

Felizmente você voltou.

ArgM-adv: Felizmente

Arg0: você

REL: voltou

Ao contrário do que você pode ter pensado, minha habilidade de dirigir um carro não foi afetada pelo acidente.

ArgM-adv: ao contrário do que você pode ter pensado

Arg1: minha habilidade de dirigir um carro

ArgM-neg: não

ArgM-pas: foi

REL: afetada

Arg0: pelo acidente

Devido a uma ambiguidade, às vezes pode haver dificuldade quanto à melhor forma de anotar um modificador, se como ArgM-mnr, que modifica o verbo ou ArgM-adv, que normalmente modifica toda a oração. É o caso dos exemplos abaixo:

Incrivelmente, ela canta.

ArgM-adv: incrivelmente (paráfrase: É surpreendente que ela consiga cantar.)

REL: canta

Arg0: ela

Ela canta incrivelmente.

Arg0: ela

REL: canta

ArgM-mnr: incrivelmente (paráfrase: Ela canta muito bem.)

Os seguintes elementos também foram considerados ArgM-adv no Portinari-base por falta de uma etiqueta mais adequada:

"Foi o pior dia da minha vida" , lembra, sobre a prisão, em 2015 .

Entre as novidades, Cambuhy destaca um trator capaz de fazer duas funções ao mesmo tempo .

A palestra de sexta-feira (15) foi promovida por uma loja maçônica de Brasília e acompanhada por integrantes do Rio de Janeiro e de Santa Catarina, entre outros.

O Apêndice 4 lista advérbios terminados em -mente anotados como ArgM-adv no Portinari-base.

3.11. ArgM-com: Companhia

Os ArgM-com Indicam com quem uma ação foi realizada. Embora o termo formal para este modificador seja "comitativo", podemos pensar neste argumento como "companhia"

Os ArgM-com responde à pergunta: Com quem?
--

Isso pode incluir pessoas ou organizações (entidades que possuem características de Agentes prototípicos: animacidade e volição), mas exclui objetos, que são considerados modificadores instrumentais (e, portanto, ArgM-mnr).

Por responder à pergunta "Com quem", a ausência de companhia, como em "sozinho", também é anotada com ArgM-com.

Hoje , ela fica sozinha com as recreadoras.

ArgM-tmp: hoje

Arg1: ela

REL: fica

ArgM-com: sozinha com as recreadoras

O presidente Michel Temer se sentará à mesa com o americano Donald Trump para jantar [...]

Então saiu da casa da mãe e foi morar sozinho , em outro bairro .

Porém, “junto”, “juntos” (e suas flexões no feminino) continuam sendo ArgM-mnr:

O trio teria , junto , cerca de 54 % de os votos .

Arg0: o trio

REL: teria

ArgM-mnr: junto

Arg1: cerca de 54 % de os votos

A dupla já havia trabalhado junta no filme " Valsa com Bashir " (2008) [...]

Escrevemos juntos , gravamos juntos.

3.12 Auxiliares: ArgM-asp, ArgM-mod, ArgM-pas e ArgM-tml

A anotação do que conta como um verbo auxiliar é uma decisão de cada projeto, pois a “lista” de verbos auxiliares não é um consenso linguístico. Os verbos auxiliares mais gramaticalizados no português são os auxiliares de tempo (“ter” e “haver”) e de voz passiva (“ser”), mas também podem ser considerados auxiliares outros verbos, como mostra a Tabela 1, que constava da versão do guia que deu base ao presente documento.

Tabela 1. VERBOS AUXILIARES

LEMA	PREP	FORMA DO AUXILIADO	ArgM-
<i>acabar</i>		<i>gerúndio</i>	ArgM-asp
<i>acabar</i>	<i>de</i>	<i>infinitivo</i>	ArgM-asp
<i>acabar</i>	<i>por</i>	<i>infinitivo</i>	ArgM-asp
<i>andar</i>		<i>gerúndio</i>	ArgM-asp
<i>cessar</i>	<i>de</i>	<i>infinitivo</i>	ArgM-asp
<i>chegar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>começar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>começar</i>	<i>por</i>	<i>infinitivo</i>	ArgM-asp
<i>continuar</i>		<i>gerúndio</i>	ArgM-asp
<i>continuar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>continuar</i>	<i>sem</i>	<i>infinitivo</i>	ArgM-asp
<i>correr</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>costumar</i>		<i>infinitivo</i>	ArgM-asp

<i>dar</i>	<i>de</i>	<i>infinitivo</i>	ArgM-asp
<i>deixar</i>	<i>de</i>	<i>infinitivo</i>	ArgM-asp
<i>desatar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>dever</i>		<i>infinitivo</i>	ArgM-mod
<i>disparar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>estar</i>	<i>para</i>	<i>infinitivo</i>	ArgM-asp
<i>estar</i>		<i>gerúndio</i>	ArgM-tml
<i>estar</i>	<i>sem</i>	<i>infinitivo</i>	ArgM-asp
<i>estar</i>	<i>por</i>	<i>infinitivo</i>	ArgM-asp
<i>ficar</i>		<i>gerúndio</i>	ArgM-asp
<i>ficar</i>	<i>sem</i>	<i>infinitivo</i>	ArgM-asp
<i>ficar</i>	<i>de</i>	<i>infinitivo</i>	ArgM-mod
<i>haver</i>	<i>de</i>	<i>infinitivo</i>	ArgM-mod
<i>haver</i>	<i>que</i>	<i>infinitivo</i>	ArgM-mod
<i>haver</i>		<i>particípio</i>	ArgM-tml
<i>ir</i>		<i>infinitivo</i>	ArgM-tml
<i>ir</i>		<i>gerúndio</i>	ArgM-tml
<i>ousar</i>		<i>infinitivo</i>	ArgM-asp
<i>parar</i>	<i>de</i>	<i>infinitivo</i>	ArgM-asp
<i>passar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>permanecer</i>		<i>gerúndio</i>	ArgM-asp
<i>poder</i>		<i>infinitivo</i>	ArgM-mod
<i>recomeçar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>sair</i>		<i>gerúndio</i>	ArgM-asp
<i>seguir</i>		<i>gerúndio</i>	ArgM-asp
<i>ser</i>		<i>particípio</i>	ArgM-pas
<i>ter</i>	<i>de</i>	<i>infinitivo</i>	ArgM-mod

<i>ter</i>	<i>que</i>	<i>infinitivo</i>	ArgM-mod
<i>ter</i>		<i>particípio</i>	ArgM-tml
<i>terminar</i>		<i>gerúndio</i>	ArgM-asp
<i>tornar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>vir</i>		<i>gerúndio</i>	ArgM-tml
<i>vir</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp
<i>viver</i>		<i>gerúndio</i>	ArgM-asp
<i>voltar</i>	<i>a</i>	<i>infinitivo</i>	ArgM-asp

Se o projeto de anotação considerar um verbo como sendo auxiliar, ele será anotado como modificador de um verbo pleno (modificador de modo: ArgM-mod, de aspecto: ArgM-asp, de tempo: ArgM-tml, ou de voz passiva: ArgM-pas). Se, contudo, o projeto de anotação considerar o verbo como um verbo pleno, esse verbo será anotado com sua própria estrutura argumental e seu complemento será o verbo que o sucede na locução verbal.

Por exemplo, se o verbo “poder” for considerado um auxiliar de modo, tem-se a seguinte anotação:

*As deduções só podem **ocorrer** em 2021.*

Arg1: as deduções

ArgM-adv: só

ArgM-mod: podem

REL: ocorrer

ArgM-tmp: em 2001

Contudo, se o verbo “poder” for considerado um verbo com estrutura argumental, tem-se a anotação de dois predicados, “poder” e “ocorrer”:

*As deduções só **podem** ocorrer em 2021.*

Arg0: as deduções

ArgM-adv: só

REL: podem

Arg1: ocorrer

ArgM-tmp: em 2001

*As deduções só podem **ocorrer** em 2021.*

Arg1: as deduções

REL: ocorrer

A anotação morfossintática e sintática do corpus Porttinari-base só foram reconhecidos como auxiliares os verbos SER (auxiliar de passiva, seguido de particípio), TER e HAVER (auxiliares de tempo, seguidos de particípio), IR (auxiliar de tempo, seguido de infinitivo) e ESTAR (auxiliar de tempo, seguido de gerúndio).

Assim, seguindo a anotação sintática do Porttinari-base, no seguinte exemplo, que contém a locução verbal “pode ter de entregar”, há três estruturas argumentais anotadas, “poder”, “ter” e “entregar”:

*Segundo ela, o país norte-americano **pode** ter de entregar os assuntos relacionados a Pyongyang ao Pentágono.*

ArgM-src: segundo ela

Arg0: o país norte-americano

REL: pode

Arg1: ter de entregar os assuntos relacionados a Pyongyang ao Pentágono.

*Segundo ela, o país norte-americano pode **ter** de entregar os assuntos relacionados a Pyongyang ao Pentágono.*

Arg0: o país norte-americano

REL: ter

Arg1: de entregar os assuntos relacionados a Pyongyang ao Pentágono.

*Segundo ela, o país norte-americano pode ter de **entregar** os assuntos relacionados a Pyongyang ao Pentágono.*

Arg0: o país norte-americano

REL: entregar

Arg1: os assuntos relacionados a Pyongyang

Arg2: ao Pentágono.

Na anotação do corpus Porttinari-base decidiu-se seguir o que foi feito no PropBank-Br, versões 1 e 2, ou seja, a anotação dos verbos auxiliares como “modificadores auxiliares”, utilizando as etiquetas ArgM-mod; ArgM-asp; ArgM-tml e ArgM-pas. Contudo, produziu-se também uma versão da anotação seguindo a classificação de auxiliares de modo e aspecto como verbos plenos, com estrutura argumental, seguindo a anotação sintática do Porttinari-base. A anotação que segue diretamente a anotação morfossintática está codificada na versão “UD” do corpus Porttinari-base Propbank e a anotação que segue mais fielmente o conceito de proposição está codificada na versão “clássica” do corpus Porttinari-base PropBank¹⁰.

¹⁰ Em Freitas (2024) estão detalhadas as especificidades das versões “ud” e “clássica”.

Independentemente das versões, é importante sempre considerar o contexto, pois uma mesma forma pode estar sendo usada com diferentes funções, como mostram os dois exemplos a seguir do verbo TER:

*Ela **tem** um carro.* => indubitavelmente, um verbo pleno

*Ela **teve** que comprar um carro* => um verbo modal, que não foi anotado como *auxiliar* na camada morfossintática do Porttinari-base – e, portanto, não anotado como auxiliar na versão “UD” do Porttinari-base PropBank, mas anotado como ArgM-mod na versão “clássica” do Porttinari-base PropBank.

*Ela **tinha** comprado um carro.* => um auxiliar temporal, totalmente gramaticalizado, anotado como *auxiliar* na camada morfossintática do Porttinari-base e anotado como ArgM-tml na versão “clássica” do Porttinari-base PropBank.

3.15 ArgM-cond: Condicionais

A etiqueta ArgM-cond é utilizada para anotar proposições condicionais. No PropBank inglês (2015), orações introduzidas por “se” devem ser anotadas com a etiqueta mais ampla ArgM-adv. Como estes casos constituem um grupo semanticamente relevante e são formalmente marcados em português com as conjunções “se” e “caso” e com o tempo verbal subjuntivo, decidiu-se separá-los de ArgM-adv e anotá-los com uma etiqueta própria: ArgM-cond.

Não podemos continuar se você não me disser a verdade.

Caso seja preciso trocá-los, o serviço custa a partir de R\$ 100

Sofresse ali o 0 a 1, o intervalo seria um inferno e o segundo tempo um sofrimento.

Ainda que o “se” seja uma pista confiável para detectar casos de ArgM-cond, é preciso sempre ter atenção ao contexto. Nos casos abaixo, o “se” é equivalente a “uma vez que” ou “já que”, e seguindo essa leitura, foi anotado como ArgM-cau no Porttinari-base:

Para que pensar, se podemos calar quem se esforça em fazê-lo?

O medo é latente, mas minha consciência diz que, se em 1985 não aconteceu nada, agora também não vai.

O mandante do crime queria contratar um menor de idade com este argumento: “Se no Iraque colocavam crianças para explodir, por que não colocar um adolescente para matar o juiz?”.

Se não colocam em risco a saúde humana ali, por que o fariam no resto do país?

3.16 ArgM-comp: Comparações

Construções envolvendo comparações são anotadas no PropBank (Bonial et al., 2015) como ArgM-ext, etiqueta aplicada a quantidades. Na anotação sintática do Portinari-base, construções comparativas são anotadas como orações adverbiais, assumindo que contêm um verbo implícito. Para os exemplos abaixo,

Brasileiro diz pagar menos propina do que vizinhos de a América Latina

O time tem dez pontos a mais que o segundo colocado em a tabela , o Grêmio .

assume-se que as frases seriam versões reduzidas de

Brasileiro diz pagar menos propina do que pagam vizinhos da América Latina.

O time tem dez pontos a mais que tem o segundo colocado na tabela , o Grêmio .

Na anotação dos papéis semânticos, os advérbios “menos” e “mais” foram anotados como ArgM-ext e o trecho sublinhado com ArgM-comp.

Também foram anotadas como comparativas ArgM-comp construções introduzidas por “assim como”:

Assim como Gustavo Franco , aproximou se de o Partido Novo .

Ela hoje exerce funções administrativas , assim como Jeferson , que foi atacado [...]

3.17 ArgM-src: Fonte da informação

De acordo com as diretrizes do PropBank inglês (Babko-Malaya, 2005; Bonial et al., 2015), sintagmas como os sublinhados a seguir não devem ser anotados.

'It is the livestock sector, according to a new report by the United Nations Food and Agriculture Organization, that generates more greenhouse gases than any other industry.'

[O setor da pecuária, de acordo com um novo relatório da Organização das Nações Unidas para a Alimentação e a Agricultura, gera mais gases com efeito de estufa do que qualquer outra indústria.]

No entanto, considerando a relevância argumentativa/retórica de estruturas como a listada acima no que se refere à detecção da fonte da informação/proposição veiculada, decidiu-se anotá-las com a etiqueta ArgM-src (para indicar “source”, fonte da informação). Construções prototípicas de ArgM-src incluem: “segundo XXX”; “conforme XXX”, “em sua opinião”, “na visão de XXX”, “na opinião de XXX”, “nas palavras de XXX”, “de acordo com XXX”, “do ponto de vista de XXX”.

Segundo o balanço do banco em junho , o BNDES tinha R\$ 173 bilhões em caixa .

De acordo com a polícia , trata-se de uma " prisão significativa " para as investigações .

Na visão feminista atual a música é machista , objetifica a mulher .

Um tremor de terra foi registrado na madrugada desta segunda-feira (18) no Paraná, de acordo com a Rede Sismográfica Brasileira (RSBR).

Incluem também contextos menos óbvios, em que não se trata exatamente da fonte da informação veiculada na proposição, mas da fonte de opinião relativa à proposição veiculada, como os sintagmas introduzidos pela preposição “para” nos exemplos a seguir.

Pergunto-me por que tudo , para o articulista , tem que se resumir a uma dicotomia PT-PSDB ?

Para os homens, vir o barbeiro e cortar o cabelo uma vez durante uma internação longa é suficiente .

Além disso, para os cidadãos cariocas, mais da metade dos moradores do Rio acredita que os usuários de droga são mais responsáveis pela violência na cidade do que os traficantes .

3.18 ArgM-conseq: Consequência

Usada para indicar consequência das proposições. Em termos tradicionais, são as **orações adverbiais consecutivas**. Na anotação original do PropBank, ficam sob o guarda-chuva dos ArgM-adv, decisão justificada pelo fato de se referirem a todo o conteúdo da oração, e não apenas ao verbo. No entanto, decidiu-se separá-las dos ArgM-adv devido tanto à relevância de seu conteúdo (a ideia de “consequência” de uma ação parece interessante de ser codificada de maneira independente dos demais casos de ArgM-adv) quanto à existência de pistas formais, em português, que auxiliam a sua detecção. Nas frases abaixo, o trecho sublinhado é a ArgM-conseq.

Mas a ciranda tecnológica vem girando tão rápido que voltar a atenção primordialmente a os estudantes deixa a já existente força de trabalho esquecida .

A partir do momento em que você fica tão rico que precisa contratar um segurança particular , seu nível de bem-estar cai substancialmente .

Também são anotadas como ArgM-conseq orações apositivas introduzidas por gerúndio que encapsulam todo o conteúdo da oração precedente. Em uma oração desenvolvida, a oração precedente seria referida com um pronome (“isso” ou “o que”):

Deveria ter sido apoteótico , mas foi apenas tépido , indicando o clima geral de a apresentação .
(= o que indica..., isso indica...)

Quatro categorias respondem , em conjunto , por 90 % de a indústria , sinalizando de maneira inequívoca a preferência de o investidor . (= o que sinaliza..., isso sinaliza...)

3.19 Casos ambíguos entre dois ou mais ArgMs

Dependendo do contexto, o léxico associado a um tipo de ArgM pode ser mais adequadamente anotado como outro tipo de ArgM. Exemplos de casos assim, encontrados durante a anotação de papéis semânticos no Porttinari-base são discutidos a seguir.

3.19.1 aqui

Embora na maioria das ocorrências o advérbio “aqui” indique local (ex: “*ela chegou aqui*”), em alguns contextos pode indicar também tempo, e nesses casos será anotado como ArgM-tmp. Em geral, nesses casos, “aqui” será equivalente a “até o momento”, ou “até agora”, mas não será equivalente a “até este local”.

Até aqui , só o ônus foi provado . (ArgM-tmp)

A investigação da Zelotes corre em Brasília e não inclui até aqui o ex-prefeito Luiz Marinho. (ArgM-tmp)

Daqui a 50 anos, esse papo furado servirá para dizer que o MBL era de esquerda, como hoje se diz do nazismo. (ArgM-tmp)

3.19.2 cada vez mais

Embora as expressões que indicam número de vezes sejam anotadas como ArgM-tmp, nas orações abaixo julgou-se que a expressão “cada vez mais” indica quantidade, motivo pelo qual foi anotada como ArgM-ext no Porttinari-base.

Ele tem jogado cada vez mais. (ArgM-ext)

Com a pista secando cada vez mais, [...] (ArgM-ext)

3.19.3 conforme

O sintagma ou oração introduzidos por “conforme” podem introduzir um ArgM-mnr, como nos dois exemplos a seguir:

Ele agiu conforme a ordem que recebeu. (ArgM-mnr)

Ele reagiu conforme esperávamos. (ArgM-mnr)

Mas “conforme” também pode sinalizar a “fonte” da informação (nos contextos em que “conforme” é equivalente a *de acordo com*), e nesse caso será anotado como ArgM-src.

As prefeituras de São Francisco decidiram que sim, conforme se lê em reportagem do jornal britânico "Independent". (ArgM-src)

(...) o convite foi enviado por e-mail, conforme decisão do cerimonial da PGR. (ArgM-src)

3.19.4 depois

Embora na maioria das ocorrências o advérbio “depois” indique tempo (ex: *ela chegou depois*), em alguns contextos pode indicar também local, e nesses casos será anotado como ArgM-loc, como nos exemplos a seguir do Porttinari-base.

Com dificuldades por causa do acidente, Vettel perdeu o controle poucos metros depois e terminou de quebrar o bico de sua Ferrari. (ArgM-loc)

Uruguai (70 %) e Costa Rica (66 %) aparecem depois do Brasil. (ArgM-loc)

3.19.5 diante de

A locução prepositiva “diante de” traz simultaneamente as ideias de local e tempo. Esta polissemia está inclusive codificada na definição de dicionários: “Serve para relacionar palavras por subordinação, dando ideia de tempo, espaço etc”¹¹. Por isso, é preciso ler atentamente a sentença a fim de selecionar a etiqueta mais adequada em cada caso. Os exemplos a seguir mostram decisões de anotação de trechos introduzidos por “diante de” no Porttinari-base.

(...) a desigualdade só cai de forma notável diante de grandes catástrofes sociais como epidemias e guerras. (ArgM-tmp)

(...) o quarteto americano Walk the Moon deve mostrar seu rock alternativo, de guitarras meio frouxas e letras cabeça, diante do provável maior público de sua carreira. (ArgM-loc)

¹¹ <https://www.aulete.com.br/diante>

(...) o árbitro Elmo Alves Resende Cunha validou o gol diante da festa dos mais de 41 mil torcedores corintianos. (ArgM-tmp) (mas ArgM-loc seria igualmente possível)

Seguranças e agentes da GCM (Guarda Civil Metropolitana) agiram para blindar o prefeito diante da aproximação dos manifestantes. (ArgM-tmp)

Parecia em choque diante do fogo, dos gritos e dos feridos. (ArgM-loc)

Conforme o contexto, a locução “diante de” pode, ainda, receber outras etiquetas: ArgM-cau ou ArgM-adv.

As duas foram revogadas diante da ampla repercussão negativa. (ArgM-cau)

Diante da ausência de um plano e de um objetivo nacional, impera a queda dos padrões morais e da noção de responsabilidade individual. (ArgM-cau)

Ele tinha mais cinco meses de mandato, e faz, então, um acordo extremamente vantajoso ao Joesley, de imunidade, diante de um material que levaria à queda do presidente. (ArgM-adv)

3.19.6 geralmente, em geral

Em alguns contextos, o advérbio “geralmente” e a locução adverbial “em geral” podem parecer ambíguos, pois são parafraseáveis tanto por “de maneira geral” (ArgM-mnr) quanto por “na maioria das vezes, quase sempre” (ArgM-tmp). Quando a leitura como “na maioria das vezes, quase sempre” é admissível, como no exemplo a seguir, optou-se por anotar ArgM-tmp no corpus Porttinari-base.

Em geral, as empresas oferecem descontos em a renovação de a apólice. (ArgM-tmp)

3.18.7 jamais, nunca

Em contextos em que **jamais** e **nunca** podem ser interpretados como ArgM-tmp ou ArgM-neg, a etiqueta ArgM-neg tem precedência.

O cara tava na pior, e uísque nunca ajudou nenhum depoente. (ArgM-neg)

O craque da seleção de 1970, por errar demais nos treinos, preferia jamais bater. (ArgM-neg)

Em outros contextos, porém, somente a leitura como ArgM-tmp é possível:

Endiabrados ou não, baladeiros nova-iorquinos passaram anos dançando sem nunca saber da regra. (ArgM-tmp)

3.19.8 tradicionalmente

O advérbio “tradicionalmente” é parafraseável por *de maneira tradicional* (ArgM-mnr) ou por *na maioria das vezes* (ArgM-tmp). No corpus Porttinari-base “tradicionalmente” foi considerado ArgM-tmp

nas duas frases em que apareceu no corpus, uma vez que, nesses contextos, equivaleria a na “maioria das vezes”.

Tradicionalmente associado ao segmento de luxo, o design de embalagens começa a ter papel estratégico para pequenas e médias empresas. (ArgM-tmp)

O Twitter, que tradicionalmente tem participação importante na cobertura da modalidade, é usado para descrever os lances de algumas partidas. (ArgM-tmp)

No entanto, em casos como o exemplo a seguir, “tradicionalmente” equivale a “de maneira tradicional” e, portanto, deve ser etiquetado como ArgM-mnr:

Vestir-se tradicionalmente é algo de que jovens e adultos japoneses se orgulham. (ArgM-mnr)

3.20 Casos Especiais

As decisões de anotação em casos mais complexos merecem ser comentadas.

3.20.1 Modificadores em frases complexas

Quando uma oração tem vários verbos, é preciso certificar-se de que o ArgM refere-se ao verbo que está sendo anotado e não a outro verbo.

Ele acordou e, por um breve instante, não enxergou nada. (Argm-tmp)

Nesse exemplo, o Arg-tmp refere-se ao verbo “enxergar” e não ao verbo “acordar”.

Uma situação mais complicada é quando há dois argumentos que têm o mesmo papel semântico (dois ArgM-tmp, por exemplo). Se eles forem independentes um do outro, cada um deverá receber uma etiqueta de ArgM-tmp, ou seja, os dois não devem ser concatenados sob uma única etiqueta.

Eles se conheceram no verão passado, durante o carnaval.

Arg0: Eles

Arg1: se

REL: conheceram

ArgM-tmp: no verão passado

ArgM-tmp: durante o carnaval.

Contudo, pode ocorrer de o verbo já prever um locativo em sua estrutura argumental. Por exemplo, o Arg2 do verbo “estacionar” é um locativo:

Eu localizei o carro estacionado na rua, em frente a um prédio de apartamentos.

Arg1: o carro

REL: estacionado

Arg2: na rua

ArgM-loc: em frente a um prédio de apartamentos

Nesse caso, tanto “na rua” quanto “em frente a um prédio de apartamentos” podem ser vistos como locativos e satisfazer o Arg2 do verbo “localizar”. A regra que se decidiu usar é anotar como argumento numerado o que estiver mais próximo do verbo e, como argumento modificador, o outro.

3.20.2 Atribuindo rótulos de papéis semânticos em interrogações

O pronome interrogativo deve receber a etiqueta do argumento desconhecido e que se procura conhecer por meio da pergunta. As perguntas *Quem* e *O quê* e suas derivadas referem-se a ArgNs e as perguntas *Onde*, *Quando*, *Como*, *Por quê*, *Quanto* e suas derivadas referem-se a ArgMs.

Do que você gosta?

Arg1: do quê

Arg0: você

REL: gosta

Quem comeu o meu queijo?

Arg0: quem

REL: comeu

Arg1: o meu queijo

Quando você chegou lá?

ArgM-tmp: quando

Arg0: você

REL: chegou

ArgM-loc: lá

De onde você vem?

ArgM-loc: de onde

Arg0: você

REL: vem

Por que não me perguntou isso?

ArgM-cau: por que

ArgM-neg: não

Arg2: me

REL: perguntou

Arg1: isso

De que modo você fez isso?

ArgM-mnr: de que modo

Arg0: você

REL: fez

Arg1: isso

3.20.3 Anotação de Agentes da passiva

Independentemente de se tratar de uma oração passiva desenvolvida ou reduzida de particípio, o Agente da passiva será anotado como Arg0:

Os lacres das embalagens foram rompidos pelos lojistas.

Arg1: os lacres das embalagens

REL: rompidos

Arg0: pelos lojistas

Denunciado por duas vítimas, não quis se pronunciar.

REL: denunciado

Arg0: por duas vítimas

3.20.4 Anotando a partícula “se”

Quando o “se” marcar a supressão de um Agente, independentemente de ser analisado como um caso de voz passiva sintética ou de indeterminação do sujeito, esse “se” deverá ser anotado como Arg0, a fim de sinalizar que o Arg0 está indisponível para ser atribuído a outro argumento da oração. Exemplos:

Mesmo abstraindo-se tudo o que a cidade representa como patrimônio cultural da humanidade, ainda assim ela resulta “paradisiaca”.

ArgM-adv: mesmo

REL: abstraindo

Arg0: se

Arg1: tudo o que a cidade representa como patrimônio cultural da humanidade

Romperam-se os lacres das embalagens.

REL: romperam

Arg0: se

Arg1: os lacres das embalagens

Isso termina por arrastá-la detestavelmente, sem que se chegue a parte alguma.

Arg0: se

REL: chegue

Arg4: a parte alguma

Perdoa-se tudo por amor.

REL: Perdoa

Arg0: se

Arg1: tudo

ArgM-cau: por amor

Pensou-se que não haveria mais confusão.

REL: pensou

Arg0: se

Arg1: que não haveria mais confusão

Fizeram-se várias tentativas e nenhuma deu certo.

REL: fizeram

Arg0: se

Arg1: várias tentativas

Quando o “se” for pronome reflexivo e preencher um argumento previsto na estrutura argumental do verbo, deve-se atribuir-lhe o respectivo ArgN. Exemplos:

Eles se deram as mãos e fizeram as pazes

Arg0: Eles

Arg2: se (um para o outro)

REL: deram

Arg1: as mãos

Ele se culpou por não ter chegado a tempo.

Arg0: Ele

Arg1: se (a si mesmo)

REL: culpou

Arg2: por não ter chegado a tempo

Quando o “se” não se enquadrar em nenhum dos casos acima, ele deve ter seu sentido descrito no Verbo-Brasil como parte de um verbo pronominal, ou seja, a forma pronominalizada do verbo terá um identificador de sentido próprio. São exemplos de “se” anotado como ArgM-nse:

O menino magoou-se com a bronca que levou. (“o menino” = Experienciador = Arg0)

O lixo se acumulou nas calçadas. (“o lixo” = Tema = Arg1)

Ele se orgulha de ter tentado. (“ele” = Experienciador = Arg0)

3.20.5 Utilização não padrão do verbo “tratar”

Construções com “tratar-se de” são construções impessoais, e o “se” é anotado como Arg0. No entanto, nem sempre o texto segue a variante padrão da língua portuguesa, e com frequência as

pessoas colocam sujeito com “tratar-se”. Nesses casos, o sujeito foi anotado como como Arg0, o “se” como ArgM-nse e o complemento de “tratar” como Arg1.

A escultura não se trata de propaganda ou placa fixa que demande parecer da CPPU .

Arg0: a escultura

ArgM-neg: não

ArgM-nse: se

REL: trata

Arg1: propaganda ou placa fixa que demande parecer da CPPU

3.20.6 Verbos que integram locuções conjuntivas ou adverbiais

Combinações como “devido a”, “a seguir” e “a partir de” são analisadas literalmente, e “dever”, “seguir” e “partir” são considerados verbos, o que acarreta a anotação de sua estrutura argumental (“devido ao anúncio”; “a partir de sábado”).

4. Anotando sobre Árvores Sintáticas

4.1 Relação entre elementos sintáticos e papéis semânticos

Normalmente, cada papel semântico corresponde a um único nó da árvore sintática. Em UD, uma gramática de dependências sintáticas, apenas o chamado *head* do constituinte sintático (núcleo da oração, núcleo do sujeito, núcleo do objeto, núcleo do adjunto adverbial etc.) recebe a etiqueta de papel semântico. Nas análises a seguir, os núcleos estão em negrito em cada uma das três estruturas argumentais anotadas, e apenas esses receberão as etiquetas de papéis semânticos.

Ohtake diz que até o compositor e cantor baiano Caetano Veloso manifestou desejo de assistir ao show de Nusrat, na segunda-feira, em São Paulo.

Arg0: Ohtake

REL: diz

*Arg1: que até o compositor e cantor baiano Caetano Veloso **manifestou** desejo de assistir ao show de Nusrat, na segunda-feira, em São Paulo*

*Arg0: o **compositor** e cantor baiano Caetano Veloso*

REL: manifestou

*Arg1: **desejo** de assistir ao show de Nusrat, na segunda-feira, em São Paulo.*

REL: assistir

*Arg1: ao **show** de Nusrat*

*Argm-tmp: na **segunda-feira***

*Argm-loc: em **São Paulo***

No entanto, nem sempre a segmentação produzida pela análise sintática coincide com a segmentação indicada no Verbo-Brasil, e isto pode acontecer devido a decisões derivadas da teoria sintática adotada ou devido a erro na análise sintática automática. Em qualquer um dos casos, a anotação de papéis semânticos no Porttinari-base seguiu a segmentação indicada no Verbo-Brasil. Alguns exemplos dessa não coincidência estão em Freitas & Pardo (2024).

4.2 O argumento pertence ou não à estrutura argumental do verbo-alvo da anotação?

Muita atenção deve ser colocada sobre os ArgMs, pois eles devem ser anotados na estrutura argumental a que se referem. Na frase abaixo, apresenta-se um exemplo de adjunto adverbial de causa que modifica um nome. Embora seja uma causa, não é uma causa relacionada ao verbo que está sendo anotado (“abalar”), ou seja, a locução prepositiva “por ter jogado dopado” é causa da “exclusão de Maradona” e não a causa de “abalou os jogadores”. O mesmo ocorreria se houvesse um adjunto adverbial de tempo ou de lugar modificando o nome “exclusão”, como ilustrado no segundo exemplo.

A expulsão de Maradona, por ter jogado dopado contra a Nigéria, abalou os jogadores”.

Arg0: A expulsão de Maradona, por ter jogado dopado contra a Nigéria,

REL: abalou

Arg1: os jogadores

A expulsão de Maradona, no jogo contra a França, abalou os jogadores”.

Arg0: A expulsão de Maradona, no jogo contra a França,

REL: abalou

Arg1: os jogadores

5. Considerações Finais

A prática da anotação sempre traz casos não previstos na teoria. A confecção de guias de anotação, como este e suas versões precedentes, que registram o aprendizado obtido com a experiência de anotação, tem por objetivo pavimentar, pelo menos em parte, o caminho a ser percorrido pelos anotadores de corpus.

No caso de projetos de anotação que envolvem vários anotadores, o guia funciona como um contrato de anotação, que evita interpretações idiossincráticas, promovendo a concordância entre anotadores.

Mesmo que decisões diferentes venham a ser tomadas por novos projetos, poder contar com um material inicial desse tipo alivia o esforço diante de uma tarefa desconhecida.

As diretrizes aqui descritas já produziram, com mínimas diferenças, três corpus de português brasileiro anotados com papéis semânticos e estão disponíveis para download:

- Propbank-Br: gênero jornalístico, árvores sintáticas de dependências revisadas. Disponível em: http://143.107.183.175:21380/portlex/images/arquivos/propbank-br/propbankbr_v1.1_conll.zip
- Propbank-Br-V.2: gênero jornalístico, árvores sintáticas de dependências não revisadas. Disponível em: http://www.nilc.icmc.usp.br/semanticnlp/index.php?id=index&id_sub=principal&dir_sub=includes/projects/propbankbr&dir=includes/projects/propbankbr&lang=pt-br
- Porttinari-base Propbank: gênero jornalístico, árvores sintáticas de constituintes revisadas. Disponível em: <https://sites.google.com/icmc.usp.br/poetisa/porttinari-base-propbank>

Referências

- BABKO-MALAYA, Olga (2005). PropBank Annotation Guidelines. Disponível em: <https://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf>
- BONIAL, C.; BONN, J.; CONGER, K.; HWANG, J. D.; PALMER, M.; REESE, N. (2015). English PropBank Annotation Guidelines. <https://verbs.colorado.edu/propbank/EPB-Annotation-Guidelines.pdf>
- DE MARNEFFE, M. C.; MANNING, C. D.; NIVRE, J.; ZEMAN, D. (2021). Universal Dependencies. *Computational linguistics*, 47(2):255–308.
- DURAN, M. S. (2010). Manual de anotação do PropBank-Br. Relatório Técnico do NILC - (NILC-TR-06/2010).
- DURAN, M. S. (2014) Manual de anotação do PropBank-Br v2. Disponível em : <http://nilc.icmc.usp.br/semanticnlp/propbankbr/manual.html>
- DURAN, M. S., ALUÍSIO, S. M. (2012). PropBank-Br: a Brazilian Treebank annotated with semantic role labels. In: PROCEEDINGS OF THE 8TH INTERNATIONAL CONFERENCE ON LANGUAGE RESOURCES AND EVALUATION, Istanbul, Turkey.
- DURAN, M. S.; SEPÚLVEDA-TORRES, L.; VIVIANI, M. C., HARTMANN, N. S., ALUÍSIO, S. M. (2014). Seleção e Preparação de Sentenças do corpus PLN-Br para compor o corpus de anotação de papéis semânticos Propbank-Br v2. Relatório Técnico do NILC (NILC-TR-07/2014).
- DURAN, M. S.; LOPES, L.; NUNES, M. G. V.; PARDO, T. (2023) The Dawn of the Porttinari Multigenre Treebank: Introducing its Journalistic Portion. In the Proceedings of the 14th Symposium in Information and Human Language Technology (STIL), pp. 115-124. September, 25-29.
- FREITAS, C. (2024). Anotação de papéis semânticos no *corpus* Porttinari-base: Procedimentos, resultados e análise. Relatório Técnico. São Carlos: ICMC-USP. (no prelo)
- FREITAS, C.; PARDO, T.A.S. (2024). PropBank e anotação de papéis semânticos para a língua portuguesa: O que há de novo?. In the Proceedings of the 15th Symposium in Information and Human Language Technology (STIL). Belém, Novembro, 17-21. (no prelo)
- LOPES, L.; PARDO, T.A.S. (2024). Towards Portparser - a highly accurate parsing system for Brazilian Portuguese following the Universal Dependencies framework. In the Proceedings of the 16th International Conference on Computational Processing of Portuguese (PROPOR), pp. 401-410. May, 13-15.
- PALMER, M.; GILDEA, D.; KINGSBURY, P. (2005). The proposition bank: An annotated corpus of semantic roles. *Computational Linguistics* 31(1), 71–106.

PARDO, T. A. S.; DURAN, M. S.; LOPES, L.; FELIPPO, A. Di; ROMAN, N. T.; NUNES, M. G. V. (2021) Porttinari - a Large Multi-genre Treebank for Brazilian Portuguese. In: Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL), 13. Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021 . p. 1-10.

SILVA, E.H.; PARDO, T. A. S.; ROMAN, N.T. (2023). Etiquetação morfossintática multigênero para o português do Brasil segundo o modelo "Universal Dependencies". In the Proceedings of the 14th Symposium in Information and Human Language Technology (STIL), pp. 63-73. September, 25-29.

Apêndice 1: Advérbios terminados em -mente anotados como ArgM-tmp

anteriormente
anualmente
eventualmente
normalmente
antecipadamente
antigamente
brevemente
comumente
frequentemente
originalmente
periodicamente
raramente
recorrentemente
repetidamente

Apêndice 2: Advérbios terminados em -mente anotados como ArgM-mnr

abertamente	exatamente	normalmente
absolutamente	exclusivamente	oficialmente
acidentalmente	explicitamente	orgulhosamente
adequadamente	extraordinariamente	ostensivamente
aleatoriamente	facilmente	pacificamente
amplamente	fisicamente	parcialmente
ansiosamente	formalmente	penalmente
ativamente	fortemente	ressoalmente
brutalmente	furiosamente	positivamente
burocraticamente	genericamente	praticamente
calmamente	globalmente	preventivamente
carinhosamente	gradualmente	promiscuamente
cientificamente	gratuitamente	propriamente
civilmente	igualmente	publicamente
claramente	ilegalmente	puramente
coletivamente	indefinidamente	rapidamente
completamente	independentemente	reiteradamente
condizentemente	individualmente	remotamente
conscientemente	inequivocamente	respectivamente
constitucionalmente	ininterruptamente	resumidamente
continuamente	intencionalmente	retroativamente
costumeiramente	interesseiramente	sabidamente
cuidadosamente	interinamente	semanalmente
definitivamente	irregularmente	severamente
desnecessariamente	irritantemente	simultaneamente
diferentemente	legalmente	sinceramente
difícilmente	literalmente	sonoramente
diretamente	livremente	totalmente
drasticamente	localmente	unanimemente
duramente	majoritariamente	unilateralmente
efetivamente	manualmente	vergonhosamente
equivocadamente	meticulosamente	visualmente
especialmente	moralmente	voluntariamente
espontaneamente	mundialmente	vulgarmente
essencialmente	musicalmente	

Apêndice 3: Advérbios terminados em -mente anotados como ArgM-dis

basicamente
certamente
curiosamente
diferentemente
difícilmente
felizmente
finalmente
geralmente
globalmente
historicamente
independentemente
infelizmente
lamentavelmente
normalmente

Apêndice 4: Advérbios terminados em -mente anotados como ArgM-adv

aparentemente
basicamente
certamente
claramente
consequentemente
definitivamente
difícilmente
especialmente
felizmente
finalmente
geralmente
infelizmente
invariavelmente
justamente
necessariamente
oficialmente
perfeitamente
possivelmente
praticamente
principalmente
provavelmente
realmente
simplesmente
surpreendentemente