UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

MANUAL DE ANOTAÇÃO DE RELAÇÕES DE DEPENDÊNCIA

MAGALI SANCHES DURAN

Nº 435

RELATÓRIOS TÉCNICOS

São Carlos – SP
Dez./2021

Natural Language Processing initiative (NLP2) of the Center for Artificial Intelligence (C4AI) of the University of São Paulo, sponsored by IBM and FAPESP

POeTiSA

POrtuguese processing – Towards Syntactic Analysis and parsing

Manual de Anotação de Relações de Dependência

Orientações para anotação de relações de dependência sintática em Língua Portuguesa, seguindo as diretrizes da abordagem *Universal Dependencies* (UD)

Magali Sanches Duran

Dezembro/2021

Relatório Técnico do

Núcleo Interinstitucional de Linguística Computacional (NILC)

Agradecimentos:

ao **C4AI**, pelo financiamento desta pesquisa,

à Prof. Dra. **Maria das Graças Volpe Nunes**, pela leitura cuidadosa e sugestões durante o processo de escrita,

à doutoranda **Marcella Monteiro Lemos Couto**, pela colaboração nos primeiros passos da construção deste Manual.

Sumário

Introdução 5

Diretrizes de Anotação de Relações de Dependência 7

acl: adnominal clause = oração adnominal 10

advcl: adverbial clause = oração adverbial 13

advmod: adverbial modifier = modificador adverbial 15

amod: adjectival modifier = modificador adjetivo 19

appos: appositional modifier = modificador apositivo 20

aux: auxiliary = verbo auxiliar 23

case: case marking = marcador de caso 25

cc: conjunction = conjunção 26

ccomp: clausal complement = complemento oracional fechado 27

compound: compound = composto 29

conj = conjunct = coordenado 30

cop: copula = verbo de cópula 32

csubj: clausal subject = sujeito oracional 33

dep: unspecified dependency = dependência não especificada 35

det: determiner = determinante 36

discourse: discourse = discurso 37

dislocated: dislocated = deslocado 39

expl: expletive = expletivo 40

flat: flat = relação plana 41

fixed: fixed = fixa 44

goeswith: goes with = tokens que vão juntos 46

iobj: indirect object = objeto indireto 47

list: list = lista 49

mark: marker = marcador de subordinação 50

nmod: nominal modifier = modificador nominal 52

nsubj: subject = sujeito 54

nummod: numeric modifier = modificador numérico 57

obj: object = objeto direto 59

obl: oblique nominal = nominal oblíquo 61

orphan: orphaned dependent = órfão 65

parataxis: parataxis = parataxis 66

punct: punctuation = pontuação 68
reparandum: overridden disfluency = disfluência 70
root: root = raiz 71
vocative: vocative = vocativo 73
xcomp: open clausal compl. = complemento oracional aberto 74
Bibliografia 78

MANUAL DE ANOTAÇÃO DE RELAÇÕES UD

Introdução

Apresenta-se neste relatório o Manual de Anotação de Relações de Dependência Sintática 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 centro faz parte do Programa de Centros de Pesquisa em Engenharia da FAPESP e está comprometido com o que há de mais moderno em pesquisa em Inteligência Artificial, explorando questões fundacionais e aplicadas. 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, visando alcançar resultados de ponta nesta área, incluindo a produção de um córpus multigênero grande e abrangente, anotado segundo o modelo *Universal Dependencies* (UD) (Nivre, 2015; Nivre at al., 2020).

O modelo UD possui um formato de anotação conhecido como CoNNL-U, constituído de 10 colunas de informações, algumas das quais exigem decisões de anotação. Este Manual tem por objetivo estabelecer diretrizes detalhadas para orientar a anotação de relações de dependência, o que envolve decidir quais são os participantes de uma relação de dependência, qual dos dois é o *head* e qual é o dependente e qual é o nome da relação que os liga. As colunas do CoNLL-U contempladas pelas diretrizes deste Manual são a sétima, a oitava e a nona.

Este Manual tem por complemento o Manual de Anotação de *PoS Tags*[1], publicado em setembro de 2021 na série de Relatórios Técnicos do ICMC sob número 434 e disponível na página do POeTiSA. O Manual de *PoS tags* contempla a quarta coluna do CoNLL-U.

A divisão das diretrizes em dois manuais inspirou-se na divisão observada nas *Guidelines* da UD e numa decisão de projeto que tomamos, de revisar as colunas de anotação por etapas, a fim de que uma etapa pudesse abreviar o esforço requerido na outra etapa. Outro propósito dessa divisão foi facilitar o treinamento dos anotadores contratados, pois a tarefa de revisar *PoS tags* e relações de dependência, separadamente, já é bastante complexa.

A anotação de córpus empreendida pelo POeTiSA visa avançar nas pesquisas nessa frente, dando continuidade a esforços anteriores, como os de construção da Floresta Sintá(c)tica (Afonso et al., 2002) e do Bosque-UD (Rademaker et al., 2017). O córpus anotado resultante desta pesquisa, chamado Porttinari (acrônimo derivado de *PORTuguese Treebank*) (Pardo et al., 2021) deve subsidiar tanto estudos linguísticos como iniciativas de pesquisa e desenvolvimento de *softwares* de análise textual automática, como *taggers* e *parsers* para o português.

A seguir, apresentamos as diretrizes de anotação de relações de dependência adotadas no projeto, que representam a instanciação, em língua portuguesa, das *syntactic relation*s contidas

[1] https://repositorio.usp.br/directbitstream/fc313d66-9ab2-4beb-af80-88d455134cf9/3043575.pdf

nas *Guidelines* da UD[2] e que se destinam a guiar anotadores humanos no processo de anotação de córpus. Mais detalhes, assim como os recursos e ferramentas associados, podem ser encontrados no portal web do projeto, acessível no link a seguir: https://sites.google.com/icmc.usp.br/poetisa.

[2] https://universaldependencies.org/guidelines.html

Diretrizes de Anotação de Relações de Dependência

Em sua versão 2, o esquema de anotação *Universal Dependencies* (UD) possui 37 etiquetas de relações de dependência, as quais são referenciadas por sua abreviação, deprel (de ***dep****endency* ***rel****ation*).

Uma deprel é uma relação que liga dois elementos de uma sentença tal que:

- Um deles é chamado de ***head*** (cabeça, governante ou núcleo da relação) e o outro é chamado de **dependente**;
- um token pode ser *head* de mais de uma relação;
- um token pode ser dependente de uma relação e *head* de outra;
- um token **não** pode ser dependente de mais de uma relação;
- o nome da relação está sempre associado à função que o dependente desempenha em relação ao *head*;
- graficamente, uma seta parte sempre do *head* em direção ao dependente da relação;
- um *head* é sempre uma palavra de conteúdo (verbo, substantivo, adjetivo, pronome, numeral e advérbio). Exceções são símbolos que podem ser expressos por palavras, como R$ (reais), % (por cento), § (parágrafo), etc.
- palavras funcionais (determinantes, preposições, conjunções) e sinais de pontuação, por sua vez, deverão ser sempre dependentes;
- quando o dependente tiver forma oracional, o elemento apontado pela seta será o predicado da oração dependente.

A Figura 1 ilustra uma sentença anotada com relações de dependência.

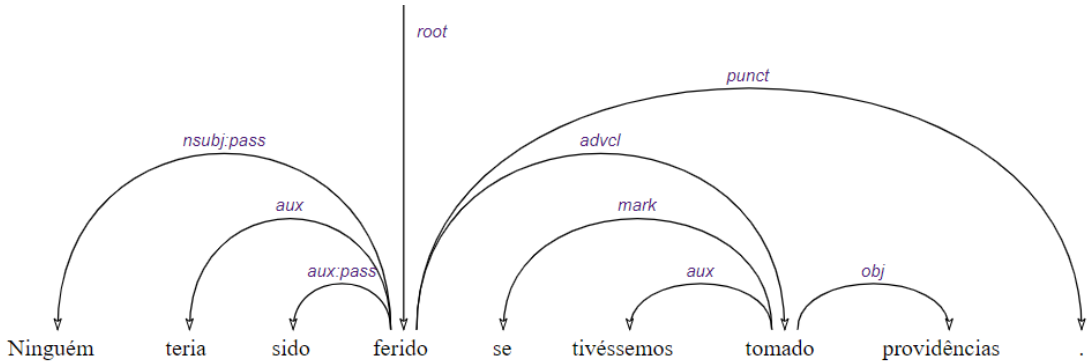


Figura 1 - Exemplo de árvore de dependências anotada com etiquetas de deprel da UD

A atribuição de relações de dependência deve observar o princípio da projetividade, ou seja, os arcos das relações não podem se cruzar. A Figura 2 mostra o que é o cruzamento de arcos, que deve ser evitado a todo custo, mesmo que isso implique uma interpretação menos fiel de alguma relação sintática.

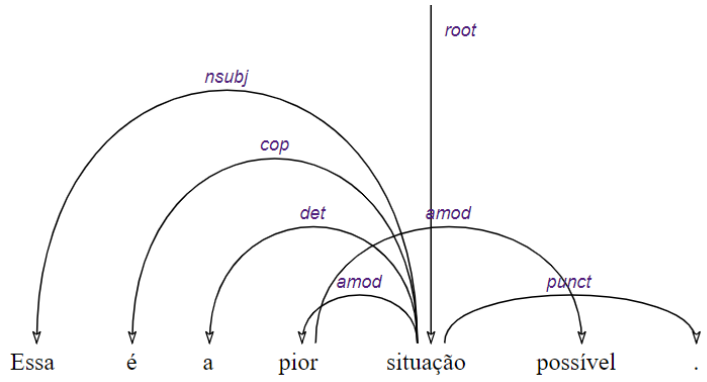


Figura 2 - Exemplo de cruzamento de arcos de relações de dependência

O quadro apresentado na Figura 3, extraído das *Guidelines* da UD[3], mostra as 37 relações de dependência da UD. Esse quadro destaca os argumentos principais dos predicados, chamados de argumentos *core*, separando-os dos demais argumentos considerados não *core*. Separa, também, os argumentos e modificadores de predicados dos modificadores de nominais. Apresenta, além disso, etiquetas diferentes quando o dependente da relação está sob forma oracional (as quais correspondem às orações subordinadas).

	Nominals	Clauses	Modifier words	Function Words
Core arguments	nsubj obj iobj	csubj ccomp xcomp		
Non-core dependents	obl vocative expl dislocated	advcl	advmod* discourse	aux cop mark
Nominal dependents	nmod appos nummod	acl	amod	det clf case
Coordination	**MWE**	**Loose**	**Special**	**Other**
conj cc	fixed flat compound	list parataxis	orphan goeswith reparandum	punct root dep

* The advmod relation is used for modifiers not only of predicates but also of other modifier words.

Figura 3 - Quadro de relações de dependência da UD

Este manual apresenta as deprel por ordem alfabética (para facilitar a busca e consulta de relações específicas), mas o leitor poderá seguir diferentes estratégias para ordenar sua leitura.

[3] https://universaldependencies.org/u/dep/index.html

Uma opção é começar pelas relações *core*, assim chamadas porque têm função central na sintaxe e porque participam de alternâncias sintáticas (mudanças na ordem dos constituintes). Essas relações são as mais importantes para se estabelecer paralelismo entre as línguas, conforme os estudos de Thompson (1997) e Andrews (2007). Há três relações *core* em que os dependentes têm forma nominal[4] (deprel: **subj**, **obj**, **iobj**) e três em que os dependentes têm forma oracional (deprel: **csubj**, **ccomp**, **xcomp**). Nessas deprel, o predicado é o *head* e os argumentos são os dependentes. Não são considerados argumentos *core* os complementos verbais introduzidos por preposição (**iobj** é restrito a pronomes dativos), nem os adjuntos adverbiais.

Em seguida, o leitor pode explorar as relações em que o predicado é o *head* e os dependentes são considerados seus modificadores (**obl**, **advmod**, **advcl**, **vocative**, **expl**, **dislocated**, **discourse**, **aux**, **cop**, **mark**). Apenas uma dessas relações tem forma oracional: **advcl.** A sequência natural dessa ordem de leitura é o conjunto de relações em que o *head* é um nominal e o dependente é um modificador (**nmod, appos, nummod, acl, amod, det, case**). O dependente nessas relações pode ser um nominal, uma palavra funcional (como as preposições, na deprel **case**, e os determinantes, na deprel **det**) ou um numeral (**nummod**). Dessas relações, apenas uma tem forma oracional: **acl**, usada para ligar nominais a orações adjetivas e completivas nominais. Há uma relação de modificador nominal que não é usada no português e por isso não está incluída neste manual: **clf**.

Depois de conhecer as relações que ligam predicados e nominais a seus complementos e modificadores, sugerimos as duas relações que tratam da coordenação de elementos da sentença, **cc** e **conj**.

E, finalmente, após conhecer as relações que têm relação com a sintaxe tradicional, recomendamos a leitura das dez relações artificiais criadas pela UD. A mais importante delas é a deprel **root**, uma relação criada para marcar a raiz da árvore sintática de dependências. Estabelecer o **root** é o primeiro passo para se fazer a anotação de uma sentença. Depois temos as relações **fixed, flat, compound, parataxis, list, orphan, goeswith, reparandum, punct** e **dep**, usadas para anotar tokens que não apresentam relação sintática com outros tokens. Como essas relações não possuem sintaxe, a decisão de qual será o *head* da deprel é arbitrária e normalmente é o primeiro dos dois tokens unidos pela relação.

Outra ordem de leitura que pode ser adotada é a que parte das relações mais frequentes: **case**, **det**, **root**, **punct**, **nmod**, **amod**, **nsubj**, **obj**, **obl**, **advmod**, **nummod**, **conj**, **cop**, **aux**, **cc**, **acl** e **appos**. Essas relações costumam representar cerca de 80% das relações anotadas. As demais relações costumam ser mais raras, respondendo cada uma por bem menos de 1% do total das deprel anotadas. Há de se ressaltar, contudo, a importância das deprel oracionais **csubj, xcomp, ccomp** e **advcl**, independentemente de serem menos frequentes. A frequência das deprel varia um pouco em função do gênero do córpus. Córpus de diálogos, por exemplo, apresentam uma frequência das deprel **vocative** e **discourse** bem acima da média. No entanto, as cinco primeiras deprel da lista apresentada são, com certeza, altamente frequentes em todos os córpus[5].

[4] Neste manual, sempre que nos referirmos a um nominal estamos considerando palavras que podem exercer as funções típicas de substantivos (substantivos, pronomes, adjetivos e numerais com função substantiva).

[5] Podemos nos arriscar a dizer que essas cinco deprel respondem por cerca de 50% de todas as deprel anotadas em um córpus.

acl: adnominal clause = oração adnominal

A deprel **acl** ocorre entre um nominal e uma oração que o modifica. A oração pode apresentar verbos na forma finita ou na forma reduzida (infinitivo, gerúndio, particípio).

Sentido da relação: a relação parte do nominal modificado em direção ao predicado da oração modificadora. A deprel **acl** admite relações nos dois sentidos: da esquerda para a direita e da direita para a esquerda. .

A oração dependente na relação **acl** pode apresentar verbos na voz ativa (Figura 4) ou na voz passiva (Figura 5).

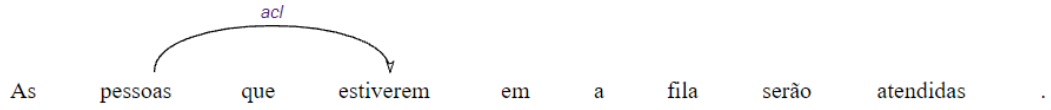


Figura 4 - deprel acl atribuída a oração com verbo na voz ativa

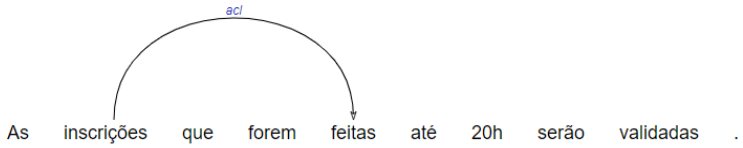


Figura 5 - deprel acl atribuída a oração com verbo na voz passiva

A oração dependente da relação **acl** pode apresentar verbos na forma finita e, nesse caso, sempre será introduzida por um pronome relativo (que, o que, a qual, cujo, etc), como mostra a Figura 6. Nessa figura, nota-se também que a oração dependente pode ter um verbo de cópula e, nesse caso, seu predicado é nominal (no caso, a palavra “político”).

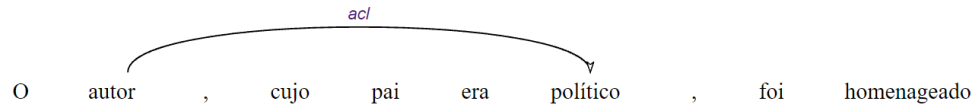


Figura 6 - deprel acl atribuída a oração relativa com verbo de cópula

A oração **acl** pode apresentar também verbos nas formas reduzidas de infinitivo, gerúndio e particípio, como ilustrado nas Figuras 7 e 8 e 9.

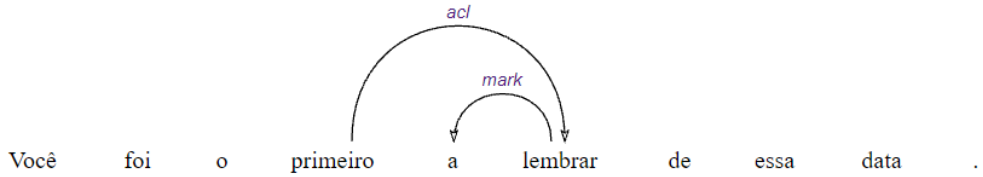


Figura 7 - deprel **acl** atribuída a oração com verbo no infinitivo

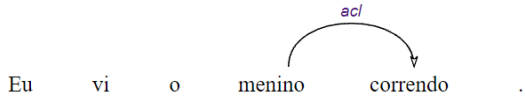


Figura 8 - deprel **acl** atribuída a oração com verbo no gerúndio

Figura 9 - deprel **acl** atribuída a oração com verbo no particípio

Além de ser usada para anotar a relação de um nominal com a oração adjetiva que o modifica, a relação **acl** também é usada para anotar a relação de substantivos, adjetivos e advérbios que possuem estrutura argumental com as orações que os complementam[6]. É o caso, por exemplo, dos substantivos, adjetivos e advérbios negritados nas sentenças a seguir e seus complementos oracionais, destacados em azul:

- Tenho **medo** de não **conseguir** a vaga. (Figura 10)
- A **esperança** de que tudo vai **mudar** nos anima.
- O **anseio** por **voltar** ao normal é geral.
- Estou **envergonhado** de **dizer** isso. (Figura 11)
- Todos ficaram **ansiosos** por **receber** notícias.
- Eu irei, **independentemente** do tempo que **fizer.**
- Chegamos **perto** de **conseguir** mais uma vitória. (Figura 12)

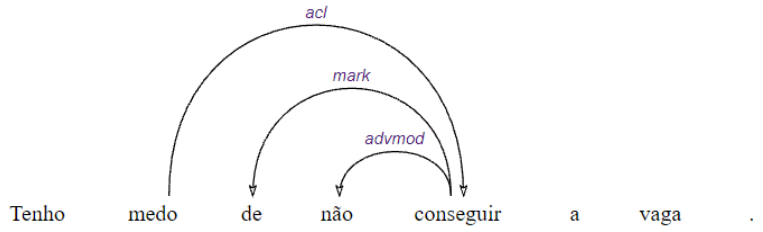


Figura 10 - atribuição da relação **acl** para complemento oracional de substantivo

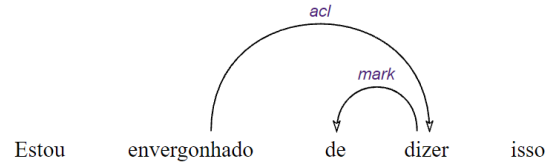


Figura 11 - atribuição da relação **acl** para complemento oracional de adjetivo

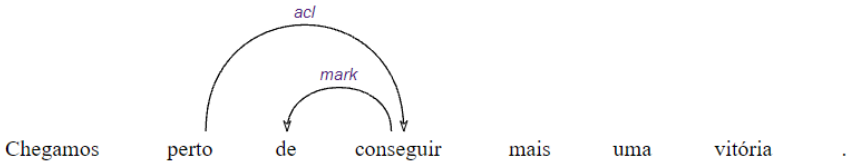


Figura 12 - atribuição da relação **acl** para complemento oracional de advérbio

acl X xcomp

Predicativos do objeto que são opcionais (que podem ser omitidos sem prejuízo da gramaticalidade), como o caso de “vazio” e “nua” nas duas sentenças a seguir, são anotados como dependentes de **acl:**

- **Encontraram** o cofre **vazio**. (O cofre foi encontrado *vazio.)

[6] São conhecidas na Gramática Tradicional como orações subordinadas completivas nominais.

- O artista **pintou** a modelo **nua**. (A modelo foi pintada *nua pelo artista)

Só são anotados como dependentes de **xcomp** os predicativos do objeto que são argumentos previstos na estrutura argumental do verbo, como é o caso de “inocente” e “encerrada” nas duas sentenças a seguir, que são argumentos dos verbo “declarar”:

- Os jurados o **declararam o réu** **inocente**.
- O presidente da comissão **declarou** a sessão **encerrada**.

advcl: adverbial clause = oração adverbial

Os adjuntos adverbiais de causa, tempo, conformidade, concessão, comparação, condição e consequência podem ser expressos sob forma de oração (Figuras 13 e 14). Sempre que isso ocorrer, o verbo da oração adverbial será dependente da deprel **advcl**, cujo *head* é o predicado da oração principal.

Sentido da relação: a relação pode ocorrer nos dois sentidos.

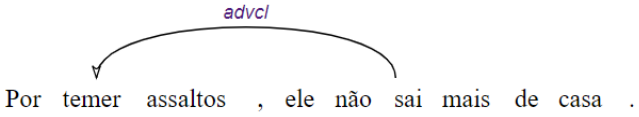


Figura 13 - Oração adverbial causal anotada com a relação **advcl**

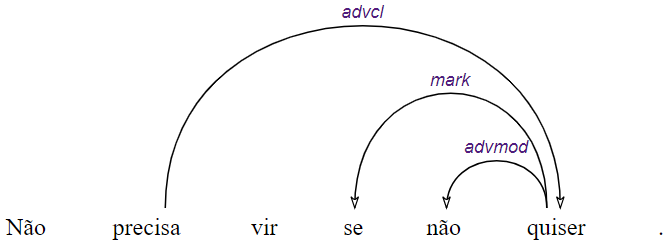


Figura 14 - Oração adverbial condicional anotada com a relação **advcl**

advcl X obl

Algumas vezes a oração adverbial é uma construção com verbo de cópula e, por isso, o predicativo é o dependente. Ocorre, no entanto, que o verbo de cópula às vezes está elíptico. Uma pista importante para identificar esse tipo de oração **advcl** e não confundi-la como um **obl** é a presença de uma conjunção subordinativa introduzindo-a, como é o caso de "embora", na Figura 15. Se o verbo de cópula da oração adverbial não estivesse elíptico, a sentença seria: "Embora **estivesse** no governo, a médica criticou a gestão da pandemia".

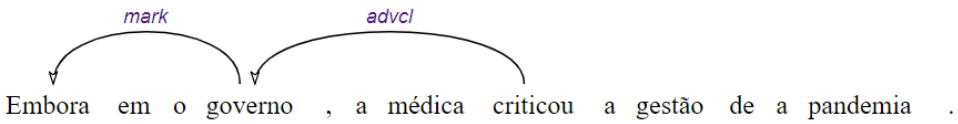


Figura 15 -Relação **advcl** em que o *head* é o predicativo com verbo de cópula elíptico

O mesmo ocorre na Figura 16, em que a conjunção adverbial concessiva "mesmo que" ocorre sem um verbo de cópula, ligando-se diretamente ao predicativo. A sentença, sem elipse do verbo de cópula, seria: "Lula não gosta de discutir, mesmo que **seja** em os bastidores, a chance de não ser candidato".

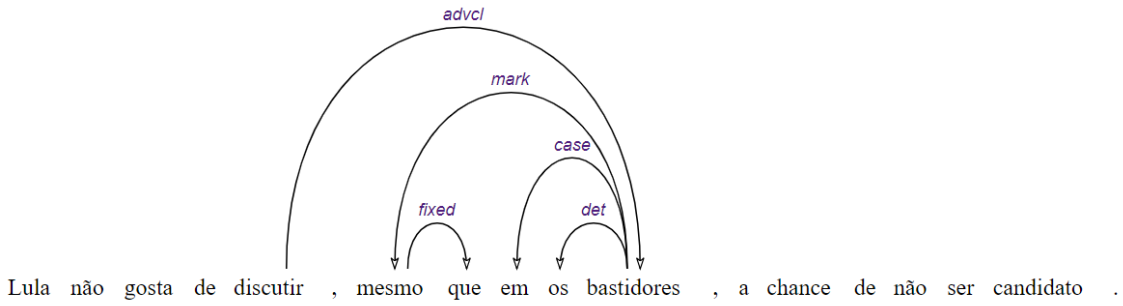


Figura 16 - Oração adverbial concessiva - relação **advcl** com verbo de cópula elíptico

advcl X advmod

Quando a oração adverbial tiver um verbo de cópula elíptico e o predicativo for um advérbio, pode ficar ambíguo se se trata de uma relação **advmod** ou de uma relação **advcl.** Por isso, mais uma vez, uma pista importante para identificar esse tipo de oração é a presença de uma conjunção subordinativa introduzindo-a, como é o caso de "ainda que", conjunção adverbial concessiva, na Figura 17. Se o verbo de cópula não estivesse elíptico, a sentença seria "Ainda que **seja** raramente, ele visita o velho amigo".

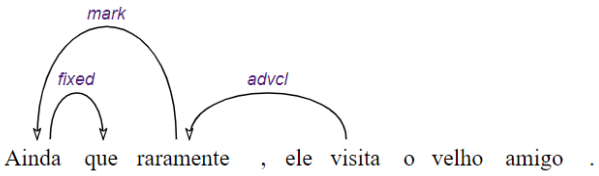


Figura 17 -Relação **advcl** em que o *head* é um predicativo em forma de advérbio

Exemplos de conjunções subordinativas (e locuções conjuntivas) que introduzem orações adverbiais nas quais o verbo de cópula pode ficar elíptico:

- temporais: quando, enquanto
- concessivas: ainda que, nem que, mesmo que, embora
- condicionais: se, desde que

A seguir puxem ser observados mais exemplos de sentenças em que a oração adverbial tem um verbo de cópula elíptico. Em azul está o dependente da relação **advcl** e em negrito, o *head*.

- Enquanto [for] **deputado**, ele **terá** foro privilegiado.
- Quando [é] **aluno**, **tem** que estudar.
- Nem que [seja] **sozinho**, **venha**.
- Mesmo que [estivesse] sem **dinheiro**, **aventurou**-se.

Outro caso de elipse de verbo nas relações advcl ocorre nas orações adverbiais comparativas. Contudo, não é o verbo de cópula que é omitido, mas o próprio verbo pleno da oração subordinada, quando ele coincide com o predicado (verbal ou nominal) da oração principal. Nesses casos, a relação **advcl** liga o **root** à palavra de maior importância da oração subordinada.

- Esse menino **canta** como um **anjo** [canta].
- Macacos **comem** bananas que nem **humanos** [comem].
- Você **gosta** de festa tanto quanto **eu** [gosto].
- A atleta é tão **alta** quanto **você** [é alto/a].

advmod: adverbial modifier = modificador adverbial

A deprel **advmod** liga uma palavra adverbial a um verbo, adjetivo ou advérbio modificado por ela (Figuras 18, 19 e 20). A principal pista para essa deprel é que todo dependente de uma relação **advmod** é anotado com a *PoS tag* **ADV**.

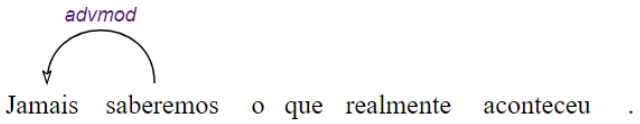


Figura 18 - Atribuição da deprel **advmod** a um *head* verbal

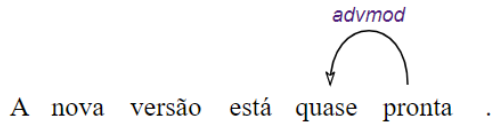


Figura 19 - Atribuição da deprel **advmod** a um *head* adjetivo

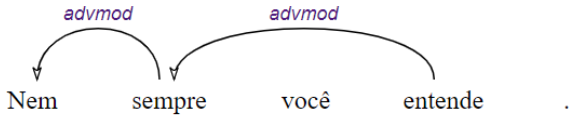


Figura 20 - Atribuição da deprel **advmod** a um *head* advérbio e a um *head* verbal

Sentido da relação: a relação pode ocorrer nos dois sentidos (Figura 21).

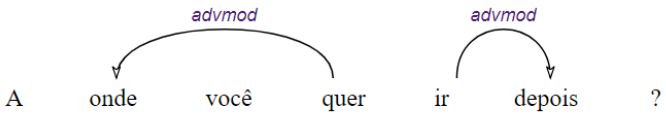


Figura 21 - Advérbios “onde” e “depois” modificando verbos

Atenção: Em raras ocasiões, uma palavra adverbial pode modificar um nominal também (Figuras 22 a 25):

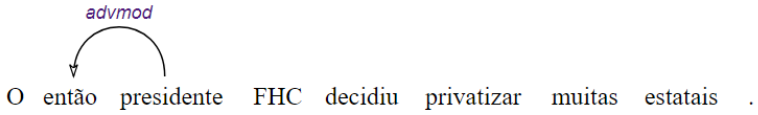


Figura 22 - Atribuição da deprel **advmod** a um *head* nominal (1)

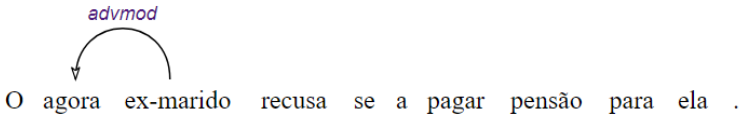


Figura 23 - Atribuição da deprel **advmod** a um *head* nominal (2)

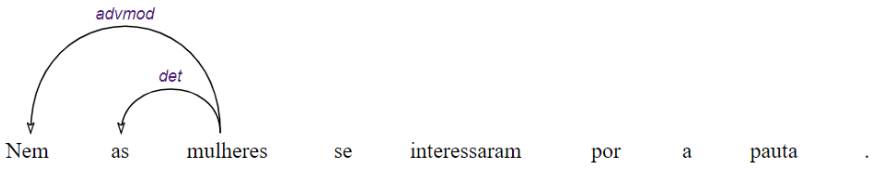


Figura 24 - Atribuição da deprel **advmod** a um *head* nominal (3)

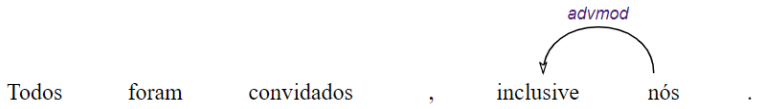


Figura 25 - Atribuição da deprel **advmod** a um *head* nominal (4)

Se um advérbio estiver atuando como predicativo, ou seja, numa construção em que há verbo de cópula, ele será o predicado da oração e não um modificador (Figuras 26, 27 e 28).

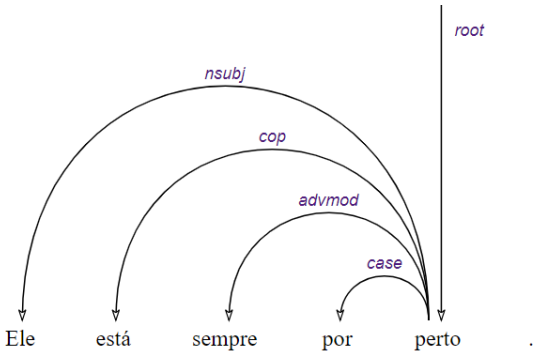


Figura 26 - Advérbio “perto” na função de predicativo

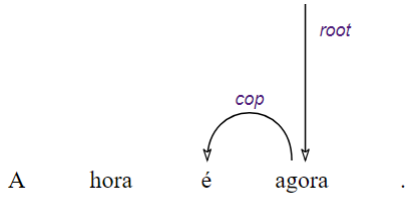


Figura 27 - Advérbio “agora” na função de predicativo

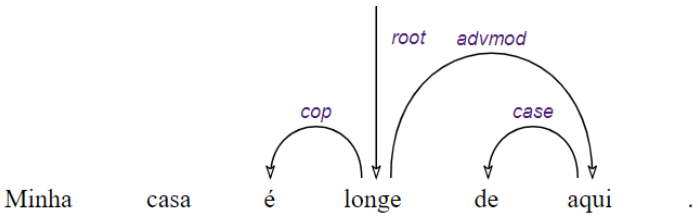


Figura 28 - Advérbio “longe” na função de predicativo

advmod X obl

Tanto **advmod** quanto **obl** ligam adjuntos adverbiais a verbos, adjetivos e advérbios. No entanto, a UD diferencia os adjuntos adverbiais constituídos por advérbios (**advmod**) dos adjuntos adverbiais constituídos por nominais (**obl**). A Figura 29 mostra um adjunto adverbial anotado com **advmod** e a Figura 30 mostra um adjunto adverbial anotado com **obl**.

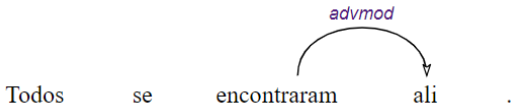


Figura 29 - Exemplo de atribuição de **advmod** a adjunto adverbial de lugar

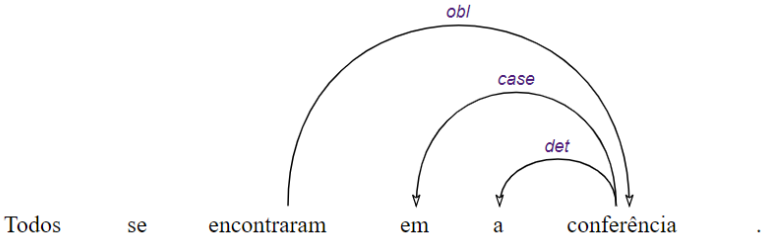


Figura 30 - Exemplo de atribuição de **obl** a adjunto adverbial de lugar

Atenção:

Quase sempre **obl** é introduzido por preposição e **advmod** não é. Mas é possível ocorrer **obl** não introduzido por preposição, como nas Figuras 31 e 32, e **advmod** introduzido por preposição, como na Figura 33.

Geralmente, **obl** não introduzido por preposição ocorre por elipse da preposição. Na Figura 31, "três vezes" corresponde a "**por** três vezes" e na Figura 32, "dia 1º" corresponde a "**em** o dia 1º". O importante é que na deprel **obl** o dependente seja sempre um nominal e na deprel **advmod** o dependente seja sempre um advérbio.

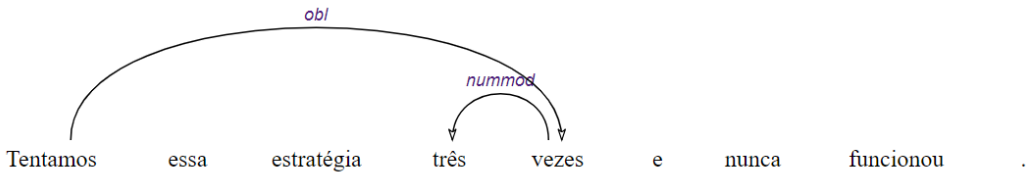


Figura 31 - Dependente não preposicionado de **obl** por motivo de elipse da preposição "por"

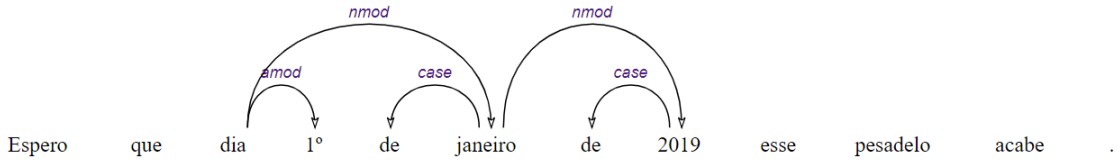


Figura 32[7] - Dependente não preposicionado de **obl** por motivo de elipse da preposição "em"

[7] A anotação de datas nas diversas línguas é um caso de metataxis (uma diferença estrutural genuína entre as línguas). Para maiores detalhes: https://github.com/UniversalDependencies/docs/issues/210

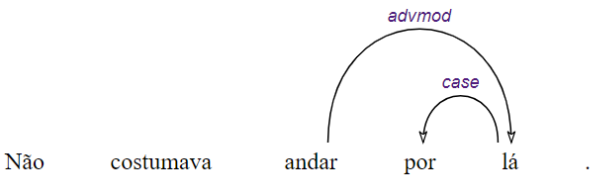


Figura 33 - Dependente preposicionado de **advmod**

amod: adjectival modifier = modificador adjetivo

A deprel **amod** ocorre entre um adjetivo e o elemento que ele qualifica ou classifica.

Sentido da relação: a relação parte do nominal modificado em direção ao adjetivo. A relação admite relações nos dois sentidos (Figuras 34 e 35), embora seja mais comum da esquerda para a direita.

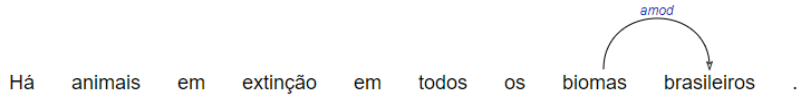


Figura 34 - adjetivo posposto ao substantivo

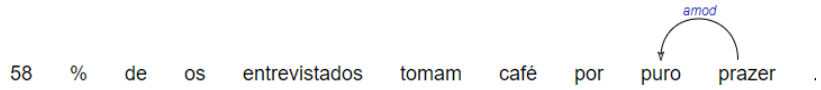


Figura 35 - adjetivo anteposto ao substantivo

A deprel **amod** pode ocorrer entre elementos não contíguos, como ilustrado na Figura 36. Na sentença, “professor - de linguística” e “professor - aposentado” são duas relações independentes, a primeira, **nmod** e a segunda, **amod**.

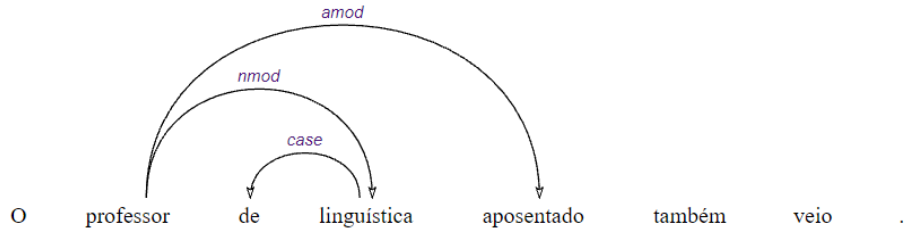


Figura 36 - Adjetivo não contíguo ao substantivo modificado

appos: appositional modifier = modificador apositivo

A deprel **appos** ocorre entre dois nominais que têm o mesmo referente extralinguístico. Ela ocorre entre siglas e suas expressões, explicações entre parênteses e entre vírgulas, por exemplo.

Sentido da relação: a relação parte do primeiro nominal em direção ao nominal à direita, seu aposto.

Importante: a relação **appos** tem sentido fixo: da esquerda para a direita. Se essa direção não for obedecida, o programa que valida a anotação da UD apontará erro.

Em geral, as posições dos dois nominais unidos por **appos** podem ser trocadas, como pode ser observado pelas Figuras 37 e 38.

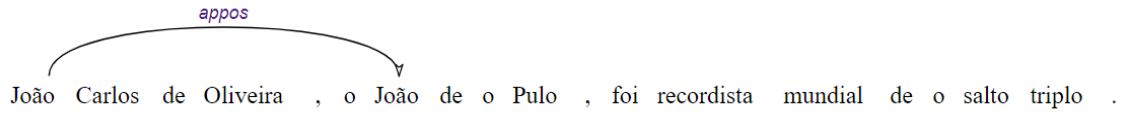


Figura 37 - relação **appos** com nome à esquerda e codinome à direita

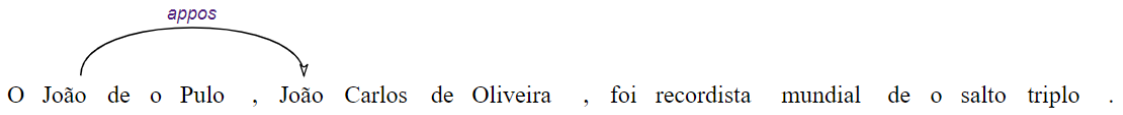


Figura 38 - relação **appos** com codinome à esquerda e nome à direita

Essa deprel pode ocorrer entre vírgulas, imediatamente após um nominal, e serve para definir, nomear ou descrever esse nominal (Figuras 39 e 40).

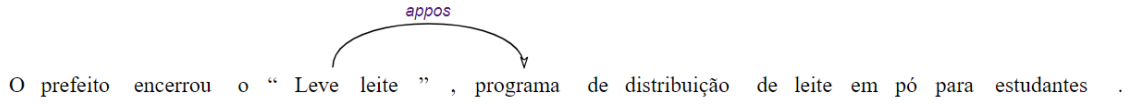


Figura 39 - relação **appos** com nome à esquerda e descrição à direita

Figura 40 - relação **appos** com definição à direita

A deprel **appos** ocorre também quando há exemplos ou definições entre parênteses (Figura 41).

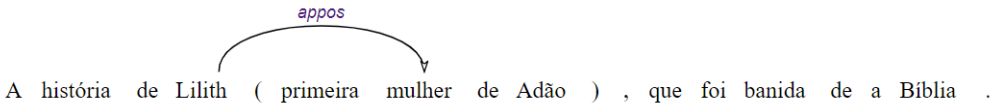


Figura 41 - relação **appos** com descrição entre parênteses

Nos raros casos de mais de um nominal apositivo, todos à direita do primeiro devem ser marcados como **appos** do primeiro nominal, como mostrado na Figura 42.

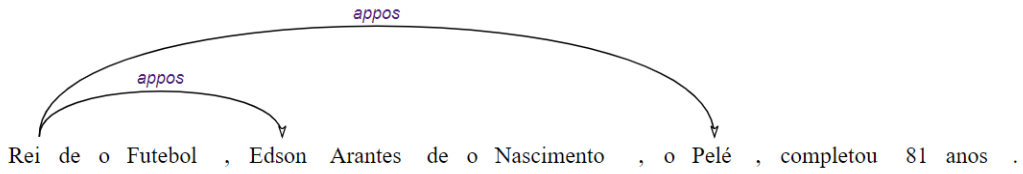


Figura 42 - Três nominais com o mesmo referente extralinguístico

A deprel **appos** serve também para vincular pares em endereços, assinaturas e afins, como mostra a Figura 43.

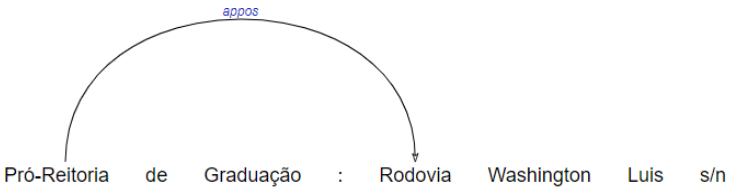


Figura 43 - Órgão e respectivo endereço à direita

appos X nmod

É preciso cuidado para não confundir essas duas deprel. Se dois nominais em sequência são intercambiáveis, trata-se de **appos** (Figura 44).

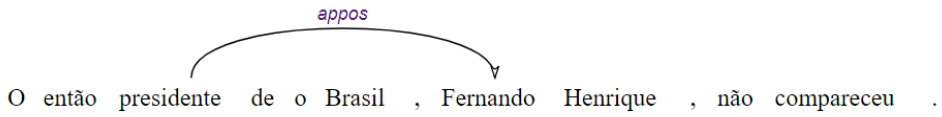


Figura 44 - Dois nominais intercambiáveis são **appos**

Na Figura 44, poderíamos ter a inversão da ordem dos nominais: “Fernando Henrique, o então presidente do Brasil, não compareceu”.

Contudo, se a ordem de dois nominais for fixa[8], trata-se de **nmod** (Figura 45). A inversão da ordem dos nominais nessa sentença ficaria, no mínimo, estranha, principalmente por não haver

[8] Essa diferenciação da UD diverge das gramáticas tradicionais do português, para as quais um nome especificador é aposto do nome especificado.

dois constituintes separados por vírgula: “Fernando Henrique o ex-presidente não compareceu”.

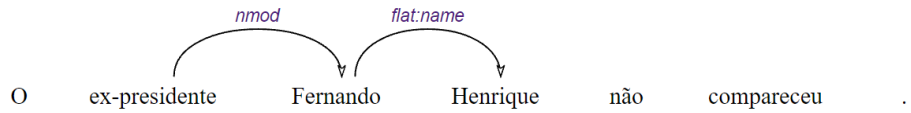


Figura 45 - Dois nominais em sequência, não intercambiáveis, são **nmod**

Também são **nmod e** não **appos**, os casos em que um nome próprio especifica um substantivo (*head* da relação em negrito e dependente em azul):

- a **cantora** **Ivete** Sangalo
- o **goleiro** **Rogério** Ceni
- o **filme** **Armageddon**
- o **jornal** **Folha** de São Paulo

aux: auxiliary = verbo auxiliar

A relação **aux** liga um verbo auxiliado a seu auxiliar. Em nosso projeto, anotamos como auxiliares apenas os auxiliares de tempo e o auxiliar de voz passiva.

ter + verbo auxiliado no particípio	Ex: Ele **tinha previsto** isso.
haver + verbo auxiliado no particípio	Ex: **Havíamos recebido** um aviso no dia anterior.
ir + verbo auxiliado no infinitivo	Ex: **Vamos fazer** uma festa amanhã.
estar + verbo auxiliado no gerúndio	Ex: **Estão tentando** de tudo para salvar vidas.
ser + verbo auxiliado no particípio	Ex: Isso **foi feito** sem autorização.

Sentido da relação: a relação parte do verbo auxiliado em direção ao verbo auxiliar. A relação é unidirecional, da direita para a esquerda, como ilustra a Figura 46.

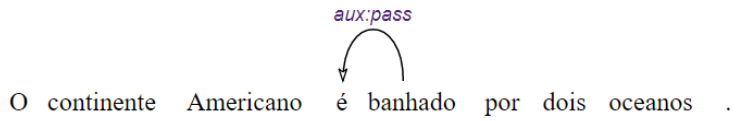


Figura 46 - Atribuição da relação aux:pass

Quando o auxiliar é de voz passiva, existe uma sub-relação **:pass** que pode ser adicionada ao nome da deprel, como mostrado acima nas Figura 46 e 47.

Pode haver tokens entre o verbo auxiliar e o verbo auxiliado, ou seja, nem sempre auxiliar e auxiliado são contíguos, como mostra a Figura 47. Essa figura mostra também dois auxiliares modificando um mesmo verbo.

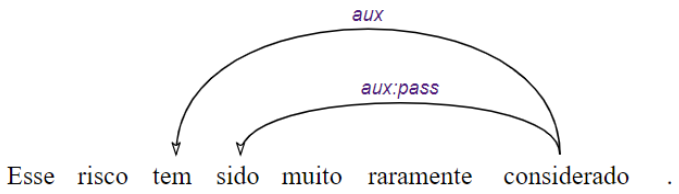


Figura 47 - Sentença com dois verbos auxiliares não contíguos com o auxiliado

O verbo “ser”, quando anotado com a PoS tag AUX, pode ser tanto um verbo auxiliar de passiva (dependente da deprel **aux**) quanto um verbo de cópula (dependente da deprel **cop**). Nos casos em que o verbo é seguido de um particípio, a situação fica ambígua. Se o particípio for um substantivo ou adjetivo na função de predicativo, o verbo “ser” será dependente de **cop**, mas se o particípio for um verbo, o verbo “ser” será dependente de **aux**. Outra pista para desambiguar as duas situações é a presença de um agente da passiva, que indica se tratar de verbo auxiliar (Figura 48) e não de cópula.

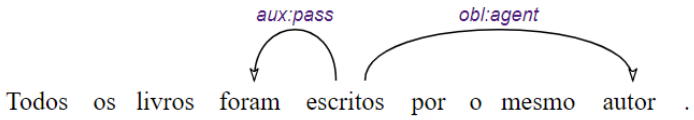


Figura 48 - Sentença com verbo “ser” auxiliar de passiva e agente expresso

case: case marking = marcador de caso

A deprel **case** liga uma preposição ao elemento que ela introduz (pode ser substantivo, adjetivo, pronome ou advérbio).

Sentido da relação: a relação parte do elemento introduzido em direção à preposição (PoS **ADP**), ou seja, é sempre unidirecional da direita para a esquerda.

As Figuras 49, 50 e 51 exemplificam a atribuição da deprel **case**.

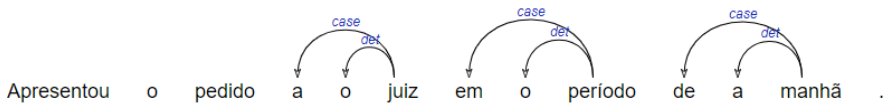


Figura 49 - Atribuição da deprel **case** a NOUN

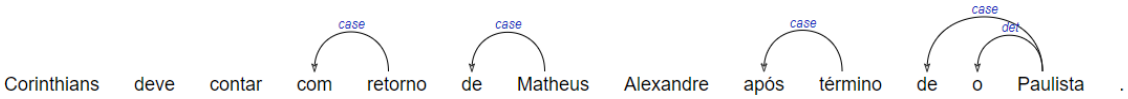


Figura 50 - Atribuição da deprel **case** a NOUN e PROPN

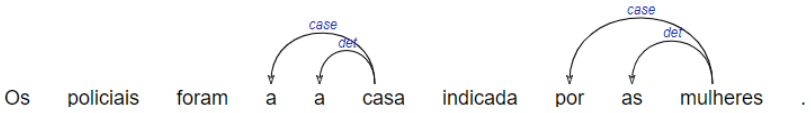


Figura 51 - Atribuição da deprel **case** a tokens de contrações

case X mark

Quando uma preposição introduzir uma oração subordinada reduzida de infinitivo 1, deverá ser anotada como **mark**, como ilustrado na Figura 52.

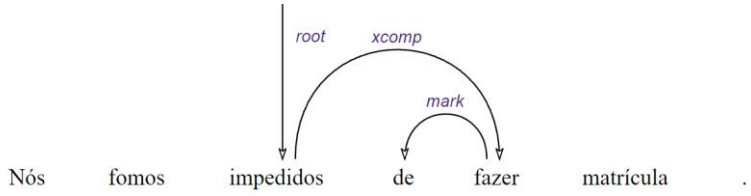


Figura 52 - Atribuição da deprel **mark** a preposição

cc: conjunction = conjunção

A deprel **cc** ocorre entre um elemento coordenado e a conjunção que o introduz numa coordenação.

Sentido da relação: a relação parte do elemento coordenado em direção à conjunção que o precede. A relação, portanto, é unidirecional da direita para a esquerda (Figuras 53 e 54).

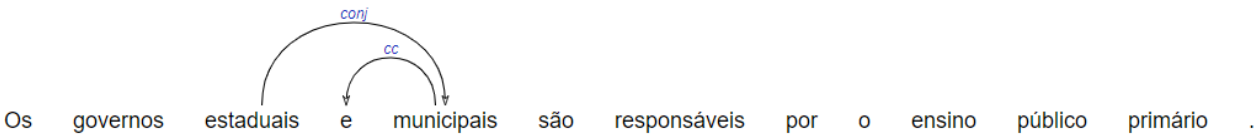


Figura 53 - Exemplo de deprel **cc** atribuída a conjunção coordenativa aditiva

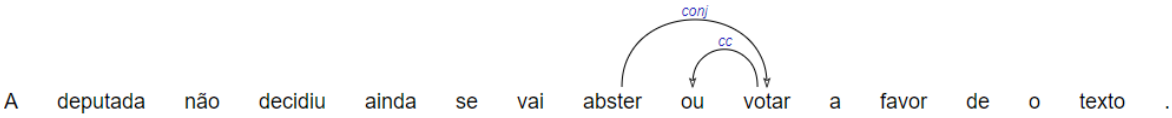


Figura 54 - Exemplo de deprel **cc** atribuída a conjunção coordenativa alternativa

Sempre que há uma conjunção coordenativa, é bem provável que haja coordenação entre dois ou mais elementos e, portanto, a atribuição da deprel **conj**. Porém, é comum haver sentenças que apresentam conjunções que se relacionam com sentenças anteriores do texto.

Como na UD só anotamos no nível de sentenças, a deprel **cc** ocorre nesses casos sem a correspondente deprel **conj**, como mostram as Figuras 55 e 56. Isso ocorre também com todas as conjunções que iniciam orações: “porém”, “todavia”, “entretanto”, etc.

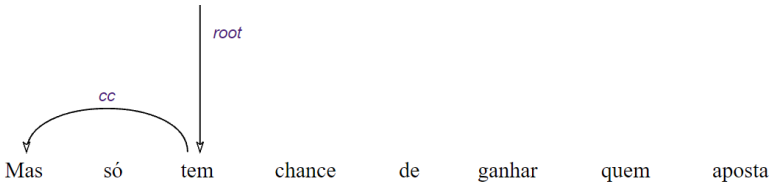


Figura 55 - Exemplo de deprel **cc** atribuída a conjunção coordenativa adversativa

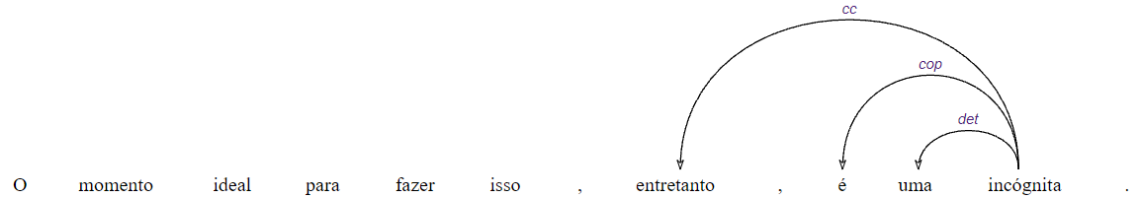


Figura 56 - Exemplo de deprel **cc** atribuída a conjunção coordenativa adversativa entre vírgulas

ccomp: clausal complement = complemento oracional fechado

Um complemento oracional fechado é um argumento *core* que complementa o sentido do predicado (verbal ou nominal) e possui um sujeito próprio, expresso ou não. Ter "sujeito próprio" significa ter um sujeito não "controlado" nem pelo sujeito nem pelo objeto da oração principal .

Sentido da relação: a relação ocorre da esquerda para a direita.

No caso do exemplo ilustrado na Figura 57, o sujeito de "disse" é "ele" e o sujeito de "quer" é "você", ou seja, as duas orações têm sujeitos diferentes. Além disso, o complemento do verbo "dizer" é a oração "que você quer falar comigo". Por isso a relação entre o predicado principal "disse" e o predicado da oração subordinada "quer" é uma relação do tipo **ccomp**.

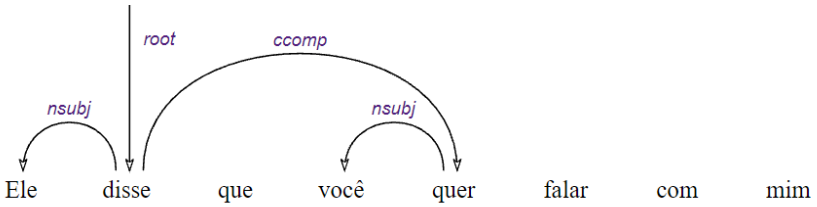


Figura 57 - Exemplo de atribuição de **ccomp** com dois sujeitos diferentes explícitos

Orações subordinadas desenvolvidas que complementam o sentido de verbos e que são introduzidas pelas conjunções subordinativas **que** ou **se** são **ccomp** (Figura 58).

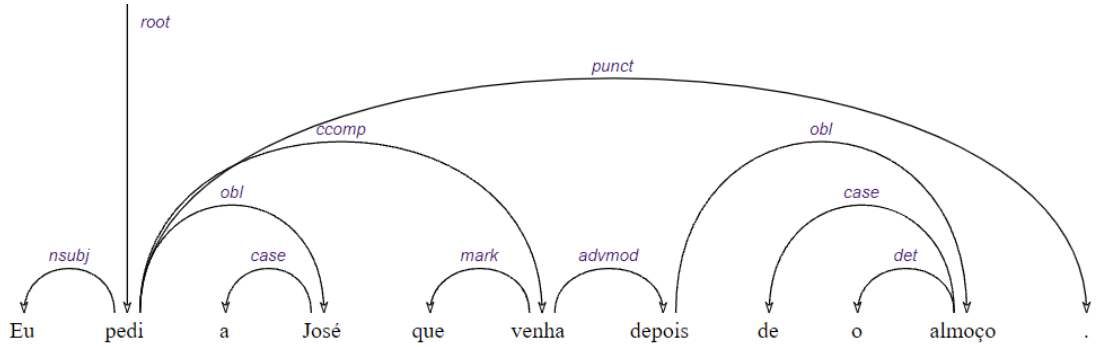


Figura 58 - Exemplo completo de atribuição de **ccomp** a oração desenvolvida

No caso ilustrado pela Figura 59, não há um **obj** de "pedi", apenas um **obl**, por isso a oração "vir" é relacionada a "pedi" com a deprel **ccomp**. O sujeito da oração subordinada poderia ser outro que não "José": "Eu pedi a José para **sua filha** vir depois do almoço"

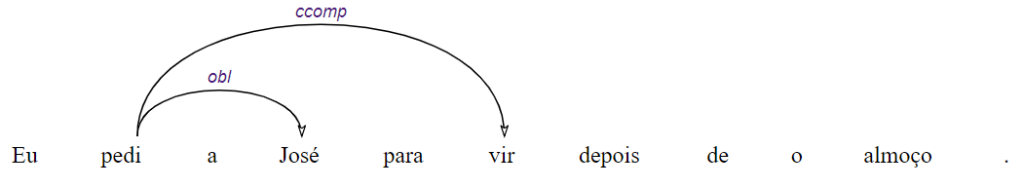


Figura 59 - Exemplo de atribuição de **ccomp** a oração reduzida de infinitivo

Outro uso de **ccomp** é nos casos em que os dois lados de uma construção com verbo de cópula são equivalentes (Figura 60). Dizemos que os dois lados são equivalentes porque poderiam ser trocados de lugar: "O mais importante **é** manter a calma" - "Manter a calma **é** o mais

importante”. Nesses casos, o verbo de cópula deverá ser **root** e, por isso, sua *PoS tag* deverá ser alterada de AUX para VERB.

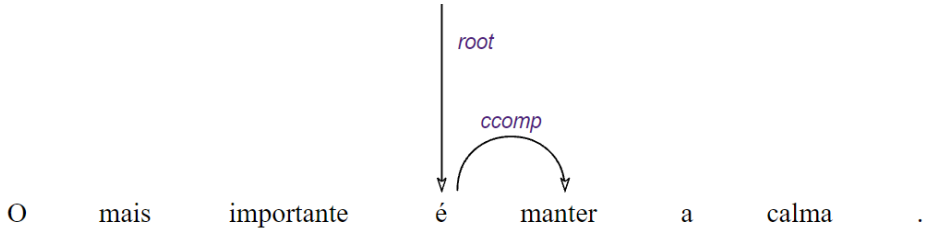


Figura 60 - Exemplo de atribuição de ccomp a verbo de cópula e oração reduzida de infinitivo

A oração dependente pode apresentar o verbo em uma forma reduzida, como na Figura 60 (acima), ou em uma forma finita, como na Figura 61 (abaixo).

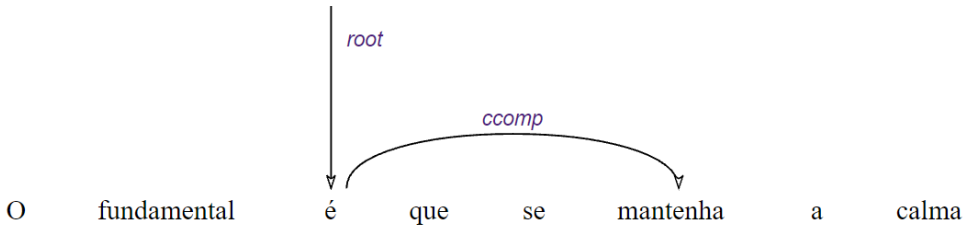


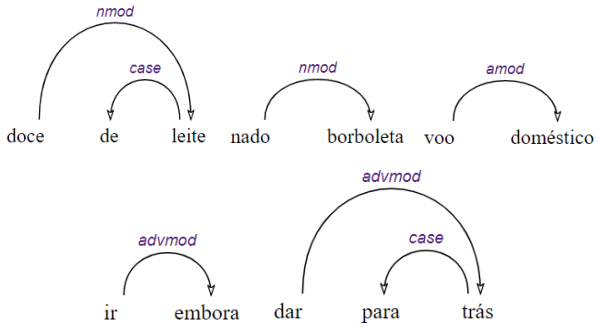
Figura 61 - Exemplo de atribuição de **ccomp** a verbo de cópula e oração desenvolvida

compound: compound = composto

A deprel **compound** foi criada para relacionar multipalavras de conteúdo (substantivos, adjetivos, verbos e advérbios) que não apresentam relação sintática entre os tokens que as compõem. No português, a princípio, não temos casos que requeiram essa deprel. As palavras compostas normalmente usam hífen.

Sentido da relação: da esquerda para a direita, como todas as demais relações não sintáticas da UD (decisão arbitrária)

A anotação sintática, sempre que existente, deverá ser priorizada. Nos exemplos abaixo, há casos de multipalavras unidas pela relação **nmod**, com preposição (doce de leite) e sem preposição (nado borboleta), pela relação **amod** (voo doméstico), e pela relação **advmod**, com preposição (dar para trás) e sem preposição (ir embora[9]).



A decisão de utilizar **compound** deverá ser feita em nível de projeto e uma lista das multipalavras que a utilizarão deverá ser fornecida. [10]

[9] nos verbos "ir embora", "vir embora", "mandar embora", "dar embora", a palavra "embora" constitui um advérbio (ADV).

[10] https://universaldependencies.org/u/dep/compound.html "Each language that uses compound should develop its own specific criteria based on morphosyntax (rather than lexicalization or semantic idiomaticity)"

conj = conjunct = coordenado

A deprel **conj** ocorre entre dois ou mais elementos coordenados. A coordenação pode ser marcada pela presença de uma conjunção (ex: *e, ou, mas*) ou não. Os elementos unidos pela relação **conj** precisam ser da mesma natureza: dois ou mais substantivos, adjetivos, predicados, etc. A única exceção a essa regra é quando o segundo elemento de uma coordenação está elíptico e, por isso, é substituído na relação por um dependente órfão[11].

Sentido da relação: a relação parte do primeiro elemento da coordenação em direção a cada um dos demais elementos. A relação, portanto, é sempre unidirecional da esquerda para a direita.

Figura 62 - Dois nomes próprios coordenados pela deprel **conj**

A coordenação pode ser assindética, o que significa que a conjunção coordenativa (deprel **cc**) é omitida, como na primeira deprel **conj** da Figura 63. Essa figura ilustra também que, entre três elementos coordenados, podem ocorrer coordenações assindéticas e sindéticas.

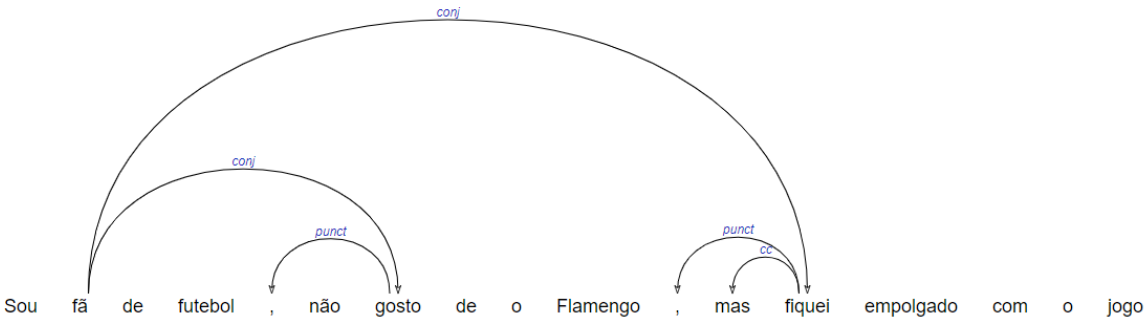


Figura 63 - Coordenação assindética e sindética entre predicado nominal e predicados verbais

Como mostra a Figura 64, mesmo quando há mais de dois elementos coordenados no mesmo nível, a relação **conj** mantém-se entre o primeiro elemento e cada um dos demais.

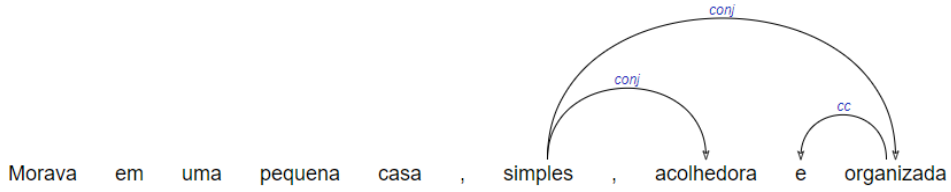


Figura 64 - Coordenações assindética e sindética entre três adjetivos

Contudo, se um elemento possui coordenação do tipo aditivo com um elemento e do tipo alternativo com outro elemento, é preciso observar a correta anotação para ser fiel ao sentido, como mostra a Figura 65. Nessa figura, “direitos e emprego” e “direitos e desemprego” são

[11] Consultar a deprel **orphan** para ver exemplo dessa situação.

coordenações aditivas diferentes, marcadas pela conjunção “e”, enquanto entre “direitos” e “direitos” há uma coordenação alternativa, marcada pela conjunção “ou”.

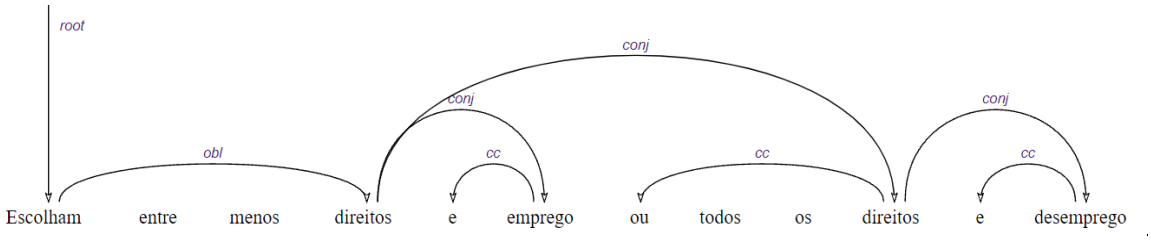


Figura 65 - Emprego de **conj** em dois tipos de coordenação, aditiva e alternativa

cop: copula = verbo de cópula

Um verbo de cópula serve para ligar o sujeito de uma oração a um predicado não verbal. No português, há dois verbos de cópula considerados pelo esquema de anotação UD: o verbo “ser” e o verbo “estar”. A relação **cop** une o predicado não verbal ao verbo de cópula, como mostram as Figura 66 e 67.

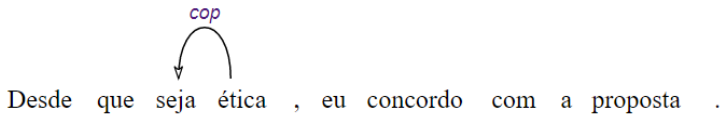


Figura 66 - Relação **cop** entre predicado não verbal e verbo de cópula

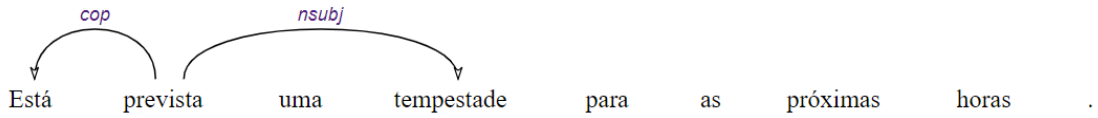


Figura 67 - Relação **cop** entre **root** não verbal e verbo de cópula

Sentido da relação: a relação parte do predicado não verbal (predicativo) em direção ao verbo de cópula. A relação é unidirecional, da direita para a esquerda.

Os verbos de cópula são anotados no nível de *PoS tag* com a mesma etiqueta dos verbos auxiliares e isso pode gerar ambiguidade no caso do verbo "ser", pois ele funciona tanto como verbo de cópula (Figura 68) quanto como auxiliar de passiva (Figura 69). O problema ocorre especificamente com relação a particípios que acompanham o verbo "ser": se eles forem anotados com *PoS tag* de verbo, trata-se de voz passiva e o verbo "ser" é dependente de **aux**; se eles forem anotados com *PoS tag* de substantivo ou adjetivo, trata-se de predicativo e o verbo "ser" é dependente de **cop**.

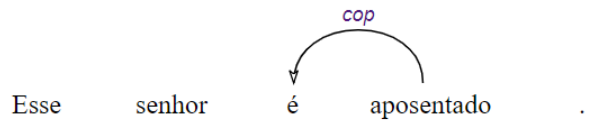


Figura 68 - Relação **cop** em que o predicado é um substantivo em forma de particípio

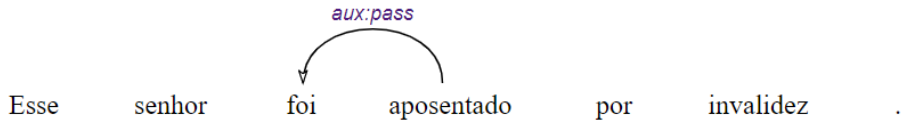


Figura 69 - Relação **aux** em que o particípio do verbo apassivado é ambíguo com substantivo

csubj: clausal subject = sujeito oracional

A deprel **csubj** é utilizada para anotar o sujeito constituído de uma oração. O *head* da relação **csubj** é o predicado da oração principal e o dependente é o predicado da oração subordinada (oração-sujeito) .

Sentido da relação: a relação parte do predicado da oração principal em direção ao predicado da oração subordinada, não importando se está à esquerda ou à direita (Figura 70).

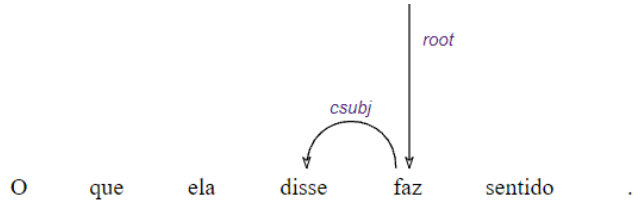


Figura 70 - Exemplo da deprel **csubj** cujo *head* é predicado verbal

Na Figura 70 (acima), o **root**, predicado da oração principal, é o verbo "faz" e o sujeito desse verbo é a oração "o que ela disse", cujo predicado é "disse". Assim, a deprel **csubj** une "faz" (*head* da relação) a "disse" (dependente da relação).

Se a oração principal tem um verbo de cópula, o predicado é nominal. Assim, o predicado da oração principal do exemplo da Figura 71 é a palavra "interessante" e o predicado da oração subordinada (oração-sujeito) é o verbo "disse". A deprel **csubj** liga o *head* da relação, "interessante", ao dependente da relação, "disse".

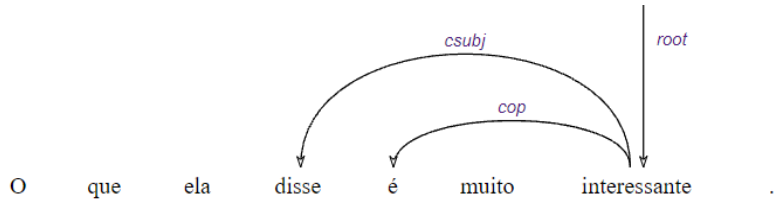


Figura 71. - Exemplo da deprel **csubj** cujo *head* é predicado nominal

O sujeito oracional pode ocorrer antes ou depois da oração principal e alguns verbos apresentam construções típicas em que o sujeito oracional é posposto ao verbo da oração principal. Por exemplo: "consta que…", "convém que…", "acontece que…", "parece que…", "pode ser que". Essa situação é ilustrada nas Figuras 72 e 73.

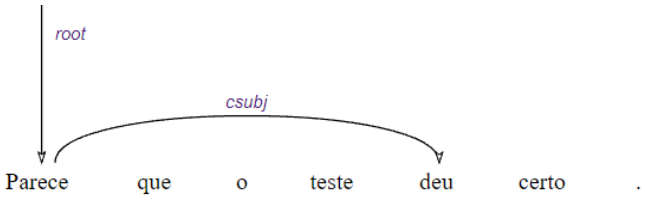


Figura 72. - Exemplo da deprel **csubj** a oração posposta ao verbo principal

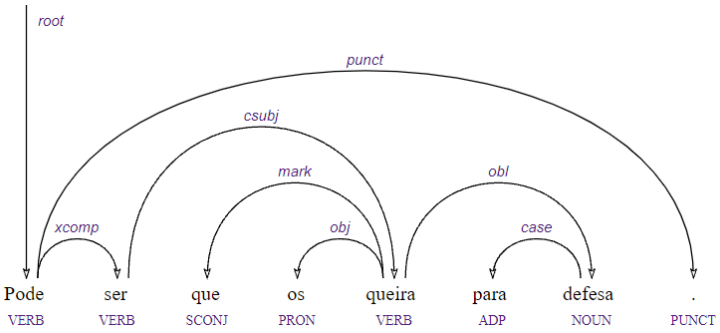


Figura 73 - Exemplo com anotação completa de sentença com a deprel **csubj**

Outra forma comum de sujeito oracional é aquele em que um verbo de cópula e um predicativo ocorrem no início da oração. Por exemplo: “é natural que…”, “está claro que…”, “ é inimaginável que…”, como na Figura 74.

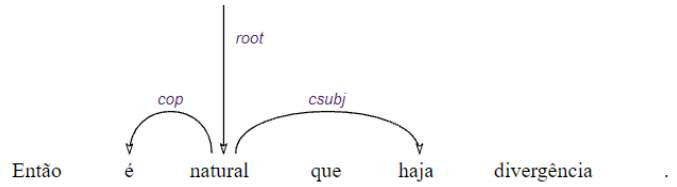
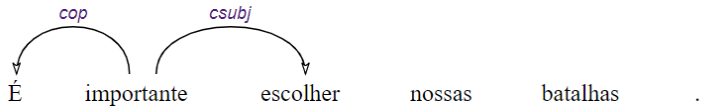


Figura 74 - Exemplo da deprel **csubj** ligada a predicado nominal

Sujeitos oracionais podem ocorrer sob a forma de um verbo no infinitivo (Figura 75)



75 - Exemplo da deprel **csubj** a oração reduzida de infinitivo

Quando o sujeito oracional for sujeito de uma construção passiva, ele deverá ser anotado como **csubj:pass**, como ilustrado na Figura 76.

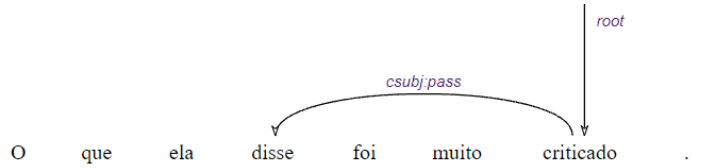


Figura 76 - Exemplo da deprel **csubj** atribuído a construção de voz passiva

dep: unspecified dependency = dependência não especificada

A deprel **dep** ocorre entre dois elementos quando nenhuma das relações previstas na lista de deprel for considerada adequada. As relações anotadas com **dep** serão revistas a fim de se discutir como deveriam ser interpretadas para se enquadrarem em uma das deprel previstas.

Sentido da relação: da esquerda para a direita, como todas as demais relações não sintáticas da UD (decisão arbitrária)

det: determiner = determinante

A deprel **det** ocorre entre um nominal e seus determinantes. A palavra anotada com *PoS tag* **DET** sempre será dependente da deprel **det**.

Sentido da relação: a relação parte do nominal em direção ao determinante (*PoS tag* **DET**). A relação admite duas direções: da direita para a esquerda (mais frequente), como mostrado na Figura 77, e da esquerda para a direita, como mostrado na Figura 78 (menos frequente e só para alguns determinantes).

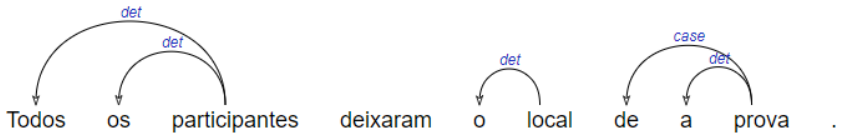


Figura 77 - Atribuição da deprel **det** à esquerda

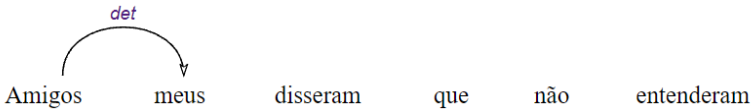


Figura 78 - Atribuição da deprel **det** à direita

discourse: discourse = discurso

A deprel **discourse** é usada para anotar interjeições e outros elementos que não têm uma relação clara com a estrutura sintática da sentença, exceto de maneira expressiva. A função dos dependentes de **discourse** é pragmática e não sintática. O elemento discursivo é anotado como dependente e o *head* é a palavra de mais alto nível na oração mais próxima.

Sentido da relação: a relação pode ocorrer nas duas direções, dependendo de onde estiver o dependente.

Temos dois casos frequentes de uso da deprel **discourse**: 1) interjeições e palavras de outras classes usadas como interjeições e 2) elementos de focalização de um dos elementos da sentença[12] .

Nas Figuras 79 e 80 estão ilustrados dois casos em que a deprel **discourse** é usada para anotar interjeição e adjetivo com função discursiva.

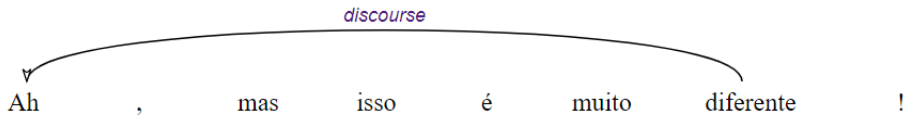


Figura 79 - Atribuição da deprel **discourse** a uma interjeição

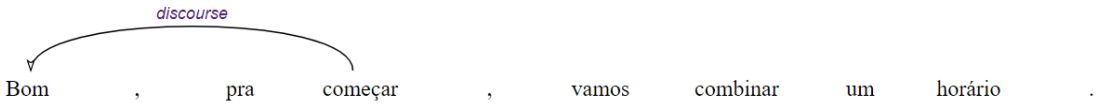


Figura 80 - Atribuição da deprel **discourse** a um adjetivo com função discursiva

Na função de **focalização**, a ordem do elemento em destaque é alterada e ele é sucedido de um "que", "é que" ou abraçado por um "é… que".

Na sentença ilustrada na Figura 81, a ordem canônica dos elementos seria "Nós soubemos do acidente só depois", porém os advérbios "só depois" foram colocados em destaque no início da sentença e, portanto, o "é" e o "que" que os sucedem têm função de focalização e são anotados com a deprel **discourse**.

[12] Esse uso da deprel **discourse** foi reportado no Relatório Técnico (Souza et al. 2020) que descreveu algumas decisões de anotação do córpus Bosque-UD. Nós a adotamos com uma leve alteração: não unimos o "é que" com a deprel **fixed**, como os autores fizeram, mas sim atribuímos a cada um dos tokens - "é" e "que", a relação **discourse**. Essa decisão se deve ao fato de que o "é que" pode ocorrer de forma não contígua e o "que" pode ocorrer com a mesma função, porém não acompanhado de "é".

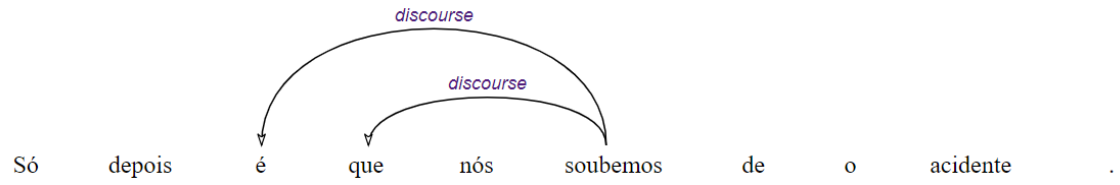


Figura 81 - Atribuição da deprel discourse a “é” e “que” com função discursiva

O mesmo ocorre na sentença da Figura 82, na qual o adjunto adverbial “desde ontem” é deslocado para o início da sentença. A ordem canônica seria “Estou esperando desde ontem” e, nesse caso, o “que” desapareceria, pois não há função discursiva de focalização.

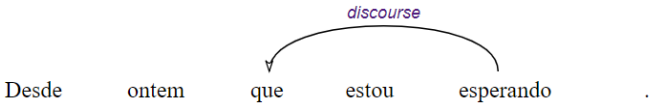


Figura 82 - Atribuição da deprel discourse a “que” focalizador

Na Figura 83, a ordem canônica dos elementos seria “Eu ganho isso por confiar nas pessoas”. Porém, o objeto direto “isso” foi colocado em destaque no início da sentença e, portanto, o “é” que o precede e o “que” que o sucede são analisados com a relação **discourse.**

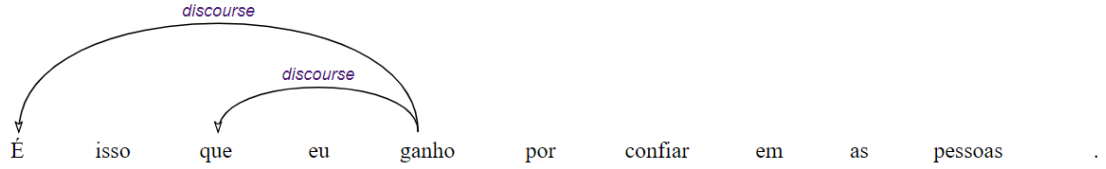


Figura 83 - Atribuição da deprel discourse a “é que” separados

dislocated: dislocated = deslocado

A deprel **dislocated** ocorre entre um elemento deslocado (talvez o melhor termo fosse reduntante) e o **root** da oração principal. A principal característica dessa deprel é o fato de o dependente repetir, com outras palavras, uma das funções sintáticas já preenchidas na árvore de dependências, caracterizando-se como uma redundância com função de tornar mais claro o token da função sintática já preenchida.

Sentido da relação: a relação pode ocorrer nas duas direções, dependendo de onde se encontra o elemento deslocado em relação ao **root**.

Normalmente o dependente de **dislocated** separa-se da oração principal por meio de vírgula. Ele repete uma das funções sintáticas já preenchidas na sentença: o sujeito, o objeto, um oblíquo e até o próprio verbo que constitui o **root**.

Na Figura 85, é o próprio verbo do **root** que aparece repetido, deslocado para o início da oração: "pensar, pensei".

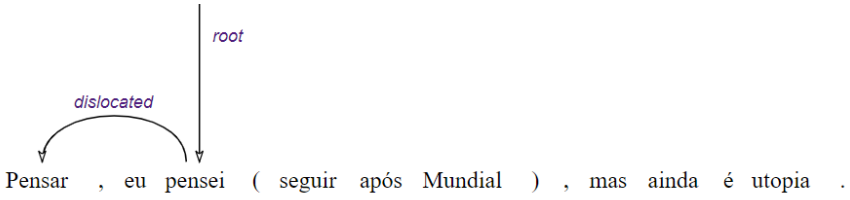


Figura 85 - Atribuição da deprel **dislocated** ligando dois verbos

Na sentença ilustrada na Figura 86, o sujeito, "elas" aparece repetido por meio de um correferente ao final da sentença: "elas" = "essas crianças".

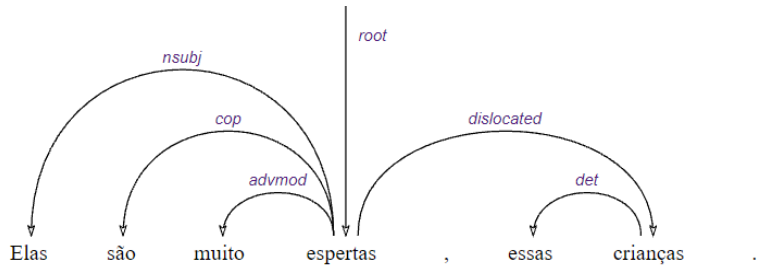


Figura 86 - Atribuição da deprel dislocated **ligando** um predicativo ADJ a um nominal

Na Figura 87, o sujeito "isso" aparece repetido por meio de um correferente anteposto ao próprio sujeito: dinheiro = isso.

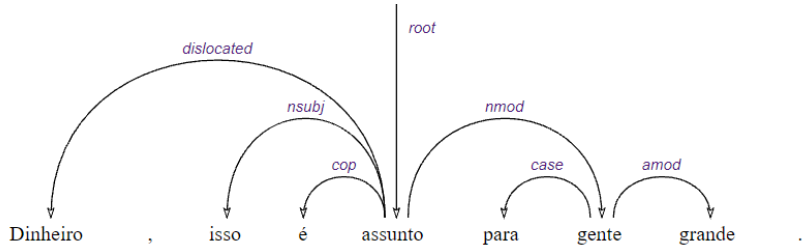


Figura 87 - Atribuição da deprel **dislocated** ligando um predicativo NOUN a um nominal

expl: expletive = expletivo

A deprel **expl** é usada para relacionar partículas expletivas, ou seja, sem um papel sintático, a um predicado.

Sentido da relação: a relação pode ocorrer nos dois sentidos.

No português, o uso dessa partícula tem três usos relacionados à anotação da partícula "se"[13]. O primeiro deles é para marcar o "se" de verbos pronominais, ou seja, casos em que o "se" não corresponde a um **obj** nem a um **iobj**, como ilustrado na Figura 88.

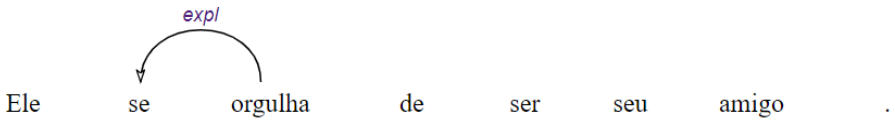


Figura 88 - Exemplo de atribuição de **expl** a verbo pronominal

O segundo uso da deprel **expl** é para marcar o "se" partícula de apassivação na voz passiva sintética, como mostrado na Figura 89.

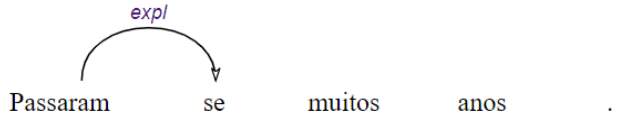


Figura 89 - Exemplo de atribuição de **expl** a "se" partícula apassivadora

E o terceiro uso da deprel **expl** é para marcar o "se" partícula de indeterminação do sujeito, como ilustra a Figura 90.

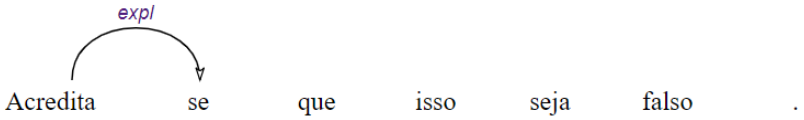


Figura 90 - Exemplo de atribuição de **expl** a "se" índice de indeterminação do sujeito

Outro uso da deprel **expl** é para relacionar a preposição "a" usada raramente para introduzir objetos de verbos transitivos diretos e que, portanto, tem caráter expletivo como mostrado na Figura 91.

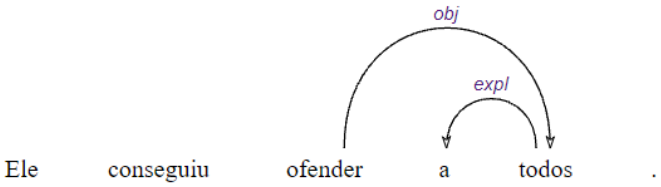


Figura 91 - Exemplo de verbo transitivo direto (ofender) com **obj** preposicionado

[13] Bouman et al. (2018) sugerem que as línguas românicas adotem uma sub-relação do expletivo: **expl:pass**, para partícula apassivadora, **expl:impers**, para índice de indeterminação do sujeito e **expl:pv** para pronomes de verbos pronominais. Essa sofisticação da anotação exige, contudo, *expertise* no assunto, motivo pelo qual recomendamos que seja implementada por linguistas.

flat: flat = relação plana

A relação **flat** foi criada para ligar elementos que têm relação entre si, mas não uma relação sintática, pois não há um *head* e um dependente naturalmente identificáveis, já que todos estão no mesmo plano. É o caso dos nomes próprios de pessoas, que além da relação **flat** recebem a sub-relação **name,** como ilustrado na Figura 92.

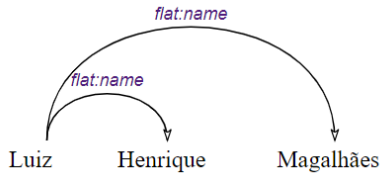


Figura 92 - Nome anotado com a relação **flat:name**

É o caso, também, dos numerais compostos, tanto cardinais quanto ordinais, como ilustrado na Figura 93.

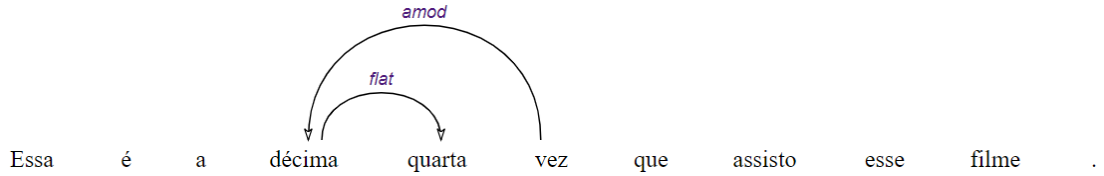


Figura 93 - Numeral ordinal composto anotado com a relação **flat**

Sentido da relação: a deprel **flat** não tem um *head* natural e por isso o *head* é arbitrariamente definido como sendo o primeiro token à esquerda. A relação parte do primeiro token em direção a cada um dos tokens do conjunto à direita. Portanto, é uma relação unidirecional da esquerda para a direita.

Se dentro de um numeral composto por extenso ou de um nome composto (Figura 94) houver sinais de pontuação, estes deverão ser anotados com a relação **punct**, pois a UD não admite outra relação para eles.[14]

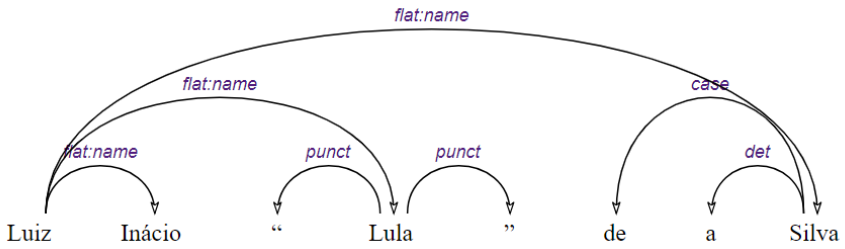


Figura 94 - Nome próprio com sinais de pontuação em seu interior

[14] O assunto é debatido no Issue 608 da UD: https://www.google.com/url?q=https://github.com/UniversalDependencies/docs/issues/608&sa=D&source=docs&ust=1636570191365000&usg=AOvVaw1Lyl_mzajlC9-kMw_NCVHf

Se dentro de um numeral composto ou de um nome composto houver conjunções coordenativas (“e”, por exemplo), elas deverão ser anotadas com a relação **cc** (Figura 95).

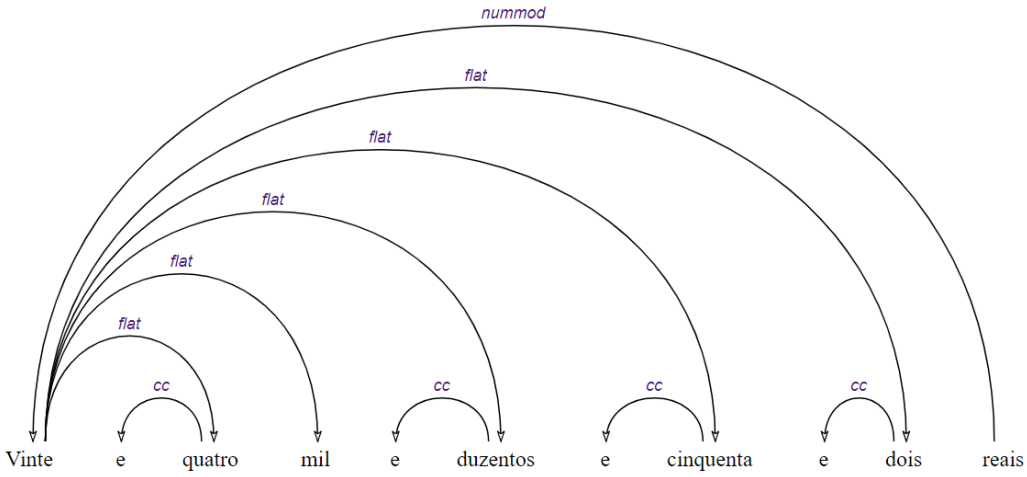


Figura 95 - Numeral composto com conjunções coordenativas em seu interior

A UD não tem uma regra clara, porém, para os casos em que “de”, “da”, “do”, “das”, “dos” ocorrem no interior de nomes próprios. Tais elementos costumam ser tokenizados (de, de a, de o, de as, de os) e anotados com suas *PoS tags* de origem (ADP e DET) e não como PROPN. Decidimos anotar essas relações interiores aos nomes próprios com suas relações exclusivas, **case** para ADP e **det** para DET, como mostrado nas Figuras 96 e 97, sempre dependentes do nome próprio à direita. Porém não anotamos esse nome próprio como **nmod** do primeiro, pois não julgamos que exista essa relação no nome próprio de pessoas. Em suma, usamos **flat** apenas para palavras de conteúdo.

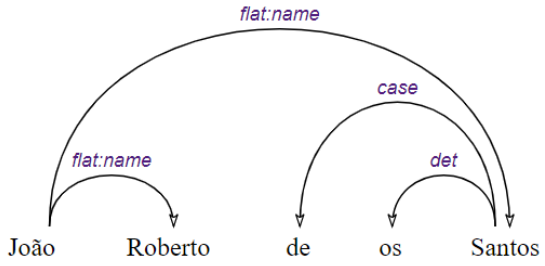


Figura 96 - Nome próprio com “de os” em seu interior

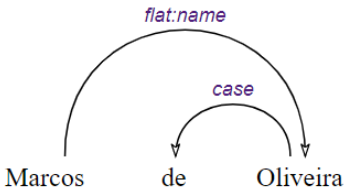


Figura 97 - Nome próprio com “de” em seu interior

Nomes próprios que apresentam uma relação sintática entre seus tokens deverão ser anotados com relações normais, como ilustra a Figura 98 e não com a relação **flat**.

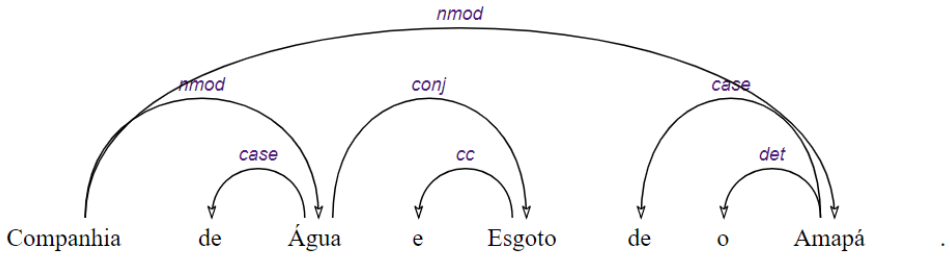


Figura 98 - Nome Próprio que apresenta relações sintáticas entre as palavras que o compõem

A relação **flat** também pode ser usada para relacionar duas palavras estrangeiras. Nesses casos, usa-se uma sub-relação: **flat:foreign**, como ilustrado na Figura 99.

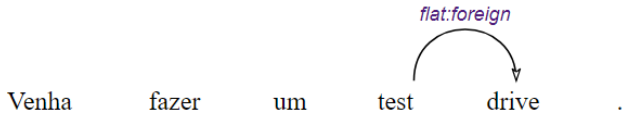


Figura 99 - Atribuição da relação **flat:foreign** a palavras estrangeiras

fixed: fixed = fixa

A deprel **fixed** foi criada para relacionar palavras funcionais (como preposições e conjunções) entre si ou a advérbios de classe fechada que não apresentam relação sintática, mas funcionam como um conjunto.

Sentido da relação: a relação parte do primeiro token em direção aos tokens seguintes que fazem parte da expressão **fixed**.

A decisão de usar a deprel **fixed** deve ser feita em nível de projeto e uma lista das expressões ligadas pela relação deve ser fornecida. Portanto, não cabe ao anotador decidir empregar a relação **fixed,** pois ela será atribuída semiautomaticamente uma vez que tenha sido definida pelo projeto.

Exemplos de expressões **fixed** são apresentados nas Figuras 100, 101, 102 e 103. A expressão fixa "desde que" funciona como conjunção subordinativa adverbial condicional ou temporal; a expressão fixa "que nem" funciona como conjunção subordinativa comparativa (sinônimo de "como"); a expressão fixa "a qual" funciona como um nominal (no exemplo, corresponde ao pronome "que"); a expressão fixa "a o", seguida de verbo no infinitivo, funciona como conjunção subordinativa temporal (corresponde a "quando").

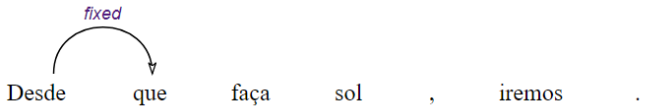


Figura 100 - Relação **fixed** entre advérbio e conjunção

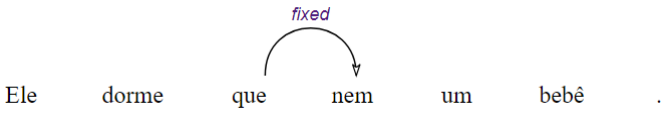


Figura 101 - Relação **fixed** entre duas conjunções

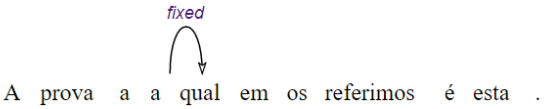


Figura 102 - Relação **fixed** entre dois pronomes[15]

[15] são considerados pronomes: o que, o qual, a qual, os quais, as quais (os dois tokens anotados como PRON)

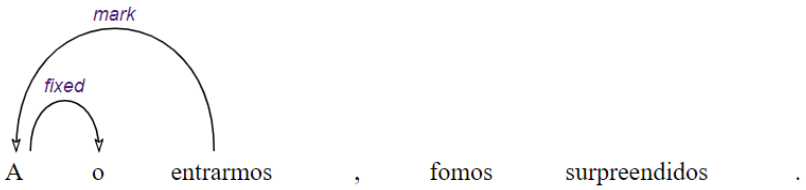


Figura 103 - Relação **fixed** entre preposição e artigo

goeswith: goes with = tokens que vão juntos

A deprel **goeswith** ocorre entre palavras ou partes de uma palavra que foram grafadas em desacordo com as normas ortográficas, ou tokenizadas indevidamente.

Sentido da relação: a relação parte do elemento da esquerda em direção ao elemento da direita.

A Figura 104 ilustra uma deprel **goeswith**: a palavra "meio-dia" é uma palavra composta, mas foi indevidamente tokenizada em duas. Unindo as duas partes com **goeswith**, preserva-se a integridade da palavra no nível sintático, impedindo que participem individualmente de outras deprel.

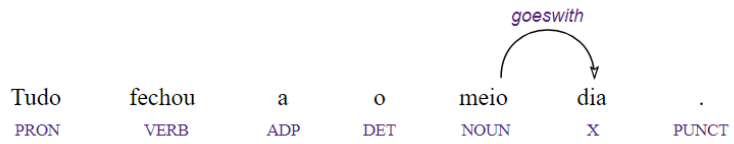


Figura 104 - Atribuição da relação goeswith a palavra composta tokenizada em duas partes

Atenção: ao usar goeswith, as *PoS tags* devem ser corrigidas simultaneamente. A regra é atribuir ao primeiro token (*head* do goeswith) a *PoS tag* da palavra correta e aos demais tokens (dependentes de goeswith) a *PoS tag* X, conforme pode ser observado na Figura 104.

iobj: indirect object = objeto indireto

A deprel **iobj** é usada para anotar a relação entre o predicado verbal e um terceiro argumento *core* (o primeiro é **nsubj** e o segundo é **obj**). Essa deprel é muito debatida na UD, pois cada língua tem uma interpretação sobre ela.

Embora o nome dessa deprel seja "objeto indireto", ela não corresponde ao conceito de objeto indireto das gramáticas tradicionais do português, ou seja, de objeto ligado ao verbo por meio de uma preposição. Na UD, os argumentos preposicionados devem ser anotados com a deprel **obl.**

Sentido da relação: a relação pode ocorrer nos dois sentidos, dependendo de o pronome vir antes ou depois do verbo.

A deprel **iobj** deve ser empregada, em português, apenas para anotar os objetos realizados sob a forma dos pronomes dativos "me, te, se, lhe, nos, vos, lhes", como mostrado na Figura 105.

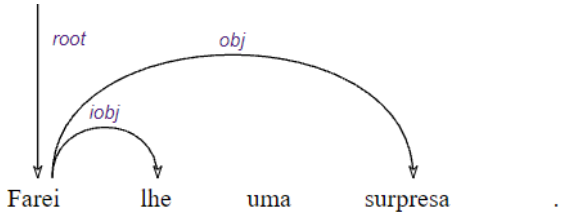


Figura 105 - Exemplo da atribuição da deprel **iobj**

Complementos verbais preposicionados devem ser anotados como **obl** e não como **iobj**. Assim, se a sentença da Figura 105 fosse reescrita sem o pronome, o complemento do verbo "fazer" teria que ser anotado com a deprel **obl** (Figura 106).

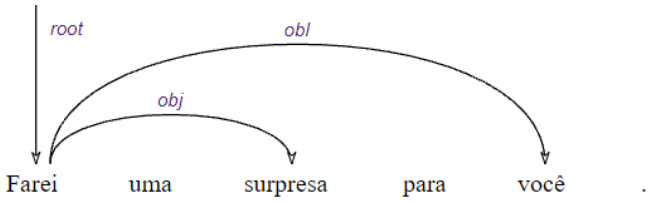


Figura 106 - Exemplo de atribuição da deprel **obl**

Atenção: os pronomes "me, te, se, nos e vos" são ambíguos, podendo atuar na relação **obj** (como ilustrado na Figura 107), ou como **iobj** (como ilustrado na Figura 108). Já os pronomes de terceira pessoa "o, a, os, as" são sempre **obj** e "lhe, lhes" são sempre **iobj**.

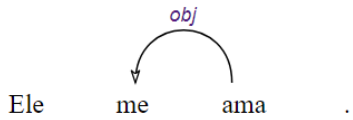


Figura 107 - Exemplo de atribuição da deprel **obj** ao pronome "me"

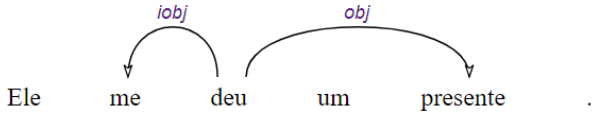


Figura 108 - Exemplo de atribuição da deprel **iobj** ao pronome “me”

list: list = lista

A deprel **list** ocorre entre os elementos que compõem uma lista. É uma relação que deve ser evitada se os itens da lista puderem ser relacionados pela deprel **conj**. Sempre que o último item da lista estiver introduzido por um "e", é sinal de que a deprel **conj** se aplica melhor que a deprel **list**.

Sentido da relação: a relação parte do primeiro elemento da lista em direção a cada um dos demais. Portanto, a deprel **list** é sempre da esquerda para a direita.

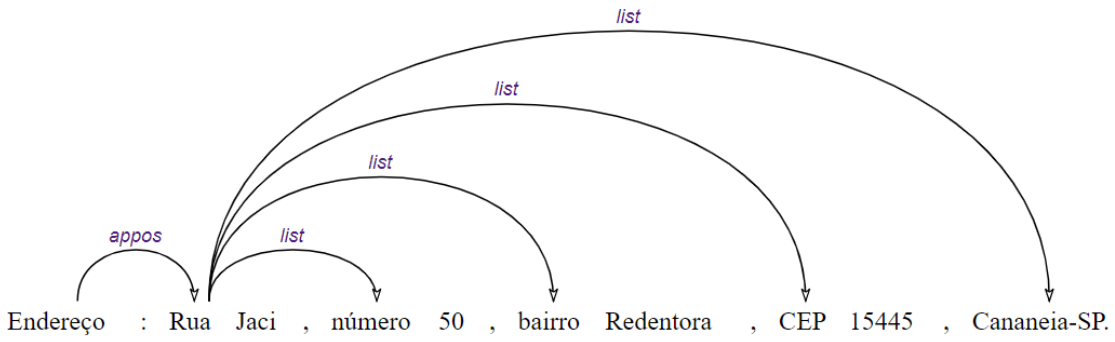


Figura 109 - Atribuição da deprel **list** a uma lista de itens que compõem um endereço

mark: marker = marcador de subordinação

A relação **mark** une a palavra que "marca" uma oração subordinada ao predicado da oração subordinada. Portanto, o *head* de **mark** é sempre um predicado, seja ele verbal ou nominal.

Sentido da relação: a relação parte do predicado da oração subordinada em direção ao marcador de subordinação. A relação é unidirecional, da direita para a esquerda.

Pelo lado do dependente da deprel **mark**, ou seja, o marcador de subordinação, há três situações clássicas:

1) As conjunções "que" e "se", que introduzem complementos oracionais dos verbos, como mostram as Figuras 110 e 111.

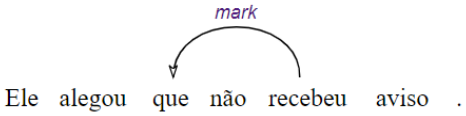


Figura 110 - Conjunção "que" introduzindo oração que complementa o sentido do verbo "alegar"

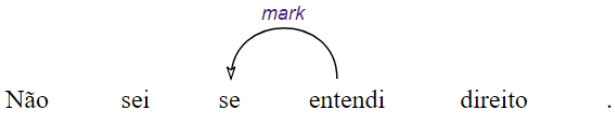


Figura 111 - Conjunção "se" introduzindo oração que complementa o sentido do verbo "saber"

2) A preposição que introduz uma oração reduzida que complementa o sentido de um verbo (Figura 112).

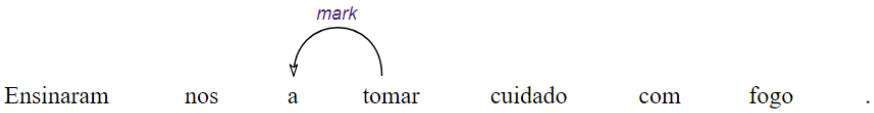


Figura 112 - Preposição "a" introduzindo oração reduzida que complementa o sentido de "ensinar"

3) E, por fim, as conjunções subordinativas que introduzem orações adverbiais, como mostrado nas Figuras 113, 114, 115 e 116.

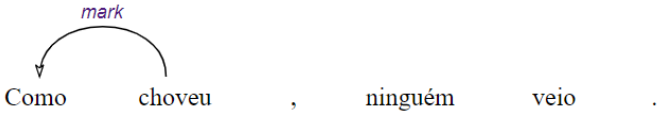


Figura 113 - Conjunção "como" introduzindo oração adverbial causal

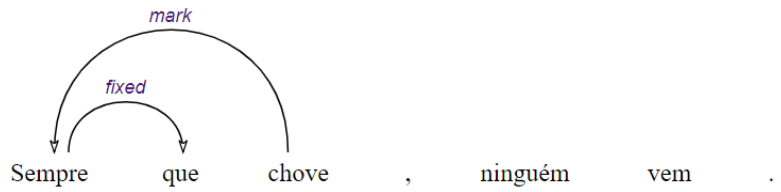


Figura 114 - Expressão fixa “sempre que” introduzindo oração adverbial temporal

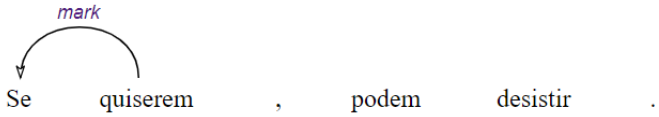


Figura 115 - “Se” introduzindo oração adverbial condicional

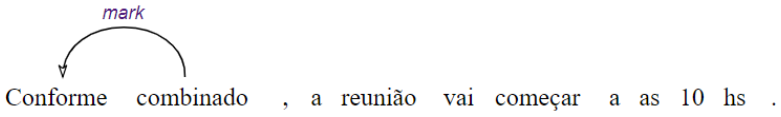


Figura 116 - “Conforme” introduzindo oração adverbial conformativa

nmod: nominal modifier = modificador nominal

A deprel **nmod** ocorre entre dois substantivos, quando um modifica o outro.

Sentido da relação: a relação parte do substantivo modificado em direção ao substantivo modificador. Em português, a relação é sempre da esquerda para a direita.

Geralmente, a deprel **nmod** é intermediada por preposição. Nesses casos, a preposição recebe a deprel **case**, que partirá do substantivo modificador, o *head* da relação, em direção à preposição (Figuras 117, 118 e 119).

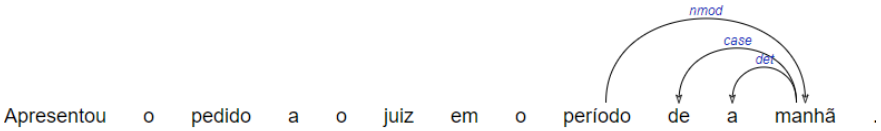


Figura 117 - Exemplo de **nmod** com a preposição *de* e artigo

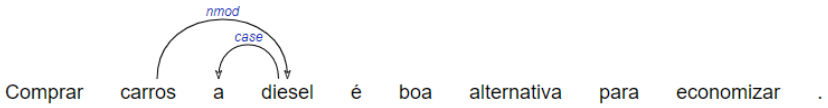


Figura 118 - Exemplo de **nmod** com a preposição *a*

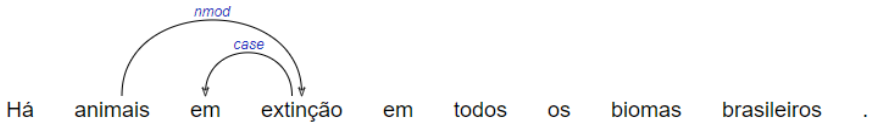


Figura 119 - Exemplo de **nmod** com a preposição *em*

Em alguns casos, a deprel **nmod** não será mediada por preposição, como na Figura 120 . Nesse caso, a relação também parte do substantivo modificado em direção ao modificador, porém não existe uma "marca" de qual é o modificador. Normalmente, no português, o dependente é o substantivo ou nome próprio à direita.

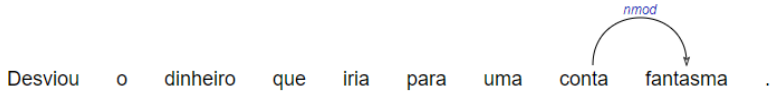


Figura 120 - Exemplo de **nmod** sem preposição

Além de ligar dois substantivos, a deprel **nmod** pode ligar um substantivo a um nome próprio, como na Figura 121.

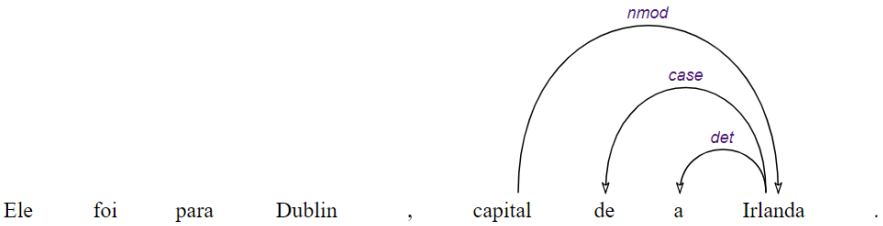


Figura 121 - Exemplo de nmod com dependente em forma de nome próprio

A Figura 122 ilustra um dependente de **nmod** que, por sua vez, também é modificado por **nmod**.

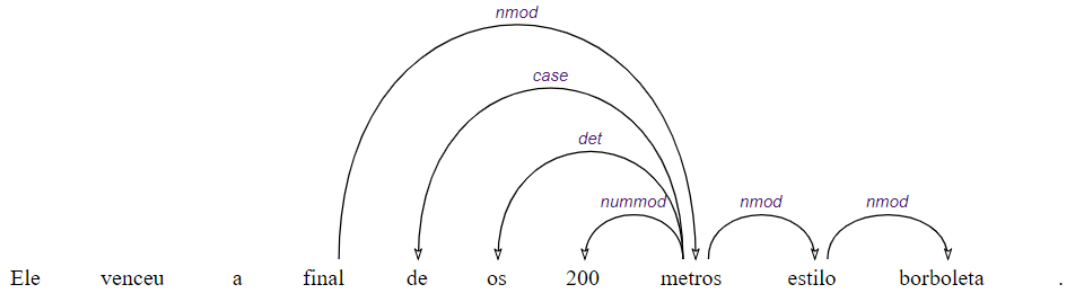


Figura 122 - Exemplo de dois **nmod**, um "dentro" do outro

Quando dois nomes próprios se ligam sem uma preposição, na função de especificação, o elemento genérico é o *head*, e o especificador é o dependente do **nmod** (Figura 123).

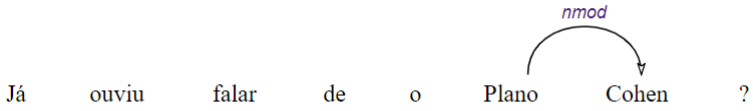


Figura 123 - Exemplo de **nmod** ligando nomes próprios

nmod X obl

Todo nominal preposicionado que modifica outro nominal é dependente da deprel **nmod**. Mas se o nominal preposicionado estiver modificando um verbo, um adjetivo ou um advérbio, será dependente da deprel **obl**, como mostrado nas sentenças a seguir.

- O **evento** em **São** Paulo foi o que recebeu mais visitantes. (em **São** Paulo = **nmod**)
- Todos **vieram** para **São** Paulo. (para **São** Paulo = **obl**)
- **Treinos** em a **piscina** são mais divertidos. (em a **piscina** = **nmod**)
- Ele **treina** em a **piscina** duas vezes na semana. (em a **piscina** = **obl**)
- **Greve** por um **aumento** no salário tem sido cogitada. (por um **aumento** = **nmod**)
- Os funcionários estão **ansiosos** por um **aumento**. (por um **aumento** = **obl**)
- **Diferenças** de **gênero** não são consideradas. (de **gênero** = **nmod**)
- Serão aceitas inscrições, **independentemente** de **gênero**. (de **gênero** = **obl**)

Como pode ser observado, os *head* que são substantivos (evento, treinos, greve e diferenças) impõem a decisão por **nmod**. Já os *head* que são verbos (vieram, treina), adjetivo (ansiosos) e advérbio (independentemente), impõem a decisão por **obl**.

nsubj: subject = sujeito

A deprel **nsubj** é usada para anotar a relação entre o predicado verbal ou nominal e o primeiro argumento *core* do verbo, que é o sujeito.

Sentido da relação: a relação parte do predicado da oração (verbal ou nominal) em direção ao sujeito. A relação pode ocorrer nos dois sentidos.

A Figura 124 ilustra um sujeito simples na ordem canônica, ou seja, antes do verbo.

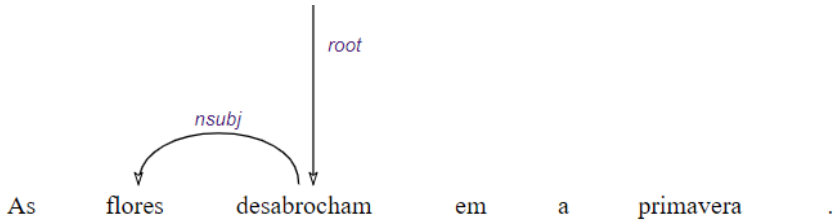


Figura 124 - Exemplo de atribuição da deprel **nsubj** a predicado verbal

Se a oração é constituída de verbo de cópula, o sujeito se liga ao predicado nominal, como nas Figuras 125 e 126.

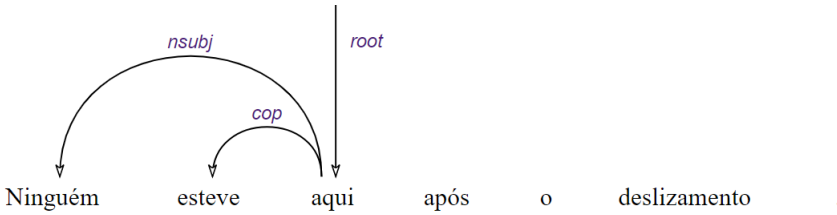


Figura 125 - Exemplo de atribuição da deprel **nsubj** a predicado nominal sob forma de advérbio

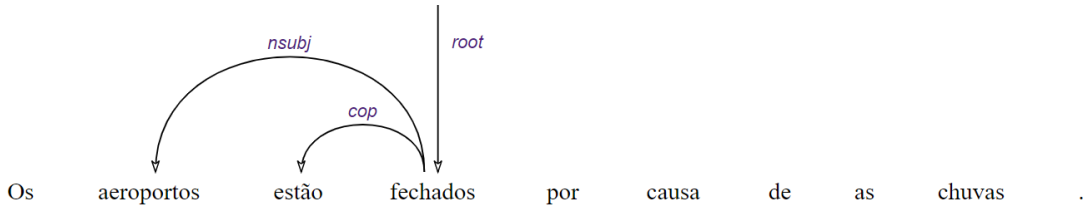


Figura 126 - Exemplo de atribuição da deprel **nsubj** a predicado nominal[16] sob forma de particípio

O sujeito da voz passiva é anotado com o subtipo **nsubj:pass.** Para isso, além de definir a relação **nsubj**, o anotador deverá definir também a sub-relação **pass** (Figura 127).

[16] As Figuras 9 e 10 apresentam o mesmo particípio, “fechados”, mas após o verbo “estar”, a palavra “fechados” é anotada como ADJ, pois “estar” é verbo de cópula (**cop**), ao passo que após o verbo “ser”, a palavra “fechados” é anotada como VERB, pois se trata de construção de voz passiva, na qual o verbo “ser” é auxiliar de passiva (**aux:pass**).

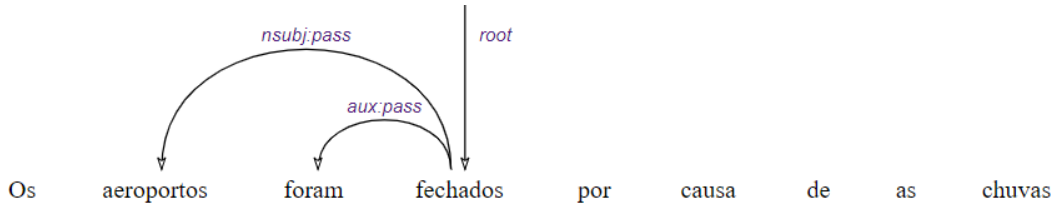


Figura 127 - Exemplo de atribuição da deprel **nsubj:pass** a predicado verbal na voz passiva analítica

O mesmo ocorre nas construções de voz passiva sintética[17], como ilustra a Figura 128.

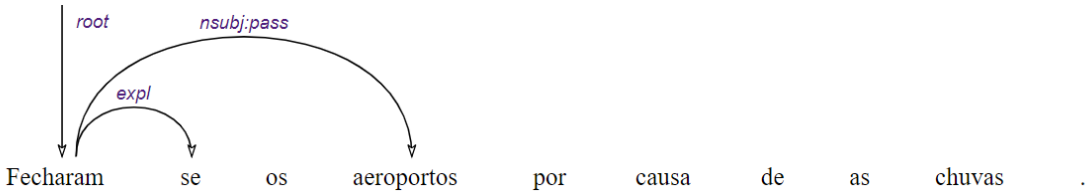


Figura 128 - Exemplo de atribuição da deprel **nsubj:pass** a predicado verbal na voz passiva sintética

Verbos existenciais que possuem sujeito[18], normalmente apresentam o sujeito posposto (colocado após o verbo), como mostra a Figura 129.

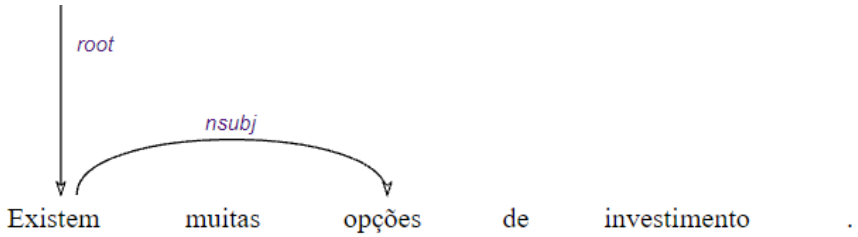


Figura 129 - Exemplo de atribuição da deprel **nsubj** a verbo existencial

Caso o verbo "ser" esteja em sentido existencial (nem como auxiliar de voz passiva nem como verbo de cópula), como na sentença ilustrada na Figura 130, ele poderá ser **root** e deverá ter a *PoS tag* VERB.

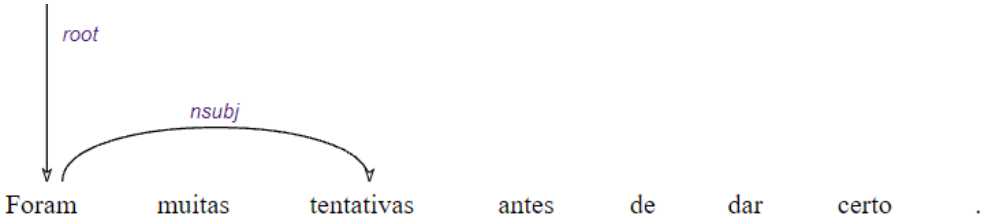


Figura 130 - Exemplo de atribuição da deprel **nsubj** ao verbo SER no sentido existencial

[17] O pronome "se", usado para construir a voz passiva sintética, é anotado com a deprel **expl**, porque não tem papel semântico.

[18] Exemplos de verbos existenciais que não têm sujeito são os verbos "ter" e "haver" no sentido de "existir", pois não concordam com o que vem após. Ex: *Havia pessoas morrendo*. *Tem horas que nada resolve.* Nesses casos, os verbos não possuem sujeito, mas possuem objeto.

Outro caso que merece destaque é quando o sujeito de uma oração subordinada é representado por um pronome relativo, como é o caso de “que”, mostrado na Figura 131.

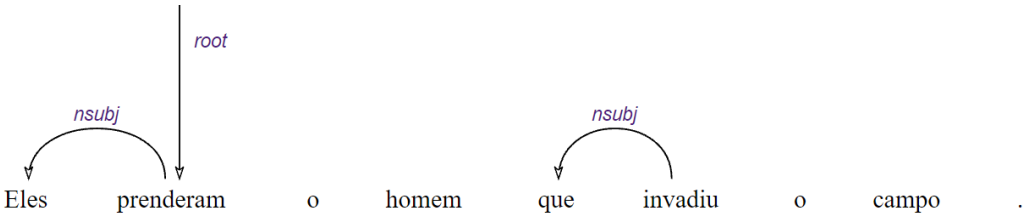


Figura 131 - Exemplo de **nsubj** atribuído a pronome relativo “que”

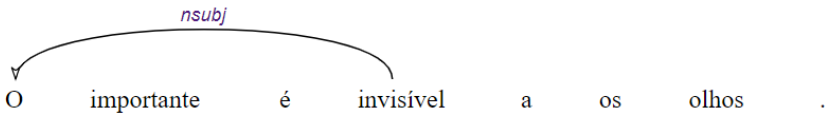


Figura 132 - Exemplo de nsubj atribuído ao pronome “o”

Em orações com verbo de cópula iniciadas por “O” e seguidas de adjetivo, o “O” é anotado como pronome no nível de *PoS tag* e como **nsubj** no nível das dependências, como ilustrado acima, na Figura 132. Outros exemplos da mesma situação são mostrados nas sentenças a seguir, onde o **nsubj** está em azul e o **root** está em negrito.

- **O** estranho é a **quantidade** de pessoas que votaram nele.
- **Os** menos favorecidos **sofrem** mais o efeito da inflação.
- **As** mais elegantes **receberam** elogios.

Se, contudo, a construção tiver um verbo de cópula e uma oração como predicativo, o verbo de cópula deverá ser anotado como **root** a fim de evitar que o verbo da oração predicativa tenha dois sujeitos. Por exemplo, na sentença ilustrada na Figura 133, se não considerássemos “é” como **root**, o verbo “perceber” teria como sujeito tanto “O” quanto “ninguém”.

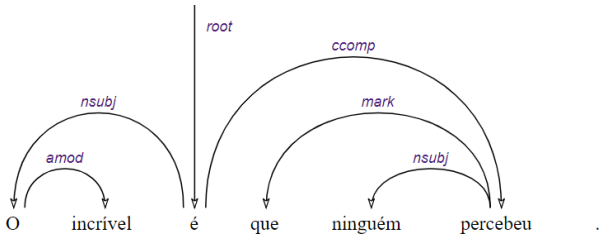


Figura 133 - Exemplo de **nsubj** atribuído ao pronome “o”

nummod: numeric modifier = modificador numérico

A deprel **nummod** ocorre entre um modificador numérico (numerais cardinais por extenso ou em algarismos) e o elemento que ele modifica, como mostrado nas Figuras 134, 135 e 136. Em geral, apenas a etiqueta de *PoS tag* NUM terá a deprel **nummod** como correspondente[19].

Sentido da relação: a relação parte do elemento modificado ou quantificado em direção ao numeral. A deprel **nummod** aceita relações da direita para a esquerda (mais frequente) e da esquerda para a direita.

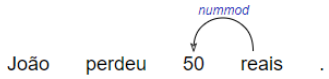


Figura 134 - exemplo de **nummod** quantificador expresso em algarismos

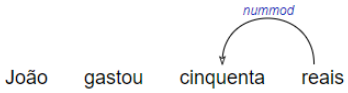


Figura 135 - exemplo de **nummod** quantificador expresso por extenso

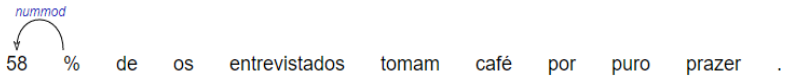


Figura 136 - exemplo de **nummod** modificando um símbolo

Quando o numeral é expresso por extenso, não por algarismos, as palavras que o compõem serão ligadas pela deprel **flat**. Eventuais conjunções dentro da expressão do número serão anotadas com a deprel **cc**. Essa situação é ilustrada na Figura 137.

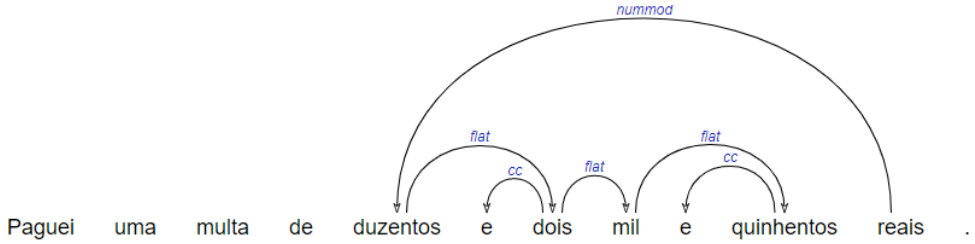


Figura 137 - exemplo de número composto modificando um substantivo

A deprel **nummod** também será usada nos casos em que o numeral classifica um substantivo, como na Figura 138, em que "4" está qualificando o substantivo "raia".

[19] Nem toda *PoS tag* NUM, porém, faz parte de uma relação **nummod**, pois NUM pode assumir também funções nominais, como as de sujeito e objeto.

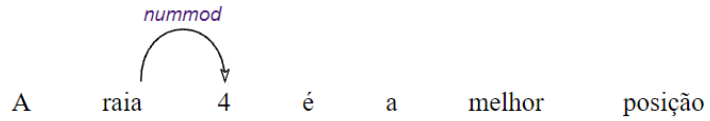


Figura 138 - exemplo de **nummod** classificando um substantivo

Outros casos de relação **nummod** em que o numeral está classificando um substantivo:

- lei **4320**
- aro **29**
- sala **8**
- vaga **57**

A mesma relação **nummod** ocorre quando um numeral está classificando um nome próprio, como em:

- BR **101**
- Rio **92**
- Windows **10**
- Copa do Mundo **2022**

A deprel **nummod** também será usada para numerais pospostos a pronomes pessoais, especificando a quantidade de pessoas representadas no pronome, como em:

- nós **três**
- eles **dois**
- vocês **quatro**

Essa situação é ilustrada pela Figura 139:

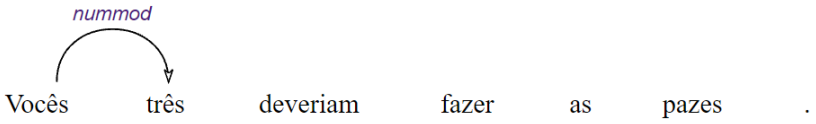


Figura 139 - exemplo de **nummod** quantificando um pronome pessoal

obj: object = objeto direto

A deprel **obj** é usada para anotar a relação entre o predicado verbal e o segundo argumento *core* do verbo (o primeiro é **nsubj**).

Sentido da relação: a relação parte do verbo em direção ao objeto que o complementa, não importando se está à direita ou à esquerda.

Um requisito de **obj** é que possa ser promovido a **nsubj** na alternância de voz passiva, como ocorre com a palavra "convite" da Figura 140 para a Figura 141.

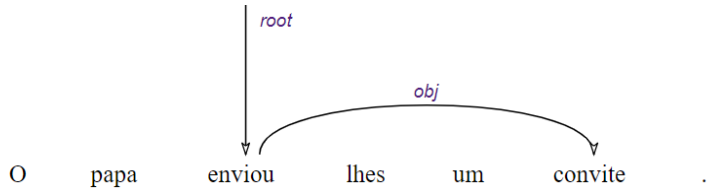


Figura 140 - Exemplo de atribuição da deprel **obj**

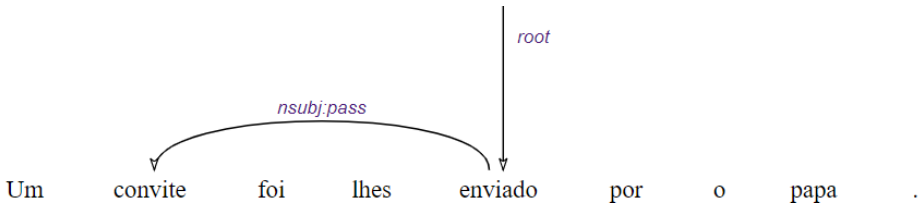


Figura 141 - Exemplo de alternância do **obj** a **nsubj** na voz passiva

Os **obj** são, via de regra, não preposicionados. Mas há casos em que a preposição "a" (e só ela) é utilizada, sem que o verbo a exija (uso expletivo). Trata-se de situações muito específicas: antes do nome de Deus (Figura 142); antes dos pronomes "quem", "ninguém", "todos" (Figura 143).

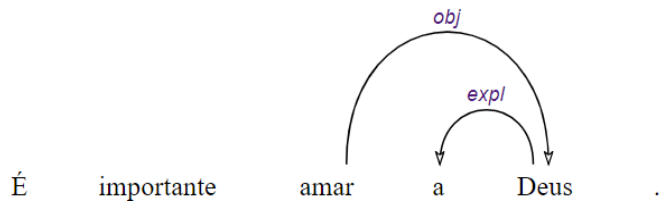


Figura 142 - Exemplo de verbo transitivo direto (amar) com **obj** preposicionado

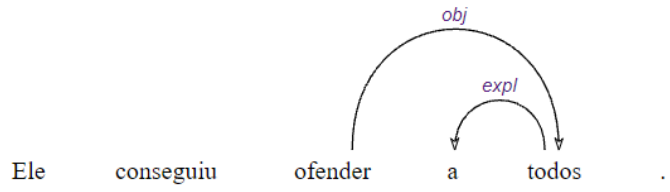


Figura 143 - Exemplo de verbo transitivo direto (ofender) com **obj** preposicionado

Os verbos “amar” na Figura 142 e “ofender”, na Figura 143, são transitivos diretos e não requerem a preposição “a” para introduzir seu complemento objeto direto. Conclui-se, portanto, que o “a”, nesses exemplos, tem função expletiva.

Nesses casos, o argumento deverá ser anotado com a deprel **obj** e a preposição “a” com a relação **expl** (ao invés de **case**, que é a deprel mais frequente das preposições).

Atenção:

Quando houver uma locução verbal (mais de um verbo em sequência), pode ocorrer de o objeto direto do último verbo estar à esquerda do primeiro verbo. Nesses casos, o **obj** deverá ser excepcionalmente ligado ao primeiro verbo da locução (lado esquerdo da Figura 144), apenas para que não haja cruzamento de arcos (lado direito da Figura 144), já que isso deve ser evitado na anotação pois atrapalha o processamento automático.

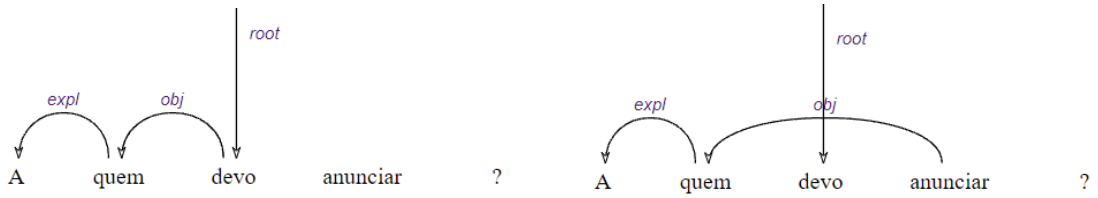


Figura 144 - Exemplo de verbo transitivo direto (anunciar) com **obj** preposicionado anteposto

obl: oblique nominal = nominal oblíquo

A deprel **obl** é usada para ligar um verbo, um adjetivo ou advérbio (*head*) a um nominal (dependente), o qual, na maioria das vezes, é preposicionado.

Sentido da relação: pode ocorrer nos dois sentidos

As Figuras 145 e 146 ilustram respectivamente um adjetivo e um advérbio na relação **obl**.

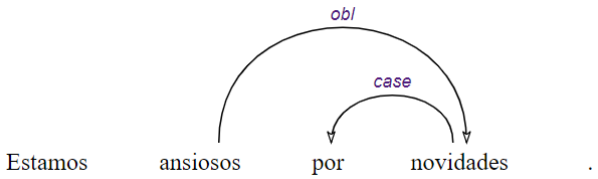


Figura 145 - Exemplo de **obl** em que o *head* é um adjetivo

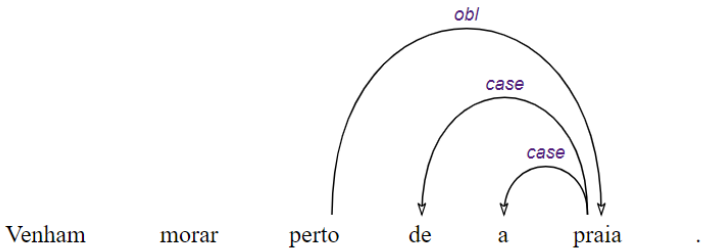


Figura 146 - Exemplo de **obl** em que o *head* é um advérbio

O nominal preposicionado pode ser um adjunto adverbial (Figura 147) ou um argumento do verbo (o que chamamos de objeto indireto na gramática tradicional), como ilustrado na Figura 148.

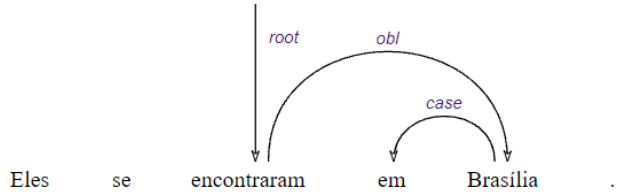


Figura 147 - Exemplo de atribuição de **obl** a um adjunto adverbial de lugar

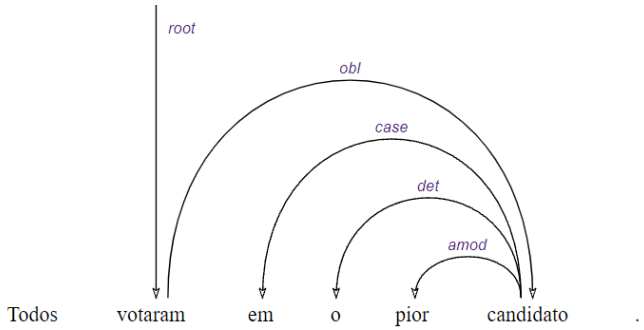


Figura 148 - Exemplo de atribuição de **obl** a argumento do verbo “votar”

Quando o **obl** corresponde ao agente da voz passiva, é possível acrescentar uma sub-relação "agent" ao **obl**, como mostrado na Figura 149.

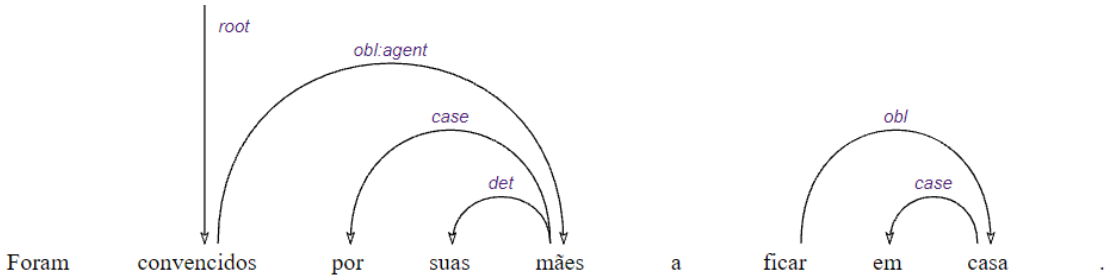


Figura 149 - Exemplo de atribuição de **obl**:agent a agente de voz passiva

obl X iobj

A deprel **iobj**, no português, é utilizada apenas para anotar pronomes que sejam objeto de um verbo transitivo indireto ou de um verbo transitivo direto e indireto, mas nunca para um argumento preposicionado. Argumentos preposicionados de verbos são dependentes da relação **obl** (Figura 150).

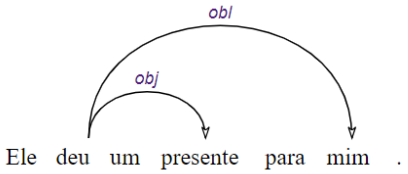


Figura 150 - Exemplo de atribuição de **obl** a pronome preposicionado

obl X nmod

Um nominal preposicionado será dependente da deprel **obl**, se estiver relacionado a um verbo, a um adjetivo ou a um advérbio, e será dependente da deprel **nmod**, se estiver relacionado a um nominal. Como pode ser observado nas sentenças a seguir, os *head* que são substantivos (evento, treinos, greve e diferenças) impõem a decisão por **nmod**. Já os *head* que são verbos (vieram, treina), adjetivo (ansiosos) e advérbio (independentemente), impõem a decisão por **obl**.

- O **evento** em **São** Paulo foi o que recebeu mais visitantes. (em **São** Paulo = **nmod**)
- Todos **vieram** para **São** Paulo. (para **São** Paulo = **obl**)
- **Treinos** em a **piscina** são mais divertidos. (em a **piscina** = **nmod**)
- Ele **treina** em a **piscina** duas vezes na semana. (em a **piscina** = **obl**)
- **Greve** por um **aumento** no salário tem sido cogitada. (por um **aumento** = **nmod**)
- Os funcionários estão **ansiosos** por um **aumento**. (por um **aumento** = **obl**)
- **Diferenças** de **gênero** não são consideradas. (de **gênero** = **nmod**)
- Serão aceitas inscrições, **independentemente** de **gênero**. (de **gênero** = **obl**)

obl X advmod

Tanto **obl** quanto **advmod** ligam adjuntos adverbiais a verbos, adjetivos e advérbios. No entanto, a UD diferencia os adjuntos adverbiais constituídos por advérbios (**advmod**), como na Figura 151, dos adjuntos adverbiais constituídos por nominais (**obl**), como na Figura 152.

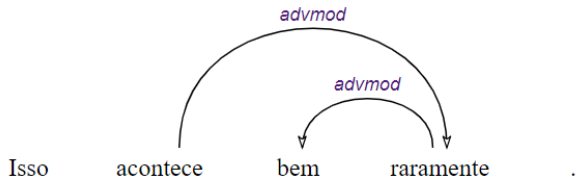


Figura 151 - Dois **advmod**: um modificando verbo e outro modificando advérbio

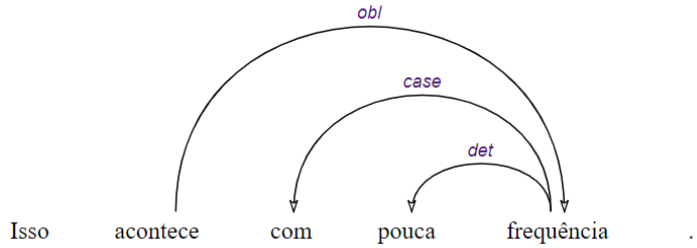


Figura 152 - Exemplo de **obl** atribuído a adjunto adverbial

Atenção:

Quase sempre **obl** é introduzido por preposição e **advmod** não é. Mas é possível ocorrer **obl** não introduzido por preposição (como na Figura 153) e **advmod** introduzido por preposição (como na Figura 154). Um **obl** não introduzido por preposição ocorre quando há elipse da preposição. Na Figura 153, por exemplo, "chegam terça-feira" corresponde a "chegam **na** terça feira). A principal diferença entre os dois é que na deprel **obl** o dependente é sempre um nominal e na deprel **advmod** o dependente é sempre um advérbio.

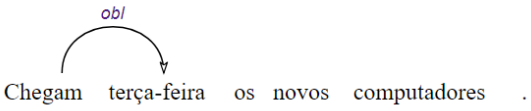


Figura 153 - Exemplo de **obl** não introduzido por preposição

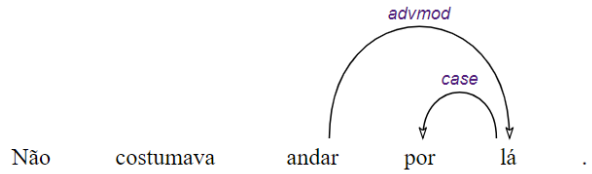


Figura 154 - Exemplo de **advmod** preposicionado

obl X advcl

Algumas vezes o verbo da oração adverbial é um verbo de cópula elíptico e, portanto, o predicativo é que será o dependente. Por isso, uma pista importante para identificar essas orações é a presença de uma conjunção subordinativa (ou locução conjuntiva) introduzindo-as, como "embora", "apesar de", "mesmo que", "embora", "ainda que", etc. (Figura 155).

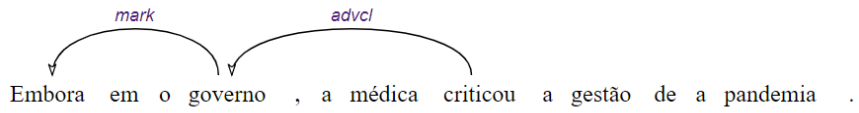


Figura 155 -Relação **advcl** em que o *head* é o predicativo com verbo de cópula elíptico

orphan: orphaned dependent = órfão

A deprel **orphan** liga dois elementos entre os quais há um ou mais elementos elípticos.

Sentido da relação: a relação parte do elemento anterior à elipse em direção ao elemento posterior à elipse. A deprel **orphan**, portanto, é sempre da esquerda para a direita.

Na Figura 156, por exemplo, há a elipse do verbo "ir" na segunda oração. A oração sem elipse seria: "João foi para Portugal e Maria **foi** para Taubaté." Se o verbo estivesse presente, haveria uma relação **nsubj** ligando "foi" a "Maria" e uma relação **obl** ligando "foi" a "Taubaté". Com a elipse do verbo, contudo, tanto "Maria" quanto "Taubaté" ficaram órfãos de *head* ("Maria" dependente de **nsubj** e "Taubaté" dependente de **obl**) Por esse motivo, os dois dependentes órfãos se unem em uma relação **orphan**, o primeiro, "Maria", assumindo a posição de *head,* e o segundo, "Taubaté", assumindo a posição de dependente.

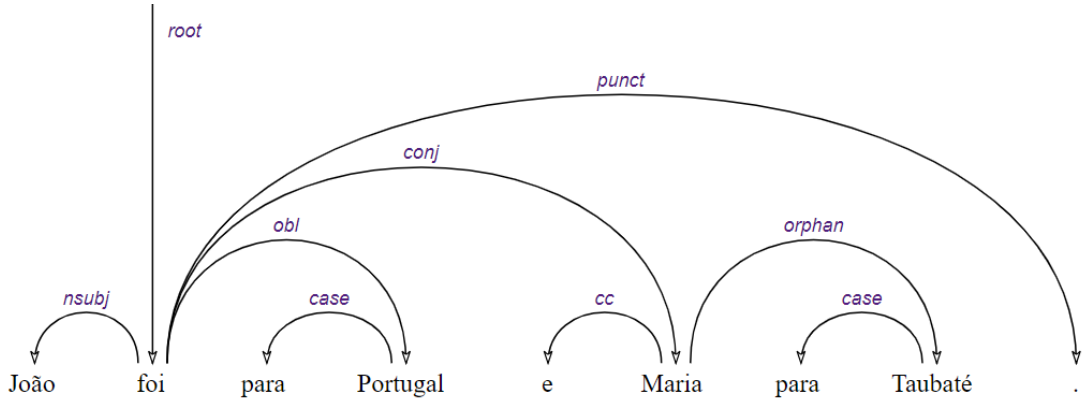


Figura 156 - Atribuição da deprel **orphan** unindo dependentes de elemento elíptico

parataxis: parataxis = parataxis

A deprel **parataxis** é uma relação entre dois elementos da sentença que poderiam ter relação sintática entre si, porém essa relação não está explicitada. É diferente de **discourse**, cujo dependente não tem função sintática, apenas expressiva (pragmática). A deprel **parataxis** é usada principalmente para ligar orações, como no exemplo ilustrado na Figura 157. Nessa figura, duas orações estão justapostas e o leitor pode inferir uma relação de causalidade entre elas, apesar de não haver uma conjunção subordinativa causal (se tivesse, a sentença seria: "Isso nunca funcionaria **porque** ninguém confia no gorverno").

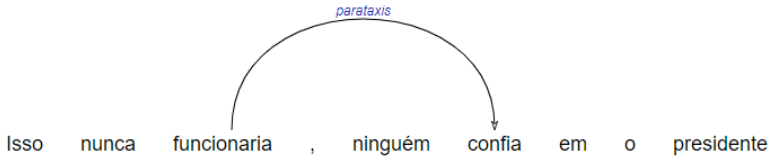


Figura 157 - Atribuição da deprel **parataxis** ligando duas orações sem conexão explícita

Sentido da relação: a deprel **parataxis** pode ocorrer nos dois sentidos, dependendo de onde se encontra o **root**.

A escolha do root: se houver várias orações, o **root** deverá ser o predicado da primeira. No caso de haver uma oração com **verbo de elocução** (afirmar, dizer, falar, etc.), introduzindo discurso direto, o **root** será o verbo de elocução sempre que ele ocorrer primeiro. Se, contudo, a oração contendo o verbo de elocução ocorrer no meio ou no final da sentença, o **root** será o verbo da oração principal do discurso direto[20]. Isso é mostrado nas três sentenças a seguir, nas quais o **root** está em negrito e o dependente em azul.

- Ele **disse**: vocês **partirão** de manhã.
- Vocês, ele **disse**, **partirão** amanhã de manhã.
- Vocês **partirão** amanhã de manhã, ele **disse**.

As Figuras 158 e 159 mostram dois exemplos da mesma sentença, uma com discurso direto e outra com discurso indireto, respectivamente. Quando se tratar de **discurso indireto**, a fala reportada está inserida em uma oração subordinada, que é um complemento do verbo de elocução, não uma **parataxis**.

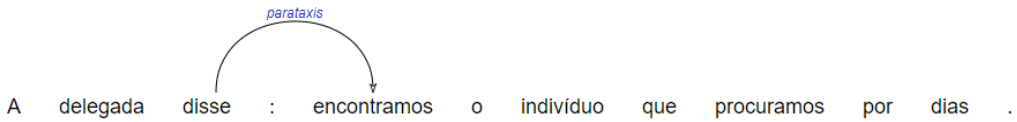


Figura 158 - Verbo de elocução ("disse") anotado como *head* de **parataxis**

[20] As guidelines da UD justificam essa opção de anotação, argumentando que : "An argument for this analysis is that in the cases analyzed as embedding, the entire clause can be further embedded, while this is not possible with medial or final placement of the speech verb."

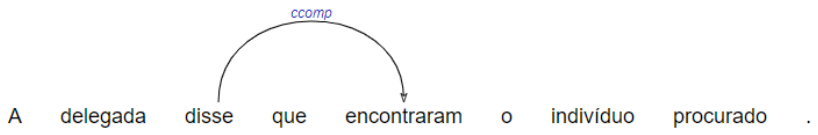


Figura 159 - Verbo de elocução (“disse”) e seu complemento oracional

Também usamos a deprel **parataxis** para anotar **afiliação expressa entre parênteses**, pois se trata de um modificador nominal não expresso como um **nmod** comum. A afiliação se liga ao nominal que modifica, como ilustrado na Figura 160.

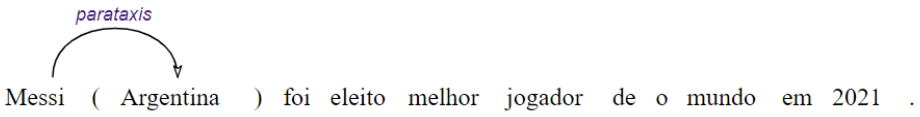


Figura 160 - Afiliação entre parênteses, modificando nominal, anotada como **parataxis**

Outro uso da deprel **parataxis** é para anotar “**conversa com o interlocutor**”, seja requisitando sua atenção, seja confirmando a recepção e entendimento da mensagem, como no exemplo ilustrado na Figura 161.

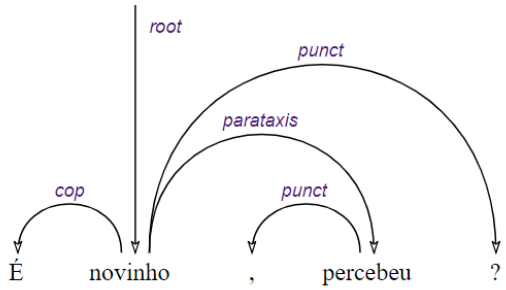


Figura 161 - Atribuição da deprel **parataxis** na confirmação da atenção do interlocutor

Outros exemplos de **confirmação** da atenção do interlocutor anotados como dependentes da deprel **parataxis** estão destacados em azul nas sentenças a seguir (o *head* está em negrito).

- Você **comprou** esse presente para ela, não **foi**? (o “não” é **advmod** de “foi”)
- Quem ia **saber**, **né**[21]?
- É **novinho**, **percebeu**?
- Vou **chegar** logo, **tá**?
- Isso não vai **dar**, tá **ligado**? (o “tá” é **cop** de “ligado”)

Exemplos de **requisição** de atenção do interlocutor foram anotados em negrito nas sentenças a seguir, pois são *head* da deprel **parataxis** por ocorrerem no início da sentença.

- **Olha**, assim não vai **dar**.
- **Preste** atenção: ninguém **deve** sair desacompanhado.

[21] A tokenização de “né” em “não é” é uma decisão de projeto. Como “né” é tipicamente usado na função discursiva, não parece haver necessidade de tokenizá-lo. Porém, se não for tokenizado, é preciso decidir a PoS tag da contração (PART poderia ser uma boa opção).

punct: punctuation = pontuação

A deprel **punct** ocorre entre um símbolo de pontuação (*PoS tag* **PUNCT**) e uma palavra de conteúdo de sua vizinhança, de acordo com algumas regras.

Sentido da relação: a relação parte da palavra de conteúdo que será *head* em direção ao sinal de pontuação. A relação admite as duas direções.

Todo sinal de pontuação é dependente de uma deprel **punct**. Existem algumas regras para decidir qual é o *head* da relação **punct**, porém, acima de tudo, está a regra de que não se pode produzir cruzamento dos arcos que representam as relações. Por isso, o ideal é relacionar os sinais de pontuação a seus respectivos *heads* ao final da anotação, quando todas as demais relações já tiverem sido anotadas.

Regras:

Ponto final, ponto de exclamação e ponto de interrogação (. ? !) que terminam uma sentença são sempre dependentes da palavra marcada como **root** da sentença (Figura 162).

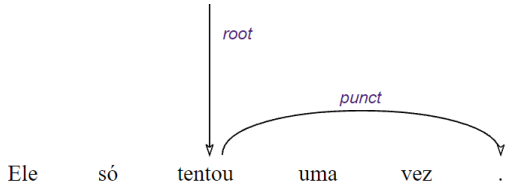


Figura 162 - a deprel **punct** ligando o **root** ao ponto final.

Sinais de pontuação em pares são sempre dependentes do *head* de tudo aquilo que estiver dentro dos pares. Ex: **( ) [ ] { } “ ” ‘ ’ < >**. As Figuras 163 e 164 mostram dois exemplos desses pares de pontos.

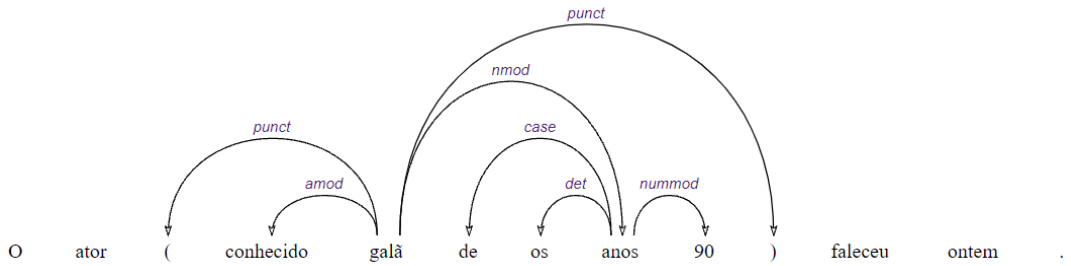


Figura 163 - Atribuição da deprel **punct** a um par de parênteses

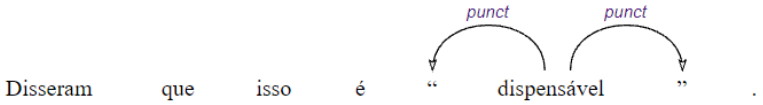


Figura 164 - Atribuição da deprel **punct** a um par de aspas

A regra vale também para pares de vírgula que separam orações intercaladas, como mostrado na Figura 165.

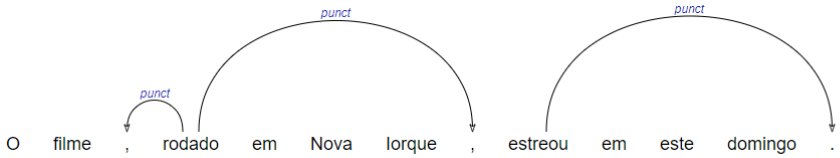


Figura 165 - Atribuição da deprel **punct** a um par de vírgulas que separam uma oração intercalada

Todos os sinais de pontuação que precedem ou sucedem um dependente devem ser ligados a esse dependente. Por exemplo, o sinal de dois pontos (:) na Figura 166, precede o dependente de uma relação **appos** e por isso é ligado ao *head* desse dependente.

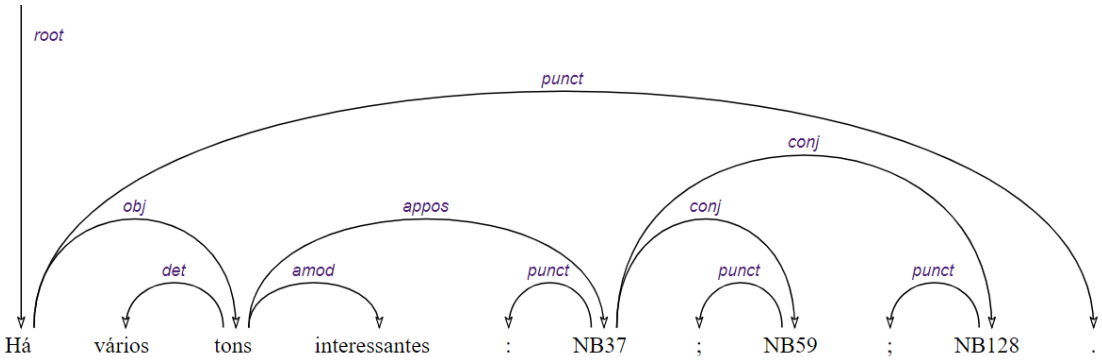


Figura 166 - deprel **punct** de dois pontos e de ponto-e-vírgula

A Figura 166 também ilustra ponto-e-vírgula que separa elementos coordenados (“NB37; NB59; NB128”). Para não produzir cruzamento de arcos, esse tipo de pontuação é sempre ligado ao próximo elemento coordenado de uma série de coordenações. O mesmo ocorre quando os elementos coordenados (“livros, cadernos, lápis e canetas”) são separados por vírgula, como mostrado na Figura 167.

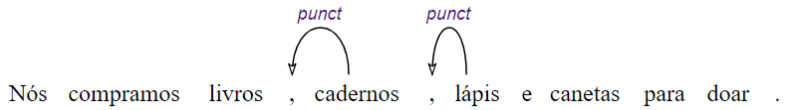


Figura 167 - deprel **punct** atribuída a vírgula que separa elementos coordenados

Para a vírgula, a regra é a mesma: se ela precede ou sucede um dependente, deve ser relacionada ao *head* desse dependente. Por exemplo, na Figura 168, a palavra “todavia” é dependente da relação **cc** e, por isso, a vírgula que a separa do restante da sentença deve ser ligada a ela.

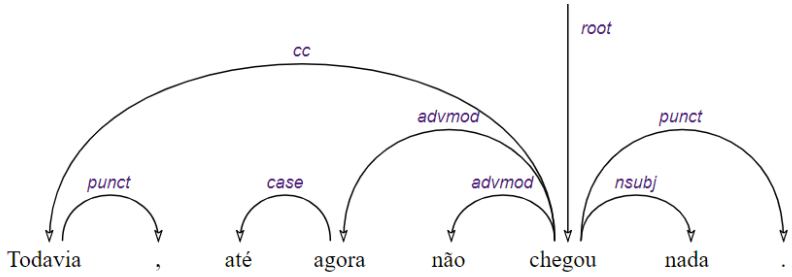


Figura 169 - deprel **punct** atribuída a vírgula que separa “todavia”, dependente de **cc**

reparandum: overridden disfluency = disfluência

A deprel **reparandum** ocorre entre dois elementos quando o segundo corrige a ocorrência do primeiro. Em geral, indica reparo de fala preservado na escrita.

Sentido da relação: a relação parte do segundo elemento (correto) em direção ao primeiro (errado), como exemplifica a Figura 170.

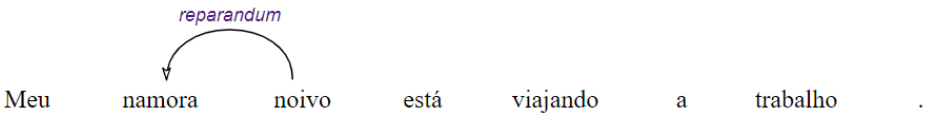


Figura 170 - Atribuição da deprel **reparandum**

root: root = raiz

A análise sintática de árvore de dependências inicia-se com a definição da raiz da sentença. Na UD, a deprel **root** tem essa função. A deprel **root** é uma relação artificial e é a única deprel que não possui um *head*. Ela vem de fora da sentença e aponta sua seta para o predicado da oração principal da sentença, como pode ser observado nas Figuras 171, 172, 173, 174 e 175.

Importante: cada sentença tem apenas um **root**. Se houver dúvida sobre qual predicado é o principal, leia a deprel **parataxis**.

Sentido da relação: a relação parte do elemento vazio em direção ao predicado mais importante da sentença.

Se o predicado da oração principal é um predicado verbal, o **root** é um verbo (Figura 171).

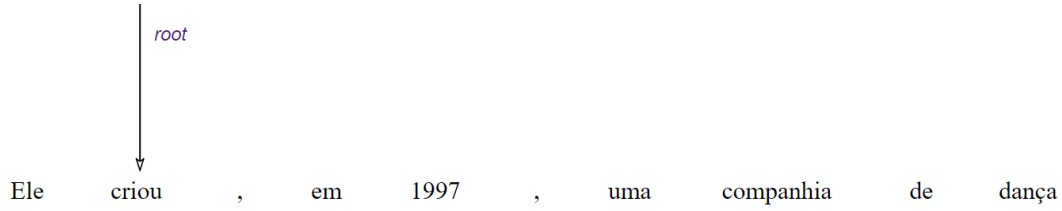


Figura 171 - Exemplo de atribuição da deprel **root** a predicado verbal

Para fins de determinação da raiz da sentença, não são considerados os verbos auxiliares (Figura 172).

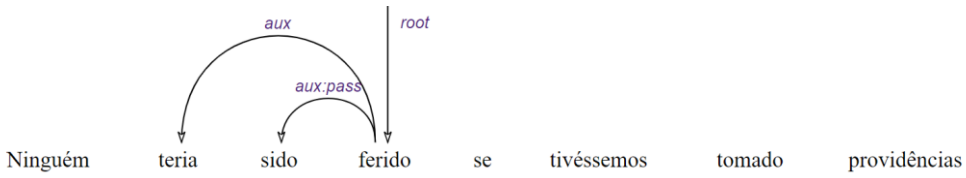


Figura 172 - Exemplo de atribuição da deprel **root** a verbo que possui dois auxiliares

É importante saber reconhecer a oração principal quando há vários verbos na sentença, pois é seu predicado que recebe a deprel **root**, como mostra a Figura 173.

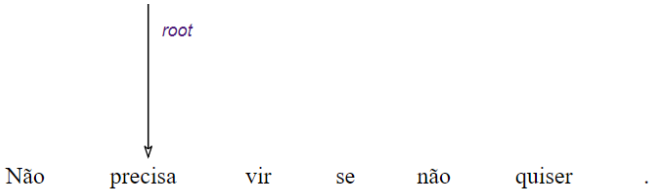


Figura 173 - Exemplo de atribuição da deprel **root** em sentença com três orações

Se o predicado da oração principal é um predicado nominal, a palavra principal do predicativo é que é o **root** da sentença (Figura 174). Lembramos que os predicados nominais são os predicativos, ligados ao sujeito pelos verbos de cópula. Na UD, só são considerados de cópula os verbos *ser* e *estar*.

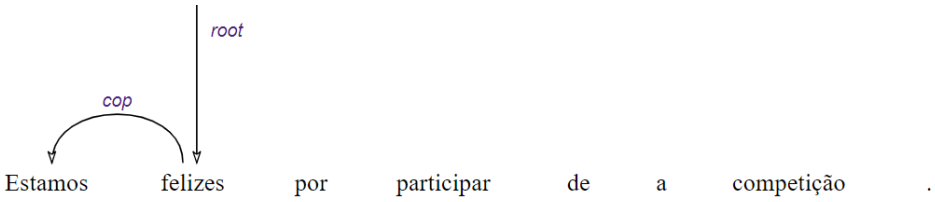


Figura 174 - Exemplo de atribuição da deprel **root** a predicado nominal

EXCEÇÃO: Na UD, nenhum verbo anotado com a *PoS tag* **AUX** pode ser **root**. Quando uma oração com verbo de cópula (**AUX**) tem um predicativo oracional, é necessário anotar o verbo de cópula como **root** para evitar associar dois sujeitos ao núcleo do predicativo (o sujeito da oração principal e o sujeito da oração subordinada). Nesses casos, a *PoS tag* do verbo de cópula terá que ser alterada de **AUX** para **VERB** (Figuras 175 e 176).

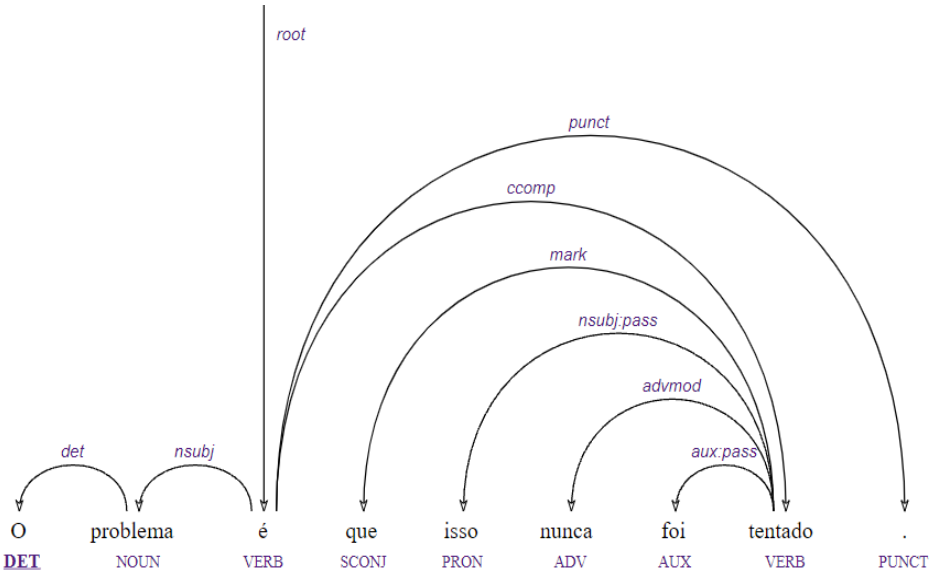


Figura 175 - Exemplo de atribuição da deprel **root** ao verbo de cópula pois o predicativo é oracional

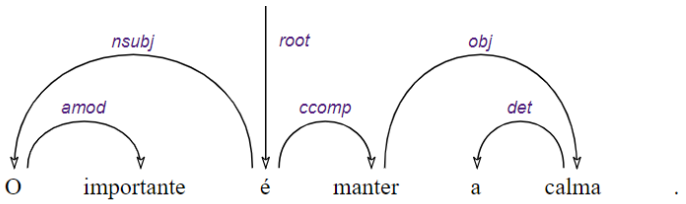


Figura 176 - Exemplo de atribuição da deprel **root** ao verbo de cópula pois o predicativo é oracional (nesse exemplo o sujeito da oração predicativa não está expresso)

vocative: vocative = vocativo

A deprel **vocative** é usada para marcar o participante do diálogo a quem se dirige a mensagem.

Sentido da relação: a deprel **vocative** pode ocorrer nos dois sentidos (Figuras 177 e 178) e o *head* é **root** da sentença.

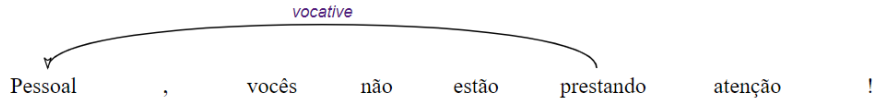


Figura 177 - Exemplo de atribuição da deprel **vocative** da esquerda para a direita

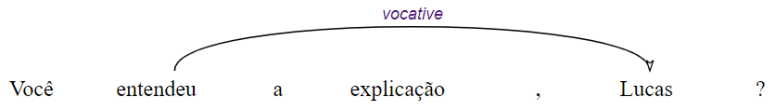


Figura 178 - Exemplo de atribuição da deprel **vocative** da direita para a esquerda

xcomp: open clausal compl. = complemento oracional aberto

Um complemento oracional aberto é uma oração que complementa o sentido de um predicado verbal ou nominal, mas não tem seu próprio sujeito. O sujeito do complemento oracional aberto é "controlado" pelo sujeito ou pelo objeto do predicado que ele complementa.[22]

Sentido da relação: a relação ocorre da esquerda para a direita.

Como pode ser observado na Figura 179, o sujeito da oração subordinada, "fazer", embora não possa ser expresso, é o mesmo de "tentaram", ou seja, "Eles".

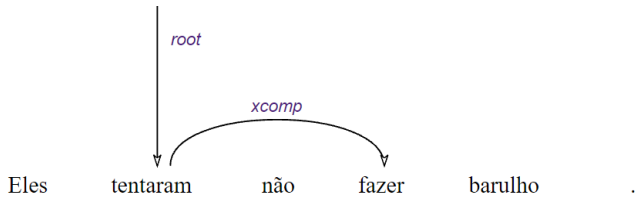


Figura 179 - atribuição de **xcomp** para a relação entre duas orações que têm o mesmo sujeito

Já na Figura 180, o sujeito nulo da oração subordinada é "controlado" pelo objeto da oração principal, ou seja, o sujeito de "abrir" é controlado por "João", que é objeto de "convenci".

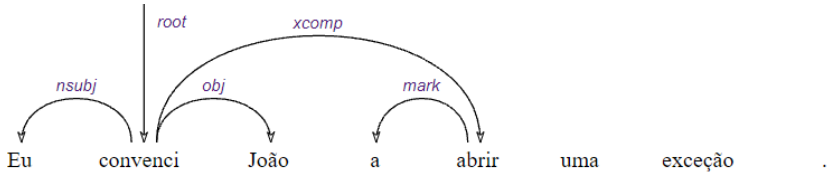


Figura 180 - sujeito da subordinada controlado pelo objeto da oração principal

Os casos de **xcomp** controlados pelo objeto da oração principal são todos de verbos bitransitivos que preveem um objeto direto e um indireto. São exemplos de verbos desse tipo: instigar, intimar, convidar, persuadir.

Há também verbos em que o objeto da oração principal "controla" o sujeito da oração subordinada sem interveniência de preposição: deixar (no sentido de permitir), mandar (no sentido de ordenar) e fazer (no sentido de causar). As Figuras 181, 182 e 183 ilustram esses três verbos.

[22] o termo **xcomp** foi tomado da LFG (*Lexical Functional Grammar*), uma evolução da teoria gerativa, e refere-se a complementos cujos sujeitos nulos têm *obligatory control*, ou seja, são controlados por uma função da oração *matrix* (oração principal ou subordinante). Os complementos **xcomp** podem ser *subject controlled* ou *object controlled* dependendo do termo da oração *matrix* que os controla, sujeito ou objeto. Para maiores detalhes: Joan Bresnan (1982). Control and Complementation. Linguistic Inquiry Vol. 13, No. 3, pp. 343-434. MIT Press.

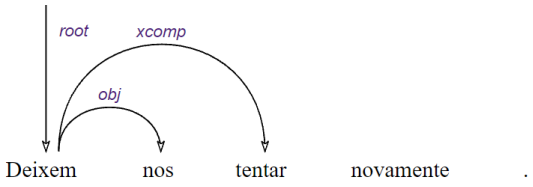


Figura 181 - sujeito nulo controlado pelo objeto do verbo “deixar”

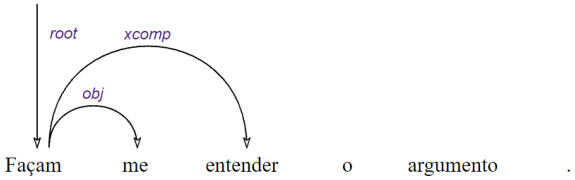


Figura 182 - sujeito nulo controlado pelo objeto do verbo “fazer”

O verbo “mandar” admite elipse do objeto, mas mesmo assim fica claro que o sujeito da oração subordinada é “controlado” pelo objeto elíptico da oração principal e por isso a relação entre as duas orações é **xcomp**, como ilustrado na Figura 183.

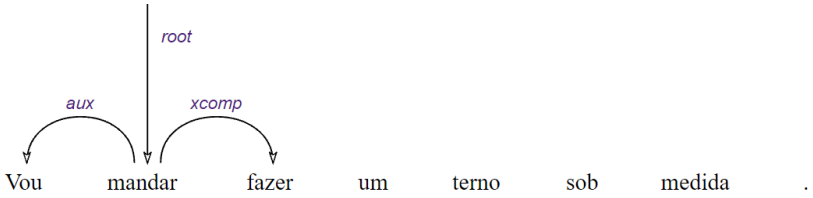


Figura 183 - objeto da oração principal elíptico, mas controla o sujeito da subordinada

Outros casos da deprel **xcomp** são observados entre os verbos em negrito nas sentenças a seguir (*head* em preto e dependente em azul).

- Todos **querem** **ganhar** um autógrafo do ator. (o sujeito de “ganhar”, embora não expresso, é o mesmo de “querem”, ou seja, “Todos”)
- Vocês **parecem** mais **magros**. (como o verbo “parecer” não é anotado como verbo de ligação na UD, ele é considerado um verbo pleno e seu complemento, representado por um adjetivo, é anotado com a deprel **xcomp**; é como se houvesse um verbo de cópula elíptico: Vocês parecem *estar* mais magros)
- Todos **começaram** a **aplaudi**-lo com entusiasmo. (o verbo “começar” seguido da preposição “a” e de outro verbo no infinitivo é um auxiliar de aspecto, porém tanto verbos aspectuais quanto modais são anotados como verbos plenos em nosso projeto)

A deprel **xcomp** também é usada para anotar o predicativo do objeto da oração principal, como mostrado na Figura 184.

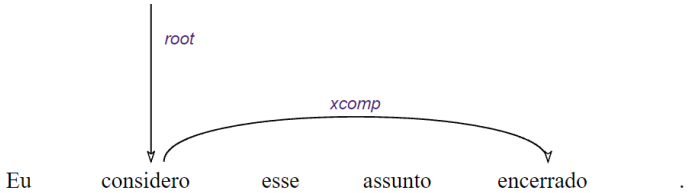


Figura 184 - Atribuição da deprel **xcomp** ao predicativo do objeto

Nesses casos, um teste para confirmar que se trata de um predicativo do objeto é transformar a sentença em voz passiva, como nos casos a seguir, nos quais o *head* está destacado em negrito e o dependente em azul.

- **Julgamos** todos vocês **responsáveis** pelo acidente. (Todos vocês foram julgados responsáveis pelo acidente.)
- **Considero** você **um irmão**. (Você é considerado um irmão.)
- O cobertor **manteve** o cachorro **aquecido**. (O cachorro foi mantido aquecido pelo cobertor.)
- Eu **considero** esse assunto **encerrado**. (O assunto foi considerado encerrado.)

- O prefeito **declarou** **aberta** a sessão. (A sessão foi declarada aberta.)
- Eles **acham** você **infantil**. (*[23]Você é achado infantil.)

Atenção: só são anotados como dependentes de **xcomp** os predicativos do objeto que são argumentos previstos na estrutura argumental do verbo. Predicativos do objeto que são opcionais (que podem ser omitidos sem prejuízo da gramaticalidade), como nas duas sentenças a seguir, são anotados como dependentes de **acl**.

- **Encontraram** o cofre **vazio**. (O cofre foi encontrado *vazio.)
- O artista **pintou** a modelo **nua**. (A modelo foi pintada *nua pelo artista)

Caso complexo: complemento oracional do verbo "considerar".

Quando o complemento do verbo "considerar" (a coisa considerada) for oracional, ao invés de ele ser **obj,** ele será **ccomp**. Ainda assim, o predicativo do objeto será **xcomp**, como mostra a Figura 185.

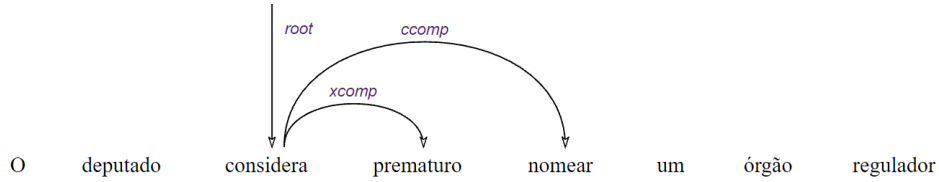


Figura 185 - Complemento oracional do verbo "considerar" (1)

Quando, contudo, o complemento inteiro do verbo principal for introduzido por uma conjunção subordinativa, como ilustrado nas Figuras 186 e 187, a oração "nomear" passa a ser **csubj** de "prematuro" e não mais complemento direto de "considera".

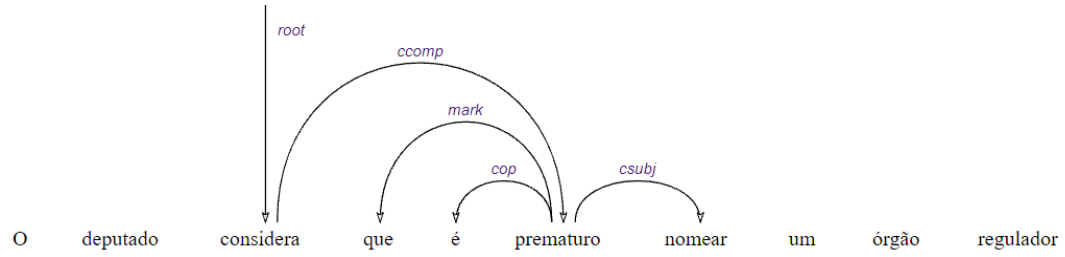


Figura 186 - Complemento oracional do verbo "considerar" (2)

[23] Embora "achar" no sentido de "considerar" admita predicativo do objeto, a passiva parece não ser aceita para evitar confusão com outros sentidos do verbo.

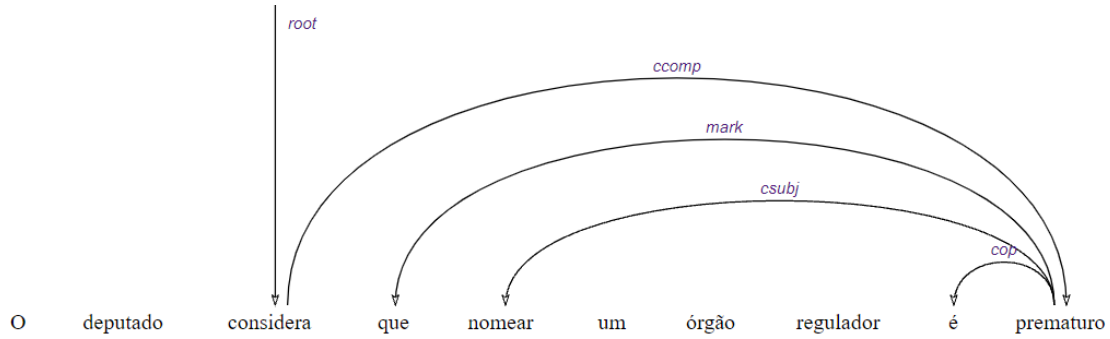


Figura 187 - Complemento oracional do verbo “considerar” (3)

Atenção: os verbos de estado (mudança de estado e manutenção de estado) não são anotados como cópula na UD, mas sim como verbos plenos. Seus complementos, sejam eles verbais ou nominais, são anotados como dependentes de **xcomp**. São exemplos desses verbos: andar, continuar, ficar, manter-se, permanecer.

- **Andam divulgando** um novo auxílio para os mais pobres.
- **Andam deprimidos** desde o acidente.
- **Continuamos torcendo** pela democracia.
- **Continuamos tristes** com o ocorrido.
- Ele **ficou olhando** para o nada.
- Ele **ficou louco** de ciúme.

Bibliografia

Afonso, S.; Bick, E.; Haber, R.; Santos, D. (2002). Floresta sintá(c)tica: um treebank para o português. In Anais do XVII Encontro Nacional da Associação Portuguesa de Linguística, pp. 533-545.

Andrews, A. D. (2007). The major functions of the noun phrase. In Timothy Shopen, editor, Language Typology and Syntactic Description. Volume I: Clause Structure. Second edition, pp. 132-223. Cambridge University Press, Cambridge, UK.

Bouma, G.; Hajic, J.; Haug, D.; Nivre, J.; Solberg, E.; Øvrelid, L. (2018). Expletives in Universal Dependency Treebanks. In the Proceedings of the Second Workshop on Universal Dependencies (UDW), pp. 18-26.

Bresnan, J. (1982). Control and Complementation. Linguistic Inquiry, Vol. 13, N. 3, pp. 343-434. The MIT Press.

Nivre, J. (2015). Towards a Universal Grammar for Natural Language Processing. In the Proceedings of the 16th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing), pp. 3-16.

Nivre, J.; Marneffe, M-C.; Ginter, F.; Hajič, J.; Manning, C.D.; Pyysalo, S.; Schuster, S.; Tyers, F.; Zeman, D. (2020). Universal Dependencies v2: An Evergrowing Multilingual Treebank Collection. In the Proceedings of the 12nd International Conference on Language Resources and Evaluation (LREC), pp. 4034-4043.

Pardo, T.A.S.; Duran, M.S.; Lopes, L.; Di Felippo, A.; Roman, N.T.; Nunes, M.G.V. (2021). Porttinari - a Large Multi-genre Treebank for Brazilian Portuguese. In the Proceedings of the XIV Symposium in Information and Human Language (STIL), pp. 1-10.

Rademaker, A.; Chalub, F.; Real, L.; Freitas, C.; Bick, E.; Paiva, V. (2017). Universal Dependencies for Portuguese. In the Proceedings of the 4th International Conference on Dependency Linguistics (Depling), pp. 197-206.

Souza, E.; Cavalcanti, T.; Silveira, A.; Evelyn, W.; Freitas, C. (2020). Diretivas e documentação de anotação UD em português (e para língua portuguesa). Disponível em: https://nbviewer.jupyter.org/github/comcorhd/Documenta-o-UD-PT/raw/master/Documenta-o-UD-PT.pdf

Thompson, S. A. (1997). Discourse Motivations for the Core-Oblique Distinction as a Language Universal Directions in Functional Linguistics. In: Akio Kamio (ed.), Studies in Language Companion, Series 36, pages 59-82. John Benjamins, Amsterdam.

Zeman, D. (2017). Core Arguments in Universal Dependencies. In the Proceedings of the Fourth International Conference on Dependency Linguistics (Depling), pp. 287-296.