
ANOTAÇÃO DE *ENHANCED DEPENDENCIES*

**ORIENTAÇÕES PARA ANOTAÇÃO DE RELAÇÕES DE DEPENDÊNCIA SINTÁTICA DO TIPO
ENHANCED EM LÍNGUA PORTUGUESA, SEGUINDO AS DIRETRIZES DA ABORDAGEM *UNIVERSAL
DEPENDENCIES* (UD)**

MAGALI SANCHES DURAN

Nº 448

RELATÓRIOS TÉCNICOS



São Carlos – SP
Ago./2024

**Relatório Técnico do
Núcleo Interinstitucional de Linguística Computacional (NILC)**

*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**

Anotação de *Enhanced Dependencies*

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

Magali Sanches Duran

Agosto/2024

Agradecimentos:

Ao meu supervisor de pós-doc, Thiago Alexandre Salgueiro Pardo, por todo apoio e incentivo nessa nova etapa de criação de diretrizes de anotação no esquema Universal Dependencies;

À colega Maria das Graças Volpe Nunes, pela leitura crítica minuciosa de todo o conteúdo, e pelas sugestões que muito contribuíram para melhorar a clareza das explanações;

À colega Adriana Silvina Pagano, pelas sugestões e por ter me incentivado a mergulhar mais fundo na exploração das *enhanced dependencies*;

Ao colega Elvis de Souza, pela leitura, sugestões e, principalmente, por ter customizado a ferramenta *enhanced-graph-visualiser*¹, utilizada para gerar as ilustrações deste documento;

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

¹ <https://souelvis.dev/enhanced-graph-visualiser/visualise.html>

Sumário

Introdução.....	5
A quem se destina esse relatório.....	5
O que são e para que servem as enhanced dependencies.....	6
Tipos de enhanced dependencies previstos nas diretrizes da UD.....	8
Automatização da anotação de enhanced dependencies.....	9
O caráter de recursividade das enhanced dependencies.....	10
Organização do relatório.....	10
1. Inserção de um token vazio para representar um predicado elíptico.....	12
1.1 Atribuição da relação orphan nas basic dependencies.....	14
1.2 Inserção do token vazio nas enhanced dependencies.....	15
1.3 Questões envolvendo os diferentes “órfãos” de um predicado elíptico.....	19
1.4 Mais de um token elíptico.....	22
1.5 A questão da posição de inserção do predicado elíptico.....	23
1.6 Predicados elípticos com mais de dois órfãos.....	24
1.7 Predicados elípticos com um único órfão.....	25
1.8 Conclusão.....	26
2. Propagação de heads de tokens coordenados pela relação conj.....	28
2.1 Propagação de head de tokens coordenados que possuem POS tags distintas.....	29
2.2 Propagação de head de tokens de palavras funcionais coordenados.....	30
2.3 Head que não propaga do head para o dependente de conj.....	31
3. Propagação de dependentes do head de conj.....	32
3.1 Propagação do sujeito do head para o dependente de conj.....	32
3.2 Propagação de case e mark.....	35
3.3 Relações que não são propagadas.....	36
3.4 Propagação condicional de dependentes de conj.....	38
3.5 Conclusões sobre a propagação de dependentes do head de conj.....	41
4. Atribuição do sujeito de xcomp.....	42
4.1 Sujeito do dependente de xcomp controlado pelo sujeito do head de xcomp.....	43
4.2 Sujeito do dependente de xcomp controlado pelo obj do head de xcomp.....	46
4.3 Sujeito do dependente de xcomp controlado por iobj ou obl do head de xcomp.....	51
4.4 Objetos diretos e indiretos antepostos ao head de xcomp.....	52
4.5 Orações infinitivas que não são xcomp.....	54
4.6 Conclusões sobre a atribuição de sujeito de xcomp.....	56
5. Correferência nas acl:relcl.....	57
5.1 Pronome relativo exercendo função de advmod.....	59
5.2 Pronomes relativos que são seu próprio antecedente.....	59
5.3 Pronomes relativos que são o predicado da acl:relcl.....	61

5.4. Pronome relativo não ligado diretamente ao predicado da acl:relcl.....	62
5.5 Conclusões sobre a enhanced dependency de correferência nas acl:relcl.....	63
6. Inserção de caso.....	64
6.1 Inserção de caso propagada em coordenações aditivas.....	65
6.2 Conclusões sobre a enhanced dependency de inserção de caso.....	66
7. Novas enhanced dependencies, não descritas na UD.....	67
7.1 Propagação do sujeito de ccomp.....	67
7.2 Propagação do sujeito de ccomp:speech.....	70
7.3 Propagação do sujeito de advcl.....	71
7.4 Atribuição de sujeito de acl.....	78
7.5 Realocação de complementos separados de seus heads lógicos.....	81
Conclusões.....	87
Bibliografia.....	88

Diretrizes de Anotação de *Enhanced Dependencies*

Introdução

Apresenta-se neste relatório os resultados de pesquisa desenvolvida no âmbito do projeto POeTiSA² (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 POeTiSA é um projeto de longo prazo que visa aumentar os recursos baseados em sintaxe e desenvolver ferramentas e aplicações relacionados à língua portuguesa do Brasil, buscando alcançar resultados de ponta nesta área. Nos últimos quatro anos o projeto vem produzindo, entre outros recursos, um córpus multigênero grande e abrangente, chamado Porttinari (Pardo et al. 2021), anotado segundo o modelo Universal Dependencies (UD) (Nivre, 2015; Nivre et al., 2020), do qual o Porttinari-base é a primeira parte concluída (Duran et al., 2023). No período compreendido entre setembro de 2023 e junho de 2024, foram dedicados esforços à definição de regras para a implementar uma nova camada de anotação no Porttinari-base, chamada de *enhanced dependencies*. Essa camada de anotação é construída sobre a camada de anotação de dependências sintáticas e representa as inferências passíveis de serem feitas a partir da sintaxe.

A quem se destina esse relatório

O relatório foi concebido como complemento aos manuais de anotação de *Pos tags*³ e de relações de dependência⁴ e tem por objetivo apresentar as diretrizes que norteiam a anotação das chamadas *enhanced dependencies* em córpus de língua portuguesa, seguindo proposta adotada pela abordagem Universal Dependencies (UD) (Nivre et al. 2018; de Marneffe et al. 2021), instanciadas em língua portuguesa por Pagano, Duran & Pardo (2023), bem como outras *enhanced dependencies* concebidas originalmente no âmbito do POeTiSA.

² <https://sites.google.com/icmc.usp.br/poetisa>

³ https://drive.google.com/file/d/1BddPswn-_Ioo-A5GsldA1cO1kqbcCahb/view?usp=sharing

⁴ <https://drive.google.com/file/d/1ile8Wfxu1qdrZOmLGqkvVuQ4fXvHqVMo/view?usp=sharing>

O conhecimento aqui descrito representa fonte de informação para: 1) quem revisa *enhanced dependencies*; 2) quem se dedica à tarefa de customizar um programa de anotação de *enhanced dependencies* para a língua portuguesa e 3) quem pretende utilizar as *enhanced dependencies* em aplicações de PLN.

O que são e para que servem as *enhanced dependencies*

As *enhanced dependencies* são uma proposta de anotação desenvolvida inicialmente para a língua inglesa (Schuster & Manning, 2016) e depois expandida pela abordagem UD (Nivre et al. 2018). Essa anotação representa um enriquecimento no arquivo utilizado pela abordagem UD para anotação de córpus, um formato conhecido como CoNLL-U (*Computational Natural Language Learning - Universal*).

O formato CoNLL-U possui 10 colunas e cada coluna é dedicada a um tipo de informação sobre cada *token* que compõe uma sentença, incluindo pontuações. As *enhanced dependencies* são anotadas na coluna 9, intitulada DEPS. A seguir, na Figura 1, são apresentadas as colunas do CoNLL-U, seus respectivos nomes e a descrição de seus respectivos conteúdos:

1	2	3	4	5	6	7	8	9	10
ID ⁵	FORM	LEMMA	UPOS	XPOS	FEATURES	HEAD	DEPREL	DEPS	MISC
número do token na sentença	forma do token	lema do token	categoria morfossintática universal do token na sentença	categoria morfossintática do token na língua (opcional)	atributos morfológicos do token na sentença	head da relação na qual o token é dependente	relação (única) da qual o token é dependente	enhanced dependencies (pode haver mais de uma para o token)	miscelânea (uso liberado)

Figura 1. Ilustração do formato de um arquivo CoNLL-U, com suas 10 colunas, respectivos nomes abreviados e respectivas descrições sobre o conteúdo.

A coluna 9, ou DEPS, do CoNNL-U:

- reproduz, para a maioria dos *tokens*, o conteúdo das colunas 7 e 8, chamadas de *basic dependencies*;
- acrescenta novas relações de dependência, inferíveis a partir da análise da sentença, mas que não são permitidas nas *basic dependencies*, pois implicam cruzamento de arcos e/ou mais de um *head* por dependente;

⁵ O ID é o número sequencial atribuído a cada palavra na sentença, sempre começando do 1. Para contrações e multipalavras, é um intervalo e para *tokens* vazios (que representam predicados elípticos nas *enhanced dependencies*) é um número decimal.

- apresenta, para alguns *tokens*, uma anotação alternativa às *basic dependencies*, particularmente para lidar com predicados elípticos e correferência em orações relativas.

Para exemplificar, tomemos a sentença “João acordou e saiu”. Nas *basic dependencies* (colunas 7 e 8 do CoNLL-U), “João” é **nsubj** de “acordou”. Já nas *enhanced dependencies* (coluna 9 do CoNLL-U) “João” é **nsubj** de “acordou” e de “saiu”. Isso pode ser observado nas Figuras 2a e 2b que ilustram respectivamente as *basic dependencies* e as *enhanced dependencies*.

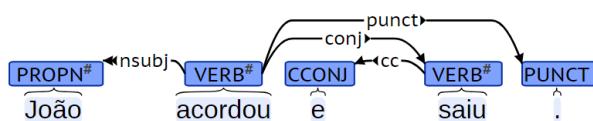


Figura 2a. *Basic dependencies* da sentença “João acordou e saiu”

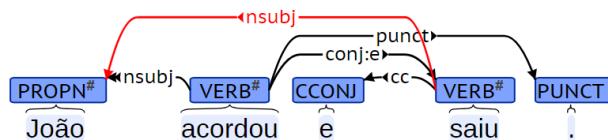


Figura 2b. *Enhanced dependencies* da sentença “João acordou e saiu”

A diferença entre as duas anotações é que nas *enhanced dependencies* foi acrescentada a relação **nsubj** entre “saiu” e “João”. Essa relação não é permitida nas *basic dependencies* porque fere o princípio de que cada dependente deva ter um único *head*, além do fato de que produz um cruzamento de arcas. Sendo assim, como numa árvore uma mesma folha não pode sair de vários galhos, as *enhanced dependencies*, embora constituam um grafo, não constituem uma árvore.

A Figura 3 mostra como fica o arquivo CoNLL-U (que representa a anotação dessa sentença em forma de tabela) após a inclusão das *enhanced dependencies*:

# text = João acordou e saiu.						
1 João	João	PROPN	_	Gender=Masc Number=Sing	2 nsubj	2:nsubj 4:nsubj TokenRange=0:4
2 acordou	acordar	VERB	_	Mood=Ind Number=Sing Person=3 Tense=Past VerbForm=Fin	0 root	0:root TokenRange=5:12
3 e	e	CCONJ	_		4 cc	4:cc TokenRange=13:14
4 saiu	sair	VERB	_	Mood=Ind Number=Sing Person=3 Tense=Past VerbForm=Fin	2 conj	2:conj:e SpaceAfter=No TokenRange=15:19
5 .	.	PUNCT	_		2 punct	2:punct SpacesAfter=r\n r\n TokenRange=19:20

Figura 3. Arquivo CoNLL-U da sentença “João acordou e saiu”

Observa-se na Figura 3 que, na coluna 9, o *token* 1, “João”, é dependente de mais de um *head*: 2:nsubj (sujeito do *token* 2) e 4:nsubj (sujeito do *token* 4).

As *enhanced dependencies* aumentam o número de informações recuperáveis a partir da anotação sintática, o que enriquece o resultado de aplicações que costumam usar a sintaxe na fase de pré-processamento, como a extração de informações e a anotação de papéis semânticos, só para citar algumas.

Um exemplo bem simples é mostrado na sentença a seguir. A sentença apresenta cinco predicados verbais (em negrito), mas apenas o primeiro deles tem sujeito anotado com a relação **nsubj**. Pode-se inferir que os outros quatro verbos têm o mesmo sujeito, porém isso não pode ser anotado nas *basic dependencies*, pois um dependente pode ter um único *head*, ou seja o sujeito só pode pertencer a um predicado. Já as *enhanced dependencies* não têm essa restrição e por isso permitem anotar os sujeitos dos outros quatro verbos. Com isso, o token “Battisti” passa a ser sujeito de cinco predicados simultaneamente, o que deixa a estrutura argumental desses verbos mais completa, passando a fornecer mais respostas para perguntas feitas a partir deles, mais suportes sobre os quais atribuir papéis semânticos, e orações mais simples para processamentos complexos. Os sujeitos implícitos, inferíveis para serem mapeados nas *enhanced dependencies*, estão representados entre colchetes, destacados em azul.

Battisti foi **detido** em a semana passada [Battisti] **acusado** de evasão de divisas ao [Battisti] **tentar** [Battisti] **atravessar** a fronteira com a Bolívia [Battisti] **portando** dólares e euros no valor equivalente a pouco mais de R\$ 23 mil.
FOLHA_DOC000007_SENT007

No arquivo CoNLL-U dessa sentença, a coluna 9 (DEPS), linha 1, relativa ao token 1 (“Battisti”) apresenta a informação de todas as relações das quais “Battisti” é dependente: 3:nsubj:pass|8:nsubj:pass|15:nsubj|16:nsubj|22:nsubj. Lê-se: sujeito da passiva do *token* 3, (“detido”); sujeito da passiva do *token* 8 (“acusado”); sujeito do *token* 15 (“tentar”); sujeito do *token* 16 (“atravessar”), e sujeito do *token* 22 (“portando”).

Tipos de *enhanced dependencies* previstos nas diretrizes da UD

Segundo as diretrizes da UD⁶, seis tipos de *enhanced dependencies* devem ser anotadas (v. Pagano, Duran & Pardo, 2023, para uma introdução):

1. Inserção de um *token* vazio para representar predicados elípticos que tenham no mínimo dois dependentes sem *head*. Nas *basic dependencies*, os dependentes de um predicado elíptico estão “órfãos” de *head* e por isso são ligados um ao outro pela

⁶ <https://universaldependencies.org/u/overview/enhanced-syntax.html>

relação **orphan**. A relação **orphan** é usada como “pista” para a inserção do *token* vazio nas *enhanced dependencies*. O *token* vazio no lugar do predicado elíptico passa a ser *head* da relação que cada dependente órfão teria com o predicado elíptico.

2. Propagação, para o dependente de **conj**, da relação que *chega* ao head de **conj**. Isso significa que, se dois elementos são coordenados, o segundo elemento é dependente da mesma relação da qual o primeiro elemento é dependente.
3. Propagação, para o dependente de **conj**, de relações que *partem* do head de **conj**. Isso significa que, se dois elementos são coordenados, o segundo elemento é *head* das mesmas relações das quais o primeiro elemento é *head*.
4. Identificação do sujeito controlador da **xcomp**. A relação **xcomp**, por definição, tem como dependente uma oração subordinada sem sujeito (sujeito nulo). Contudo, é possível inferir qual *token* da oração matriz é o “sujeito controlador” desse sujeito nulo.
5. Correferência em orações relativas (introdução da relação **ref**). O pronome relativo de uma **acl:relcl** é ligado a seu antecedente pela relação **ref** e a relação da qual era dependente é transferida para o antecedente.
6. Inclusão do *token* dependente de **mark**, **cc** e **case** nas relações modificadoras cujos dependentes são introduzidos por eles. É uma informação extra em relações já existentes nas relações básicas.

Automatização da anotação de *enhanced dependencies*

Como as *enhanced dependencies* podem ser deduzidas logicamente a partir das *basic dependencies*, costuma-se utilizar um programa, baseado em regras, para fazer uma pré-anotação. Na sequência, os anotadores humanos conferem a anotação e fazem correções, se necessário.

Em 2020 houve uma *shared task* para tratar da automatização das *enhanced dependencies* (Bouma, Seddah & Zeman, 2020). A base de dados para avaliação foi constituída de 28 córpuses de UD anotados com *enhanced dependencies*, num total de 17 línguas (para algumas línguas havia mais de um córpus). Nenhum córpus de português fez parte do conjunto usado na tarefa. A proposta da *shared task* foi chegar à mesma anotação apresentada no conjunto de dados, usando regras.

Essa *shared task* mostrou que a dificuldade de implementar cada tipo de *enhanced dependency* varia, e que a *enhanced* número 1, que insere um *token* vazio para representar um predicado elíptico, é a mais complexa.

O caráter de recursividade das *enhanced dependencies*

Uma característica das *enhanced* é seu efeito em cascata, pois, uma vez inserida uma *enhanced*, cria-se uma base para atribuições de outras *enhanced*. Por exemplo, se há uma cadeia de orações que compartilham o mesmo sujeito, mas só a primeira apresenta sujeito explícito, é preciso que esse sujeito seja propagado para a segunda oração antes de ser propagado para a terceira e assim por diante. A sentença a seguir apresenta essa situação, pois “governo” é sujeito das 3 orações cujos predicados estão em negrito:

“Dois dias após a publicação do material pelo jornal, o **governo** paulista **revogou** os decretos de sigilo e se **comprometeu** a **revisar** todos os termos semelhantes impostos até então.”

As *enhanced dependencies* ilustradas na Figura 4 apresentam o resultado dessa propagação de sujeito em cadeia. Primeiro é propagado o sujeito do *head* de **conj** “revogou” para o dependente de **conj** “comprometeu” e a seguir é propagado o sujeito do *head* de **xcomp** “comprometeu” para o dependente de **xcomp** “revisar”. Não haveria como inverter essa ordem⁷ porque “revisar” não está ligado diretamente a “revogou”.

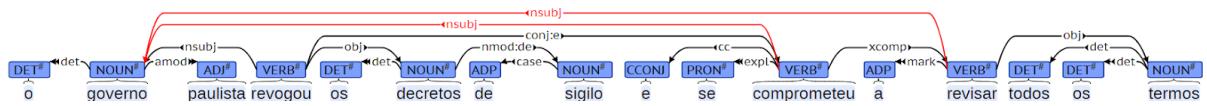


Figura 4. Propagação de sujeito em cadeia

Organização do relatório

Cada tipo de *enhanced* possui particularidades e, para detalhá-las, decidiu-se dedicar uma seção a cada uma delas. Além dos seis tipos de *enhanced* já descritos anteriormente, percebeu-se a possibilidade de propagar sujeitos de outros tipos de orações: **ccomp**, **acl** e **advcl**, uma característica típica das línguas que admitem elipse do sujeito (chamadas “línguas *pro-drop*”). Uma outra oportunidade de *enhanced* no português é a realocação do *head* de relações que geram cruzamento de arcos (anotados indevidamente apenas para evitar cruzamento de arcos nas *basic dependencies*).

⁷ Como não há uma sequência ideal para inserir cada tipo de *enhanced*, um programa gerador de *enhanced* deve passar várias vezes pelo mesmo córpus e a cada vez pode apresentar novas *enhanced*, derivadas das *enhanced* das rodadas anteriores, parando quando na rodada que não apresentar mais nenhum incremento.

Como as diretrizes da UD não autorizam⁸, até o momento, outras *enhanced* além dos seis tipos já definidos, todas as novas *enhanced* foram reunidas numa seção denominada “Novas *Enhanced Dependencies*”. Este documento apresenta, portanto, sete seções:

1. Inserção de um *token* vazio para representar um predicado elíptico;
2. Propagação de *heads* de *tokens* coordenados pela relação **conj**;
3. Propagação de dependentes do *head* de **conj**;
4. Atribuição do sujeito de **xcomp**;
5. Correferência nas **acl:relcl**;
6. Inserção de caso;
7. Novas *enhanced dependencies* não descritas na UD.

⁸ A UD não autoriza que os córpus disponibilizados em seu site tenham novas *enhanced*, mas há algumas que são universais e estão em discussão pelo grupo gestor. Assim, pode-se ter uma versão de anotação *enhanced* para disponibilizar na UD e outra versão, mais enriquecida, para uso independente. Vide: <https://github.com/UniversalDependencies/docs/issues/1038>

1. Inserção de um *token* vazio para representar um predicado elíptico

A mais complexa das *enhanced dependencies* é a de inserção de *token* vazio para representar um predicado elíptico. Primeiramente, é preciso entender a motivação dessa *enhanced*. Como o predicado é *head* das várias relações estabelecidas dentro de uma oração, sua elipse impede que essas relações sejam estabelecidas. Assim, inserir um *token* vazio para representar um predicado elíptico é essencial para explicitar todas as funções sintáticas dentro da oração.

Como regra geral, um predicado só pode sofrer elipse se for redundante no contexto. Daí a hipótese de que todo predicado elíptico possa ser inferido a partir do contexto. Essa operação de “preencher” as elipses é feita naturalmente pelos falantes da língua. Porém, quando a finalidade do insumo linguístico é ser processado por máquinas, ou ser consumido por pessoas com baixo letramento ou com dificuldades cognitivas, o ideal seria que as elipses de predicado não ocorressem, ou seja, o ideal seria que essas informações estivessem explícitas. Preencher os vazios deixados pelos predicados elípticos, a partir de “pistas” do contexto, é o objetivo da *enhanced dependency* discutida nesta seção.

Os casos de predicados elípticos tratados pelas *enhanced dependencies* ocorrem majoritariamente em orações coordenadas (ligadas por **conj**), sendo que o predicado elíptico na oração dependente de **conj** coincide com o predicado explícito na oração *head* de **conj**. Os exemplos a seguir buscam ilustrar essa situação: o predicado explícito, na primeira coordenada, foi destacado em vermelho, e o predicado elíptico (não presente na segunda oração coordenada) foi preenchido entre colchetes, em azul.

- Estou **indo** para os EUA, não **[indo]** para a Coreia do Norte, afirmou.
- A adoção tem que ser **pensada** com a cabeça, e não **[pensada]** só com o coração.
- A Netflix **vem** em seguida, com 91; depois **[vem]** a NBC, com 60.
- Um copo de mate **custa** R\$ 7, e um chope **[custa]**, R\$12.
- Por que esses países **conseguiram**, e não **[conseguiram]** os catalães?

Em outras palavras, o predicado da oração dependente de **conj** está elíptico porque foi considerado redundante e é redundante porque é igual ao predicado da oração *head* de **conj**. Embora eventualmente o local da elipse seja marcado por vírgula, esse sinal de pontuação nem sempre está presente para ser considerado uma característica definidora do fenômeno de elipse de predicado.

Na anotação das *basic dependencies*, toda vez que ocorre um predicado elíptico, os *tokens* restantes da oração são unidos entre si pela relação **orphan**. A relação **orphan** une, por

exemplo, um *token* que seria sujeito do predicado elíptico a um *token* que seria objeto do predicado elíptico. A Figura 5 ilustra a anotação de **orphan**:

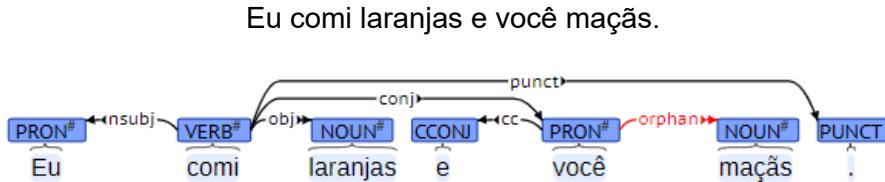


Figura 5. Anotação da relação **orphan** nas *basic dependencies*

O *token* “você” e o *token* “maçãs” são, respectivamente, sujeito e objeto de um mesmo predicado, porém não podem ser dependentes das relações **nsubj** e **obj**, pois o predicado *head* dessas relações (“comeu”) está elíptico. Diz-se, nessa situação, que “você” e “maçãs” são “órfãos” de um mesmo *head* e, por isso, são ligados pela relação **orphan**.

A relação **orphan** não corresponde a uma função sintática e foi criada apenas para dar uma solução de anotação para *tokens* “órfãos” de um *head* e para marcar onde houve elipse de um predicado.

Por isso, a relação **orphan** é o ponto de partida para a construção da *enhanced dependency* que insere um *token* vazio para representar o predicado elíptico.

A Figura 6 mostra a anotação, nas *enhanced dependencies*, da mesma sentença ilustrada na figura anterior. A relação **orphan** das *basic dependencies* não é reproduzida nas *enhanced dependencies* e um *token* vazio, chamado arbitrariamente de **NOFORM**, é inserido para representar o predicado elíptico (“comeu”).

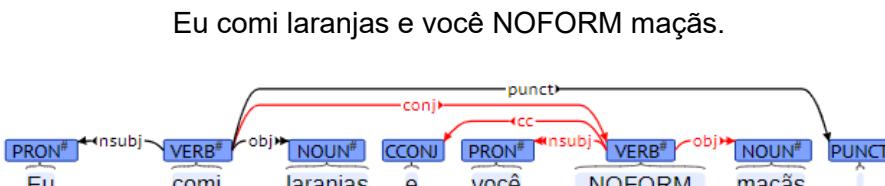


Figura 6. Inserção de *token* vazio **NOFORM** no lugar do predicado elíptico

Como pode ser observado, a inserção do *token* vazio permitiu revelar a relação de **nsubj** da qual o *token* “você” é dependente, bem como a relação de **obj**, da qual o *token* “maçãs” é dependente.

Dada a importância da relação **orphan** para a *enhanced dependency* discutida nesta seção, julgou-se procedente retomar, na Subseção 1.1, o tópico de atribuição de **orphan**. A inserção do *token* vazio da *enhanced dependency* é detalhada na Subseção 1.2.

1.1 Atribuição da relação **orphan** nas *basic dependencies*

A relação **orphan** liga os *tokens* que estão “órfãos” devido à elipse do predicado. O *head* da relação **orphan** é o núcleo da oração cujo predicado está elíptico. A escolha de qual dos dependentes “órfãos” será o *head* da relação **orphan** exige que o anotador “imagine” as relações que ligariam cada *token* ao predicado caso o mesmo não estivesse elíptico. Uma vez “imaginadas” tais relações, segue-se uma ordem de prioridade. Essa ordem estabelece que o sujeito da oração (**nsubj** ou **nsubj:pass**) é o herdeiro natural do predicado elíptico; na ausência do sujeito, o herdeiro natural é o objeto (**obj**) e assim por diante. A ordem de sucessão dos herdeiros (“órfãos”) do predicado elíptico, para fins de determinar o *head* da relação **orphan**, é a seguinte:

nsubj > obj > iobj > obl > advmod > csubj > xcomp > ccomp > advcl

Atribuir a relação **orphan** já é, em si, uma tarefa difícil durante a anotação das *basic dependencies*, e muitas vezes essa relação é atribuída indevidamente, pois embora a UD estabeleça que apenas a elipse de predicado deva ser tratada com **orphan**, alguns anotadores, equivocadamente, atribuem **orphan** a outros casos de elipse.

Há dois casos de atribuição indevida que se observa nos córpuses de português na UD: 1) elipse de verbo de cópula ou de verbo auxiliar e 2) elipse do verbo pleno e manutenção do verbo de cópula ou auxiliar. Nas duas subseções a seguir esses dois casos serão discutidos e exemplificados.

1.1.1 Elipse de verbo de cópula ou verbo auxiliar

Se o predicado de uma oração está explícito, não há motivo para anotar a relação **orphan**, mesmo que um verbo de cópula ou um verbo auxiliar estejam claramente elípticos. As Figuras 7 e 8 ilustram casos assim.

Aqui o espaço é limitado, e a convivência, [é] intensa.

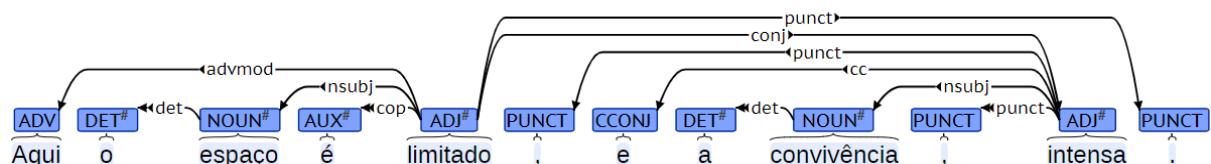


Figura 7. Oração dependente de **conj** com elipse de verbo de cópula

Ele tinha acusado o menino e eu, [tinha] defendido.

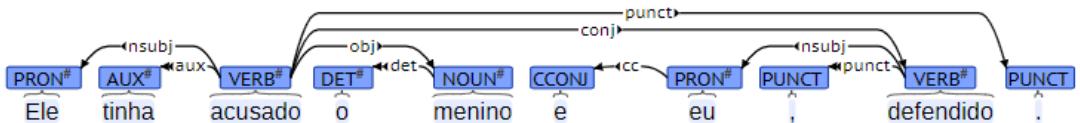


Figura 8. Oração dependente de **conj** com elipse de verbo auxiliar (**aux**)

1.1.2 Elipse do verbo pleno e presença do verbo de cópula ou do verbo auxiliar

Os verbos de cópula e os verbos auxiliares não são considerados órfãos dos predicados elípticos, mas seus substitutos naturais. Assim, se um verbo pleno estiver elíptico, mas houver um verbo de cópula ou um verbo auxiliar remanescente, o verbo remanescente passará a desempenhar as funções de núcleo da oração. As Figuras 9 e 10 ilustram essa situação.

Sua casa nunca foi invadida, mas a minha já foi [invadida].

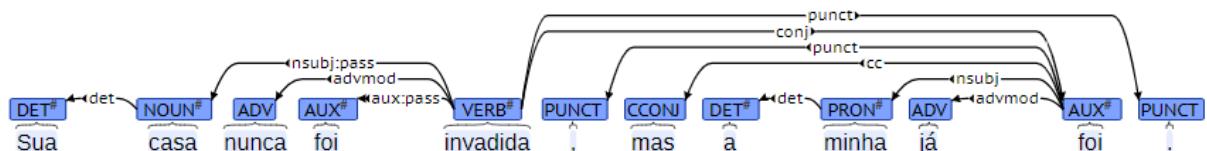


Figura 9. Oração dependente de **conj** com elipse do verbo e manutenção do **auxiliar**

Você não é amigo dele, mas eu sou [amigo dele].

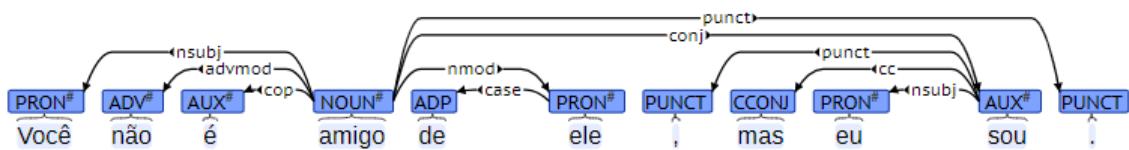


Figura 10. Oração dependente de **conj** com elipse do verbo e manutenção do **cópula**

1.2 Inserção do *token* vazio nas *enhanced dependencies*

A inserção de um *token* vazio para representar um predicado elíptico não altera o córpus, uma vez que o *token* vazio inserido não recebe um número de *token* válido. A inserção do *token* vazio (chamado genericamente de NOFORM⁹) se dá por meio da criação de um sub-*token* “pendurado” no *token* à esquerda do local da inserção. O arquivo CoNLL-U (Figura 11) mostra isso: como o *token* à esquerda de NORFORM é o de número 5, o *token* vazio inserido recebe o número 5.1.

⁹ Na verdade, esse é o nome do *token* inserido na *Shared Task*.

# text = Eu comi laranjas e você maçãs.									
1	Eu	eu	PRON	_	Case=Nom Gender=Masc Number=Sing Person=1 PronType=Prs	2	nsubj	2:nsubj	TokenRange=0:2
2	comi	comer	VERB	_	Mood=Ind Number=Sing Person=1 Tense=Past VerbForm=Fin	0	root	0:root	TokenRange=3:7
3	laranjas	laranja	NOUN	_	Gender=Fem Number=Plur	2	obj	2:obj	TokenRange=8:16
4	e	e	CCONJ	_	_	5	cc	5.1:cc	TokenRange=17:18
5	você	você	PRON	_	Case=Nom Number=Sing Person=3 PronType=Prs	2	conj	5.1:nsubj	TokenRange=19:23
5.1	_NOFORM_	comer	VERB	_	_	_	_	wordform=_EMPTY_	
6	maçãs	maçã	NOUN	_	Gender=Fem Number=Plur	5	orphan	5.1:obj	SpaceAfter=No TokenRange=24:29
7	.	.	PUNCT	_	_	2	punct	2:punct	SpacesAfter=\n TokenRange=29:30

Figura 11. Arquivo CoNLL-U mostrando a inserção de um *token* vazio “NOFORM”.

Como o *token* vazio é um *token* artificial, decidiu-se manter sua identidade artificial com a forma NOFORM. Se, contudo, o objetivo fosse realmente alterar a sentença do córpus, inserindo o *token* vazio como um *token* numerado da sentença, a forma do verbo “comer” teria que ser ajustada da primeira pessoa do singular (“comi”), explícita na oração *head* de **conj**, para a terceira pessoa do singular (“comeu”).

Pode-se observar também no arquivo CoNLL-U, ilustrado na Figura 11, que a relação **orphan** existente na coluna 8 (*basic dependencies*) não é reproduzida na coluna 9 (*enhanced dependencies*), embora seja fundamental como “pista” para a inserção do *token* vazio NOFORM na coluna 9. Portanto, a primeira consequência de se inserir um *token* vazio para representar um predicado elíptico é eliminar a necessidade da relação **orphan**. Por isso, na coluna das *enhanced dependencies*, não há nenhuma relação **orphan** anotada.

O *token* vazio inserido passa a representar o predicado elíptico da oração, assumindo a função de dependente de **conj** (relação que anteriormente chegava no *head* de **orphan**) e a função de *head* de **cc**¹⁰ (relação que anteriormente partia do *head* de **orphan**). Em seguida, novas relações são estabelecidas entre o *token* vazio inserido e cada um dos elementos que participavam da relação **orphan** nas *basic dependencies*.

Usando uma analogia, os dependentes “órfãos” do predicado elíptico, que se ligavam pela relação **orphan** nas *basic dependencies*, são “adotados” pelo novo “pai” (o *token* vazio), e passam a ser seus dependentes.

Uma vez inserido o *token* vazio e estabelecidas suas relações com seus dependentes, a próxima tarefa é nomear essas novas relações. Essa tarefa pode ser relativamente simples para um anotador humano, porém é difícil de ser automatizada. Dada a simetria entre as duas orações (o que possibilitou a elipse), espera-se que o *token* vazio, na estrutura oracional que está sendo reconstruída, seja *head* de relações do mesmo tipo que aquelas das quais o predicado da primeira oração é *head*. Em outras palavras, as etiquetas

¹⁰ Desde que se trate de uma coordenação sindética, isto é, feita por meio de uma conjunção.

disponíveis para nomear as relações estabelecidas entre o *token* vazio e seus dependentes são, a princípio, as etiquetas das relações presentes no *head* de **conj**.

Por exemplo, na sentença ilustrada na Figura 12a, pode-se observar que a primeira oração apresenta um **advmod** (“aqui”), um **nsubj** (gente) e uma **xcomp** (ficar). Espera-se, portanto, que a oração que está sendo reconstruída a partir da inserção do *token* vazio também apresente um **advmod**, um **nsubj** e uma **xcomp**. Contudo, observa-se que na segunda oração só há um sintagma nominal e uma oração, mas nenhum ADV para assumir a função de **advmod**. Presume-se, portanto, que o sintagma nominal seja **nsubj** e a oração seja **xcomp**¹¹. Na sentença, o *token* elíptico está entre colchetes e com destaque em azul. Na Figura 12a o *token* elíptico é marcado pela relação **orphan**; já na Figura 12b, o *token* elíptico é representado pelo *token* vazio NOFORM.

Aqui a gente fica preso e os bandidos, [ficam] soltos.

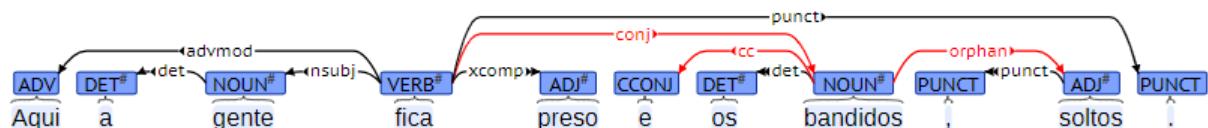


Figura 12a. Relação **orphan** anotada nas *basic dependencies*

Aqui a gente fica preso e os bandidos, NOFORM soltos.

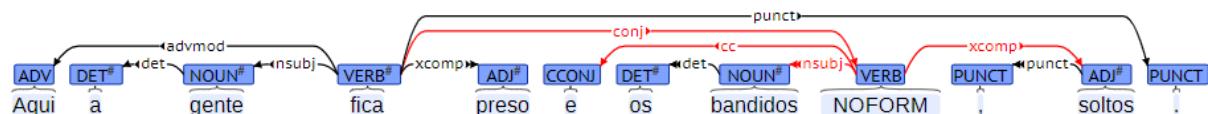


Figura 12b. Novas¹² relações *enhanced* estabelecidas após inserção de *token* vazio

O problema é que, algumas vezes, os dependentes do *token* vazio não são absolutamente o que se esperava a partir da hipótese de simetria entre as duas orações coordenadas. Os dependentes podem ser em menor ou em maior número e podem apresentar *POS tags* diferentes. Isso pode ser observado na sentença a seguir, ilustrada nas Figuras 13a e 13b.

¹¹ Como não há um ADV disponível, o ADV “aqui”, da primeira oração, passará a ser **advmod** também da segunda oração mas isso será visto nas próximas *enhanced dependencies*.

¹² Não foram incluídas as outras relações *enhanced* para não poluir o exemplo.

Na hora ela não pensou em nada, [na hora] [ela] só [pensou] em tirar as crianças.

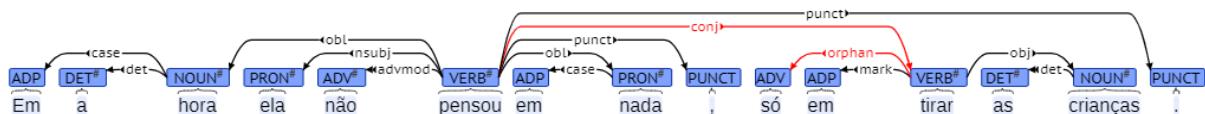


Figura 13a. Orações coordenadas assimétricas, com **orphan**, nas *basic dependencies*

Na hora ela não pensou em nada, só NOFORM em tirar as crianças.

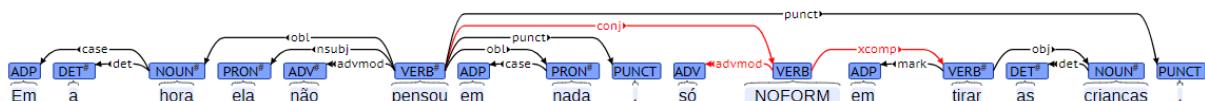


Figura 13b. Orações coordenadas assimétricas, com *token* vazio, nas *enhanced dep*.

A primeira oração da sentença ilustrada nas Figuras 13a e 13b apresenta dois **obl** (“hora” e “nada”), um **nsbj** [“ela”] e um **advmod** (“não”). A segunda oração¹³, contudo, só apresenta um **advmod** (“só”) e uma oração (“tirar”). A relação **advmod** está presente, mas onde estão os dois **obl** e o **nsbj**? Que tipo de relação deve ser atribuída à oração “tirar”, dependente do *token* vazio? Para humanos, é relativamente simples descobrir que o **obl** “hora” e o **nsbj** “ela” estão elípticos na segunda oração. É também relativamente simples deduzir que o **obl** argumental “nada” da primeira oração tem como equivalente uma **xcomp** (“tirar”) na segunda oração. No entanto, fazer isso automaticamente, por meio de regras, pode ser muito mais complexo.

Uma alternativa à construção de regras para automatizar a inserção de *tokens* vazios para representar predicados elípticos seria usar técnicas de aprendizado de máquina. Contudo, a elipse de predicado não é um fenômeno muito frequente em córpus de português e, por esse motivo, há poucos dados para o aprendizado automático.

Pelo que se observou, nenhum *parser* treinado em córpus do português anotado com UD atribui a relação **orphan**, e provavelmente isso se deva tanto à esparsidade de dados no treinamento quanto à diversidade de padrões das orações que requerem a anotação da relação. A relação **orphan** está relacionada a uma ausência (ausência do predicado) e talvez isso dificulte ainda mais o aprendizado.

¹³ Destaca-se o fato de ser uma coordenada assindética, ou seja, não introduzida por uma conjunção, e, portanto, sem a relação **cc**.

1.3 Questões envolvendo os diferentes “órfãos” de um predicado elíptico

Nos códigos de português disponíveis no site da UD, o *token* mais frequente como *head* da relação **orphan** é o sujeito da voz ativa (**nsubj**), seguido do objeto direto (**obj**) e do sujeito da voz passiva (**nsubj:pass**), conforme ilustra a Tabela 1.

Head de orphan	Bosque	Petrogold	<u>Porttinari</u>	PUD
nsubj	7	4	29	4
obj	0	2	9	0
nsubj:pass	0	3	2	0
obl	0	2	5	0
advmod	0	0	2	0
outros	0	0	7	0

Tabela 1. Distribuição do *head* de **orphan** por tipo de relação, nos códigos de português¹⁴

Reconhecer a função dos *tokens* “órfãos” em relação ao predicado elíptico é importante em dois momentos: primeiro no momento de decidir qual órfão vai ser *head* da relação **orphan**, já que há uma ordem de sucessão a ser obedecida, como visto na seção 1.1; segundo, no momento de decidir qual o nome das relações que unem o *token* vazio inserido a cada um dos (ex) participantes da relação **orphan**, como visto na Subseção 1.2.

A seguir são ilustrados e comentados casos com diferentes combinações de funções dos órfãos do predicado elíptico. Na sentença de cada exemplo, o *token* elíptico é mostrado entre colchetes, em azul. Na linha de sucessão mostrada junto a cada exemplo, o *head* de **orphan** está destacado em vermelho e o dependente de **orphan** em negrito. As Figuras que ilustram os gráficos da anotação de cada exemplo nas *basic* e nas *enhanced dependencies* trazem em vermelho as relações que mudam de uma para outra.

No exemplo a seguir, os órfãos são o **obj** e o **obl** do predicado elíptico. Nesse cenário, a linha de sucessão estabelece que o *head* de **orphan** seja o **obj**.

A prefeitura chamará 300 aprovados em março e [chamará] os demais até julho.

- órfãos do predicado elíptico: **obj** (“os demais”) e **obl** (“até março”)
- **nsubj > obj > iobj > obl > advmod > csubj > xcomp > ccomp > advcl**

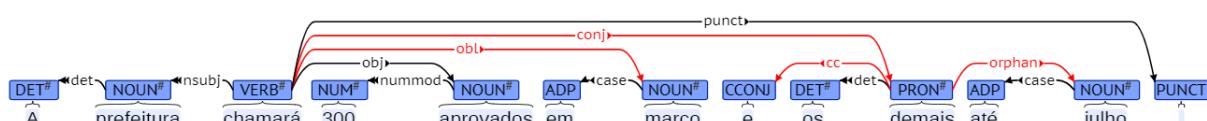


Figura 14a. Basic dependencies: head de **orphan** é o **obj** do predicado elíptico.

¹⁴ O código CINTIL não apresenta nenhuma relação **orphan** anotada.

A prefeitura chamará 300 aprovados em março e NOFORM os demais até julho.

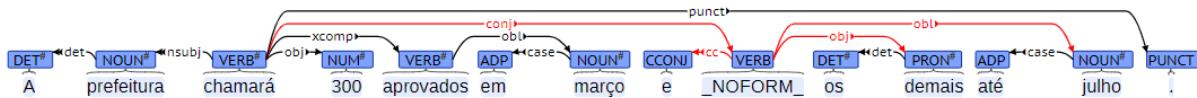


Figura 14b. Enhanced dependency com as funções dos órfãos **obj** e **obl** explicitadas¹⁵.

Observa-se, nas Figuras 14a e 14b, que a oração *head* de **conj** apresenta **nsubj**, **obj** e **obl**, mas a oração dependente de **conj** apresenta um único sintagma nominal não prepostionado (“os demais”) e um sintagma prepostionado (“até julho”). O sintagma prepostionado é facilmente identificado como **obl** do predicado elíptico, mas foi preciso decidir se o sintagma não prepostionado era **nsubj** (hipótese A) ou **obj** (hipótese B). Na hipótese A, haveria elipse do **obj** na oração dependente de **conj** e na hipótese B, haveria elipse do **nsubj** na oração dependente de **conj**, como mostrado a seguir:

A prefeitura chamará 300 aprovados em março e os demais até julho.

- A prefeitura chamará 300 aprovados em março e **os demais [chamarão]** [300 aprovados] até julho.
- A prefeitura chamará 300 aprovados em março e [a prefeitura] [chamará] **os demais** até julho.

Via de regra, nesses casos, decide-se que o sintagma não prepostionado é **obj** (opção B), pois sempre há a possibilidade de que o sujeito da primeira oração seja também o sujeito da segunda oração.

O próximo exemplo (Figuras 15a e 15b) ilustra um caso em que os órfãos são o **obl** e o **csubj** do predicado elíptico. Nesse caso, a linha de sucessão determina que o *head* de **orphan** seja o **obl**.

Cabe ao Legislativo criar leis e, [cabe] à sociedade1, cumpri-las.

- órfãos do predicado elíptico: **obl** (“ao Legislativo”) e **csubj** (“criar leis”)
- nsubj** > **obj** > **iobj** > **obl** > **advmmod** > **csubj** > **xcomp** > **ccomp** > **advcl**

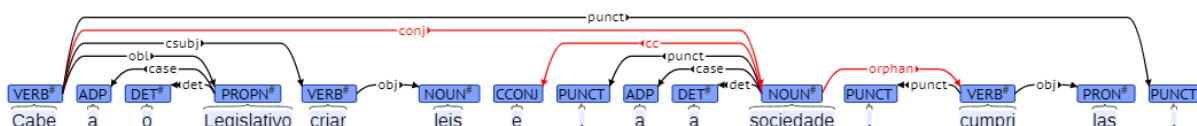


Figura 15a. Basic dependency: head de **orphan** é **obl** do predicado elíptico

¹⁵ Para fins clareza, omitiu-se na figura a *enhanced dependency* que propaga o sujeito “prefeitura” como sujeito do *token* inserido NOFORM.

Cabe ao Legislativo criar leis e, NOFORM à sociedade, cumpri-las.

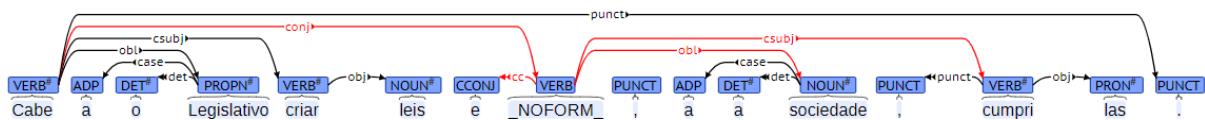


Figura 15b. Enhanced dependency com função dos órfãos **obl** e **csub** explicitadas

Os exemplos de diferentes órfãos de predicados elípticos mostram que as funções presentes na oração *head* de **conj** são quase sempre projetadas nos *tokens* órfãos do predicado elíptico na oração dependente de **conj**.

Há casos, no entanto, em que as funções presentes no *head* de **conj** não têm equivalentes com a mesma forma de expressão sintática na oração dependente de **conj**. É o caso da sentença ilustrada pelas Figuras 16a e 16b. O **advmmod** na primeira oração ("agora") tem por equivalente uma **obl** na segunda oração ("no ano"). Embora tenham *POS tags* diferentes, ambos os *tokens* têm papel semântico de Tempo.

Ele vai se formar agora e eu [formar] só no ano que vem.

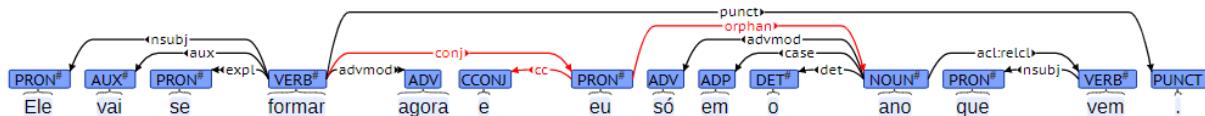


Figura 16a. Basic dependency: head de **conj** com **advmmod** e dependente de **conj** com **obl**

Ele vai se formar agora e eu NOFORM só no ano que vem.

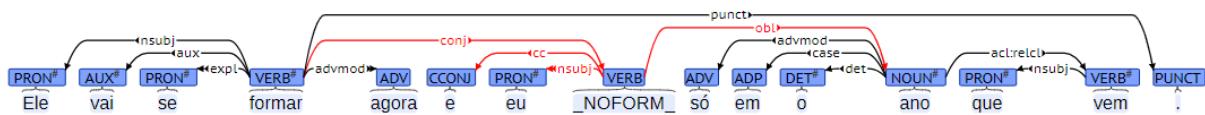


Figura 16b. Enhanced dependency com função do órfão **obl** explicitada

Além de **advmmod** e **obl**, outras relações de dependência podem ligar equivalentes semânticos não simétricos sintaticamente, como :

advmmod e advcl:

Vou jantar aqui **hoje** e [jantar] na sua casa quando você me convidar.

obl e advcl:

Eu convidei a família para o almoço e [convidei] vocês para jogar à noite.

obl e xcomp:

Ele se arrependeu da viagem e eu **[arrependi]** de não ter viajado.

1.4 Mais de um token elíptico

Muitas vezes, além do predicado, há outros elementos elípticos na oração dependente de **conj**. Os exemplos a seguir exemplificam essa situação: o predicado explícito, na primeira coordenada, foi destacado em vermelho; o predicado elíptico (não presente na segunda oração coordenada) foi preenchido entre colchetes, em azul e negrito, e os demais elementos elípticos na segunda oração foram preenchidos entre colchetes, em azul.

Você **sente falta** de afeto e ele **[sente] [falta]** de dinheiro.

Hoje **vou usar** a blusa azul e amanhã **[usar]** a **[blusa]** verde.

Ele **deu atenção** aos pais e eu **[dei] [atenção]** aos filhos.

Emma **ganhou o Oscar** de melhor atriz e Cillian **[ganhou] [o Oscar]** de melhor ator.

A Figura 17 ilustra um caso assim:

Tenho medo de fantasmas e ela **[tem] [medo]** de ser mordida por cobra.

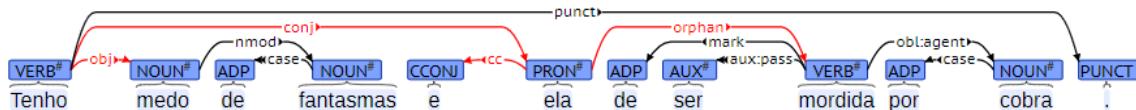


Figura 17a. Basic dependency: elipse do predicado e do **obj**

Na segunda oração da sentença ilustrada na Figura 17a, o predicado ("tem") e o **obj** ("medo") estão elípticos, ao passo que o **nsubj** ("ela") e a **acl** que modifica o **obj** ("de ser mordida por cobra") estão explícitos. Como a inserção de um *token* vazio só ocorre para representar o predicado elíptico, mantém-se a elipse do **obj** e adapta-se a anotação da **acl** que seria dependente do **obj** elíptico como uma relação plausível (**xcomp**, no caso, pois, assim como **obj**, complementa o sentido de verbos), conforme ilustrado na Figura 17b.

Tenho medo de fantasmas e ela NOFORM de ser mordida por cobra.

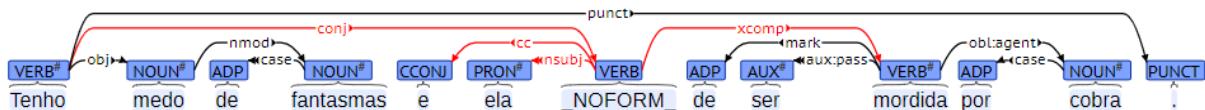


Figura 17b. Enhanced dependency: **acl** de "medo" anotada como **xcomp** do *token* vazio

É possível, ainda, que haja mais de um predicado elíptico numa mesma oração e isso exige a inserção de dois tokens vazios para representá-los, como ilustrado nas Figuras 18a e 18b.

João decidiu cursar Direito e seu irmão [decidiu] [cursar] Administração.

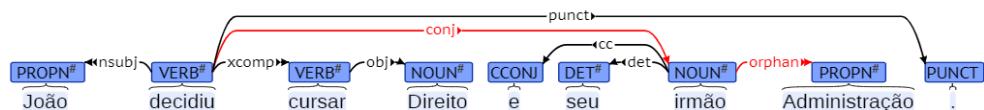


Figura 18a. *Basic dependency*: elipse de dois predicados

João decidiu cursar Direito e seu irmão NOFORM NOFORM Administração.

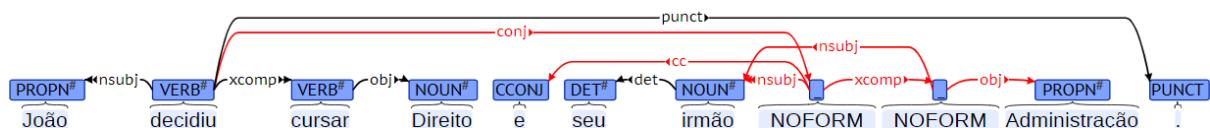


Figura 18b. *Enhanced dependency*: inserção de dois tokens vazios

Casos como esse aumentam a complexidade da tarefa de reconstruir a estrutura argumental dos predicados elípticos.

Atenção: Não são considerados órfãos do predicado elíptico os modificadores naturais dos próprios órfãos, como determinantes, preposições, conjunções e expletivos.

1.5 A questão da posição de inserção do predicado elíptico

Decidir a posição mais “natural” para inserir o *token* vazio é uma questão discutível, pois há mais de uma posição gramaticalmente aceitável. Na *shared task* promovida para desenvolver programas que anotassem automaticamente as *enhanced dependencies* (Bouma, Seddah & Zeman, 2020) essa questão foi, inclusive, ignorada pelos avaliadores da tarefa.

Em se tratando de anotação humana, acredita-se que o *token* vazio sempre será inserido numa posição adequada. Porém, se o objetivo for informar um programa sobre onde inserir o *token* vazio, é preciso colocar algumas restrições.

Uma das restrições é que o *token* vazio (NOFORM) nunca deve ser inserido entre tokens já ligados por outras relações que não a **orphan**. Isso significa que não se deve separar os *tokens* órfãos de seus próprios dependentes (já anotados, por exemplo, com **det**, **case**, **amod**, **nmod**, **nummod**, **mark**, **advmod**, **fixed**, **flat**, etc.). A seguir é apresentada uma sentença com o *token* vazio inserido de forma incorreta, seguida de uma sentença com a inserção correta:

Incorreto:

Em Porto Rico, 900 mil pessoas ficaram sem energia e cerca de 50 mil sem NOFORM abastecimento de água.

=> o *token* vazio não pode ser inserido entre “sem” e “abastecimento”, pois “sem” é dependente de “abastecimento” na relação **case**.

Correto:

Em Porto Rico, 900 mil pessoas ficaram sem energia e cerca de 50 mil NOFORM sem abastecimento de água.

=> o *token* vazio foi inserido entre os dois integrantes da relação **orphan** (“50’ e “abastecimento”), respeitados seus respectivos dependentes individuais.

A outra restrição é que, se um dos *tokens* participantes da relação **orphan** for um advérbio de negação (“não”, “nunca”), o *token* vazio deverá ser inserido imediatamente à sua direita:

Pensamos em poupar tempo, mas **não** NOFORM em **poupar** aborrecimentos.

João falta das aulas, mas **Pedro, nunca** NOFORM.

De modo geral, a inserção do *token* vazio entre os participantes da relação **orphan**, independentemente de qual seja o *head* da relação, costuma ser uma boa decisão, principalmente porque, em córpus do português, o principal herdeiro do predicado elíptico é o sujeito (ou seja, é muito comum haver sujeito em orações com elipse de predicado), o que garante a ordem SVO (Sujeito, Verbo, Objeto).

Contudo, quando o *head* da relação **orphan** não é o sujeito, a inserção do predicado elíptico entre os participantes de **orphan** pode gerar uma ordem não muito natural, mas nem por isso inaceitável. Isso acontece nas sentenças a seguir, em que o *head* da relação **orphan** era o objeto direto (“demais” na primeira e “vocês” na segunda sentença).

A prefeitura chamará 300 aprovados em março e os **demais** NOFORM até **julho**.

= A prefeitura chamará 300 aprovados em março e os **demais** chamará até **julho**.

= A prefeitura chamará 300 aprovados em março e **chamará** os **demais** até **julho**.

Eu convidei a família para o almoço e **vocês** NOFORM para **jogar** à noite.

= Eu convidei a família para o almoço e **vocês** convidei para **jogar** à noite.

= Eu convidei a família para o almoço e **convidei** **vocês** para **jogar** à noite.

1.6 Predicados elípticos com mais de dois órfãos

Há casos em que o predicado elíptico deixa mais de dois órfãos. Nesses casos, é estabelecida uma relação **orphan** para cada dependente de **orphan**, mas o *head* de **orphan** permanece o mesmo, como ilustra a Figura 19a.

Ele ganhou três prêmios neste ano e ela [ganhou] dois no ano passado.

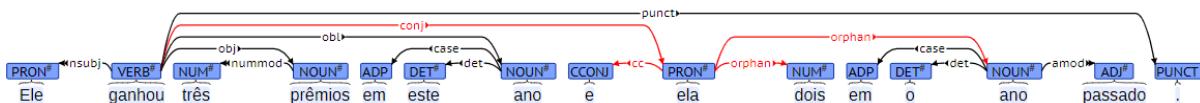


Figura 19a. Basic dependencies com 2 relações **orphan** num mesmo *head*

Nas *enhanced dependencies* (Figura 19b), os três participantes das duas relações **orphan** têm suas funções reveladas ao serem ligados ao *token* vazio: **nsubj** (“ela”), **obj** (“dois”) e **obl** (“ano”) do predicado elíptico.

Ele ganhou três prêmios neste ano e ela NOFORM dois no ano passado.

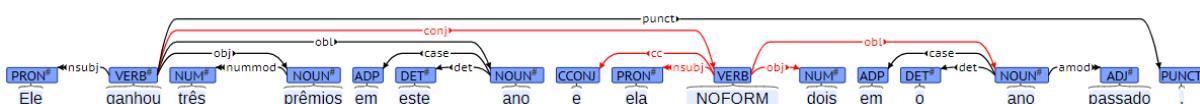


Figura 19b. Enhanced dependencies com 3 órfãos “adotados” pelo token vazio

A UD não comenta casos com mais de dois órfãos do predicado elíptico.

1.7 Predicados elípticos com um único órfão

Há casos em que, na oração que apresenta predicado elíptico, há apenas um órfão remanescente. Se a relação de **cc** que introduz esse órfão remanescente é do tipo aditiva (“e”), a relação de coordenação se estabelece no nível em que esse órfão se encontra, ignorando-se a possibilidade de haver um predicado elíptico:

Vimos **peixes** e **baleias**. => a **conj** liga “peixes” e “baleias”

Reclamei **do formato** e **freelamei** da **cor**. => a **conj** liga “formato” e “cor”

Porém, se a relação de **cc** é do tipo adversativa (“mas”) e a primeira oração está na negativa, é preciso reconhecer o predicado elíptico:

Não **vimos** peixes, mas **[vimos]** baleias. => a **conj** liga “vimos” e “baleias”

Não **reclamei** do formato, mas [reclamei] da cor. => a **conj** liga “reclamei” e “cor”

Nesses casos, contudo, não há relação **orphan** anotada, porque a relação **orphan** requer pelo menos dois órfãos do predicado elíptico (um para ser o *head* e outro para ser o dependente da relação). Sendo assim, não há a pista da relação **orphan** para motivar a inserção do *token* vazio. Uma nova pista poderia ser a coordenação entre *tokens* de diferentes *POS tags*, porém, dada a existência de predicados verbais e nominais coordenados, a precisão dessa pista seria baixa.

Também são exemplos desse fenômeno, no português, as orações adverbiais comparativas. As Figuras 20a e 20b ilustram um caso desses nas *basic* e nas *enhanced dependencies*.

Você correu mais hoje do que eu [corri].

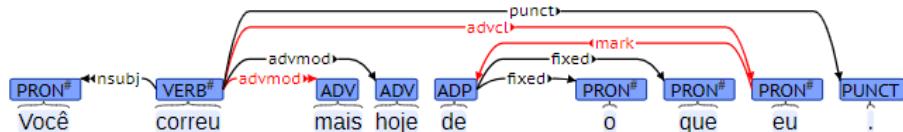


Figura 19a. oração **advcl** comparativa com predicado elíptico e um único órfão.

Você correu mais hoje do que eu NOFORM.

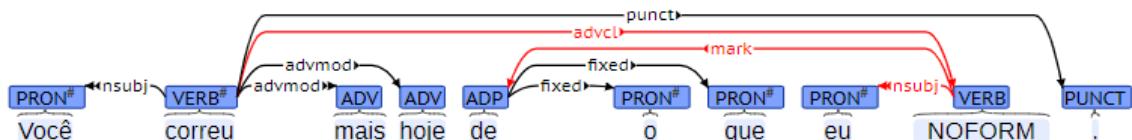


Figura 20b. Inserção do *token* vazio e explicitação do papel do órfão

A UD, contudo, não comenta casos de predicados elípticos com um único órfão nem predicados elípticos fora das orações coordenadas.

1.8 Conclusão

Em resumo, a inserção do *token* vazio nas *enhanced dependencies* da UD envolve os seguintes passos:

- identificar a relação **orphan** nas *basic dependencies*
- determinar a posição onde será inserido o *token* vazio
- estabelecer uma relação entre o *head* de **conj** e o *token* vazio
- estabelecer uma relação entre o *token* vazio e o dependente de **cc**¹⁶
- estabelecer uma relação entre *token* vazio e o (ex) *head* de **orphan**
- estabelecer uma relação entre o *token* vazio e o (ex) *dependente* de **orphan**

Dada a baixa frequência do fenômeno e diversidade de contextos em que ocorre, esta é uma *enhanced dependency* que merece ser anotada manualmente.

O conhecimento sobre o comportamento das elipses, contudo, abre possibilidades para explorá-las computacionalmente. Seria necessário, para isso, reunir um córpus de casos

¹⁶ Desde que a oração seja coordenada por meio de uma conjunção (oração coordenada sindética)

anotados que contivesse, ao mesmo tempo, diversidade de padrões (ou seja, diferentes combinações de funções de órfãos) e uma quantidade razoável de sentenças para cada um dos padrões.

2. Propagação de *heads* de *tokens* coordenados pela relação conj

Esse tipo de *enhanced dependency* trata da propagação da relação que “chega” no *head* de **conj**, ou seja, a relação da qual o *head* de **conj** é dependente. Essa *enhanced* se distingue daquela que será abordada na Seção 3, que tratará da propagação das relações que “partem” do *head* de **conj**, ou seja, relações cujo *head* é o token *head* de **conj**. Para não causar confusão entre essas duas *enhanced dependencies*, nesta e na próxima seção serão ilustradas exclusivamente a anotação das *enhanced* em foco (em vermelho nas figuras).

Os termos “chegar” e “partir” podem ser melhor entendidos ao se observar as ilustrações dos arcos das relações: se a flecha “chega” no *head* de **conj**, a *enhanced* é de propagação de *head*, assunto desta seção; se a flecha “parte” do *head* de **conj** em direção a outros *tokens*, a *enhanced* é de propagação de dependentes, assunto da Seção 3.

A *enhanced* de propagação de *head* estabelece que, se dois *tokens* estão ligados pela relação **conj**, ambos compartilham o mesmo *head*. A Figura 21 ilustra essa propriedade

Gosto de **macarrão** com **molho** e **queijo** ralado.

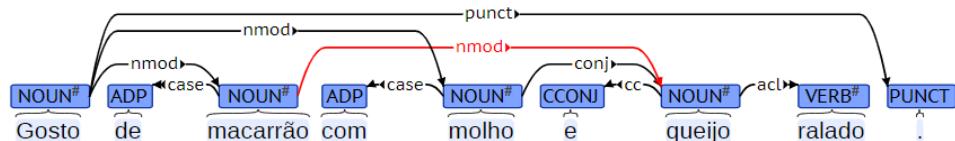


Figura 21. Propagação de head do head de **conj** para o dependente de **conj**

Na Figura 21, como “macarrão” é *head* de “molho” na relação **nmod** e “molho” é *head* de “queijo” na relação **conj**, então “macarrão” também é *head* de “queijo” numa relação **nmod**. Assim, nas *enhanced dependencies* cria-se uma relação do mesmo tipo, **nmod**, ligando o *head* “macarrão” a “queijo” (relação em vermelho).

O mesmo se aplica a outros *tokens* coordenados, dependentes de outros tipos de relações, como os ilustrados nas Figuras 22 e 23.

Ele **recebeu** **vaias** e **aplausos**.

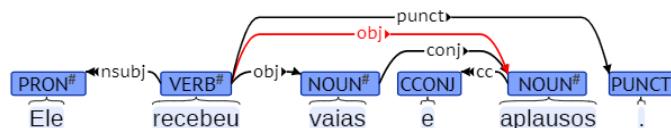


Figura 22. Propagação da relação **obj** para o dependente de **conj**

Na Figura 22, como “recebeu” é head de “vaias” numa relação **obj** e “vaias” é *head* de **conj** com “aplausos”, então “recebeu” também é *head* de “insetos” numa relação **obj** nas *enhanced dependencies*.

Ele **reclamou** do calor e dos insetos.

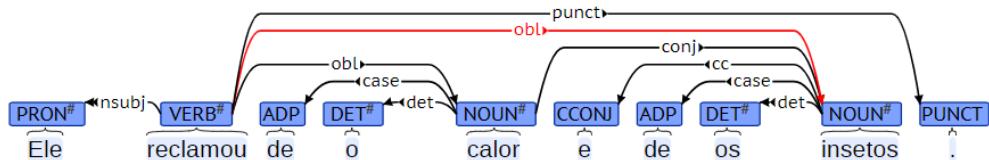


Figura 23. Propagação da relação **obl** para o dependente de **conj**

Na Figura 23, como “reclamou” é *head* de “calor” numa relação **obl** e “calor” é *head* de **conj** com “insetos”, então “reclamou” também é *head* de “insetos” numa relação **obl** nas *enhanced dependencies*.

2.1 Propagação de *head* de *tokens* coordenados que possuem POS tags distintas

Os exemplos apresentados até aqui são casos em que os dois *tokens* coordenados apresentam a mesma *POS tag*, porém há casos de *tokens* de diferentes *POS tags* ligados pela relação **conj**. Nesses casos, propaga-se o **head**, porém é preciso adequar o nome da relação de dependência, como mostra a Figura 24.

Ele **reclamou** do calor e de ter que andar muito.

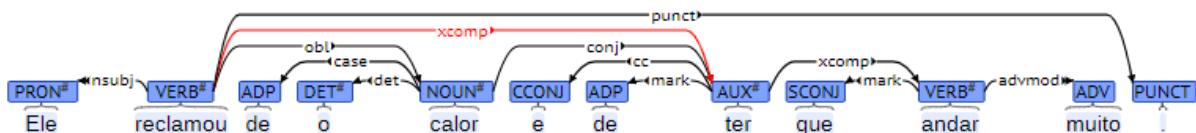


Figura 24. Propagação de *head* entre *tokens* NOUN e VERB coordenados

Na Figura 24, um NOUN e um VERB estão ligados pela relação **conj**. Como “reclamou” é *head* do NOUN, “reclamou” também tem que ser *head* do VERB, porém, em relação ao NOUN, “reclamou” é *head* de **obl** e em relação ao VERB, “reclamou” é *head* de **xcomp**.

Tome esse remédio **hoje** e sempre que **tiver** dores semelhantes

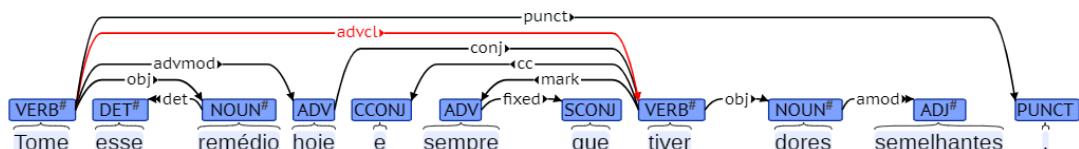


Figura 25. Propagação de *head* entre *tokens* ADV e VERB coordenados

Na Figura 25, um ADV e um VERB estão ligados pela relação **conj**. Como “tome” é *head* do ADV, “tome” também tem que ser *head* do VERB (“tiver”), porém, em relação ao ADV, “tome” é *head* de **advmmod**, e em relação ao VERB, “tome” é *head* de **advcl**.

Casos de *tokens* de diferentes *POS tags* coordenados devem receber atenção redobrada na revisão da anotação, pois propagam o *head* do primeiro coordenado para os demais coordenados, mas não propagam o mesmo tipo de relação de dependência.

2.2 Propagação de *head* de *tokens* de palavras funcionais coordenados

Há casos em que os *tokens* coordenados não são palavras plenas, mas sim palavras funcionais, como preposições e conjunções, como pode ser visto na Figura 26.

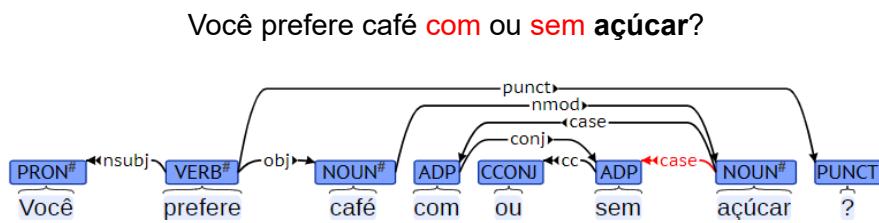


Figura 26. Propagação do *head* de **conj** entre duas preposições (ADP)

Na Figura 26, como “açúcar” é *head* de “com” numa relação **case**, então “açúcar” também é *head* de “sem” numa relação **case** nas *enhanced dependencies*. Como consequência da propagação do **case**, a anotação passará a ter não apenas o **nmod** “com açúcar”, mas também o **nmod** “sem açúcar” para modificar “café”¹⁷.

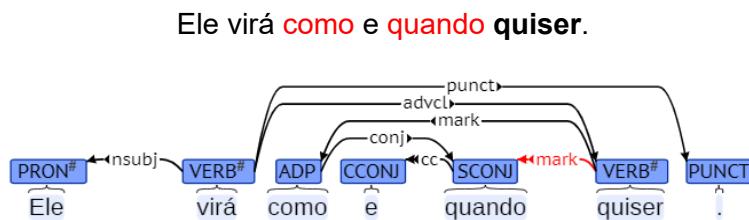


Figura 27. Propagação do *head* de **conj** entre duas conjunções (SCONJ)

Na Figura 27, como “querer” é *head* de “como” numa relação **mark**, então “querer” também é *head* de “quando” nas *enhanced dependencies*. Sendo assim, além da **advcl** “como querer”,

¹⁷ Isso ficará claro na extensão dos dois **nmod**, incorporando a forma da preposição que introduz cada um deles (**nmod:com** e **nmod:sem**). Esse tipo de extensão constitui o tipo de *enhanced dependencies* discutido na Seção 6 deste relatório.

a anotação passa a apresentar também a **advcl** “quando quiser” para modificar “virá”: “como quiser” e “quando quiser”¹⁸.

2.3 Head que não propaga do **head** para o dependente de **conj**

A relação **root** é a única da qual o **head** de **conj** pode ser dependente mas que não propaga para o dependente de **conj**. Isso porque a relação **root** não é uma relação sintática, mas sim um recurso computacional para indicar a raiz da árvore sintática. Não há, portanto, nenhum sentido em manter mais de um **root** em uma sentença.

¹⁸ Novamente, isso ficará claro na extensão das duas **advcl**, incorporando a forma das conjunções subordinativas que introduzem cada uma delas (**advcl:como** e **advcl:quando**). Esse tipo de extensão constitui o tipo de *enhanced* discutido na Seção 6 deste relatório.

3. Propagação de dependentes do head de conj

A *enhanced dependency* de propagação de dependentes de **conj** estabelece que, se dois *tokens* estão ligados pela relação **conj**, alguns dependentes do *head* de **conj** são também dependentes do dependente de **conj**. A Figura 28 ilustra essa propriedade para o dependente do tipo **nsubj**.

Essa **música** é **alegre** e **animada**.

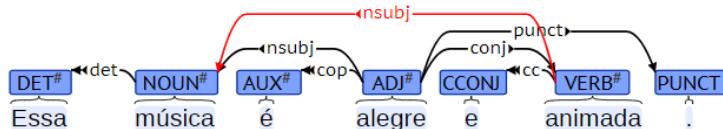


Figura 28. Propagação da relação **nsubj** para o dependente de **conj**

Na Figura 28, como “música” é dependente **nsubj** de “alegre” e “alegre” é *head* na relação **conj** com “animada”, então “música” é também **nsubj** de “animada” (relação *enhanced* em vermelho).

Um *token* pode ser *head* de várias relações, mas nem todas podem ser propagadas automaticamente do *head* de **conj** para o dependente de **conj**. Para explicar o que pode ser propagado, o que não deve ser propagado e o que pode ser propagado dentro de determinadas condições, o restante desta seção foi dividido em cinco subseções: 3.1 - propagação de sujeito (**nsubj**, **nsubj:pass**, **csubj**, **csubj:pass**); 3.2 - propagação de **case** e **mark**; 3.3 - relações que são propagadas dependendo da posição do dependente (**obj**, **iobj**, **obl**, **nmod**, **amod**, **advmod**, **ccomp**, **xcomp**, **advcl**, **acl**, **acl:relcl**); 3.4 - relações que não são propagadas, algumas por decisão de projeto (**aux**, **cop**, **det**, **expl**) e outras por não fazer sentido (ex: **punct**, **discourse**, **reparandum**, **dislocated**, etc.). Por fim, em 3.4.5, são feitas considerações gerais acerca da *enhanced* aqui em foco.

3.1 Propagação do sujeito do head para o dependente de conj

A condição básica para propagação é que o *head* de **conj** apresente um sujeito explícito (**nsubj**, **nsubj:pass**, **csubj**, **csubj:pass**) e o dependente de **conj** não apresente sujeito explícito, como nas Figuras 29, 30, 31 e 32.

João viajou na sexta e não voltou até agora.

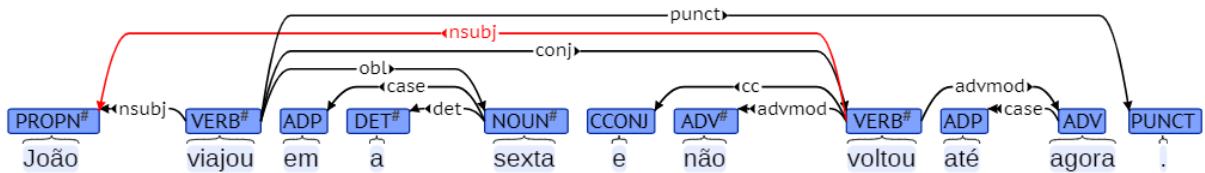


Figura 29. Propagação de **nsubj** do *head* para o dependente de **conj**

Os fazendeiros foram surpreendidos e prejudicados pelas enchentes.

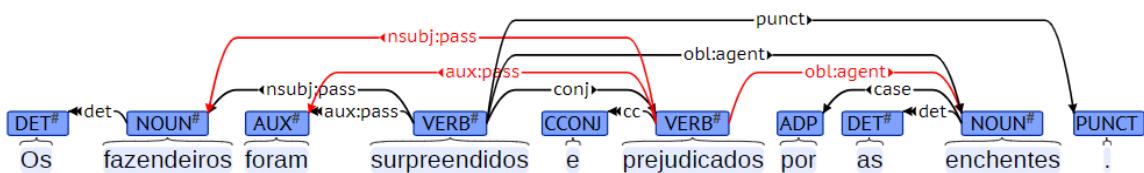


Figura 30. Propagação de **nsubj:pass** do *head* para o dependente de **conj**

É difícil e improvável reverter o desmatamento.

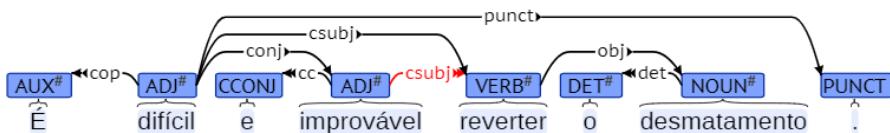


Figura 31. Propagação de **csubj** do *head* para o dependente de **conj**

Foi permitido e até recomendado usar shorts no trabalho.

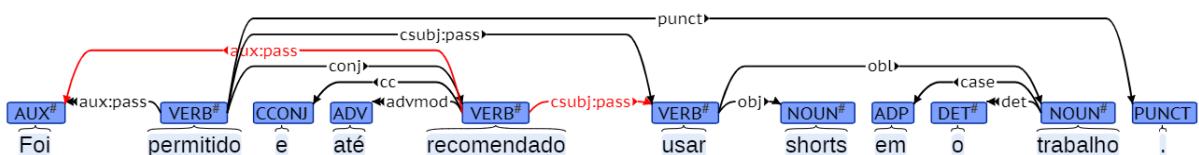


Figura 32: Propagação de **csubj:pass** do *head* para o dependente de **conj**

3.1.1 Não propagação de sujeito do *head* para o dependente de **conj**

Se o *head* de **conj** não tiver sujeito, não haverá sujeito a ser propagado, conforme ilustra a Figura 33.

Acordei cedo e logo **saí** para caminhar.

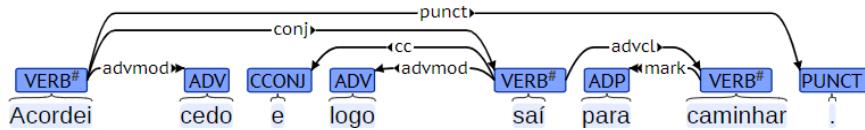


Figura 33. Não propagação de sujeito: *head* de **conj** não possui sujeito

Da mesma forma, se o dependente de **conj** já tiver seu próprio sujeito, não poderá receber a propagação, como mostra a Figura 34.

A comida era gratuita e a bebida era barata.

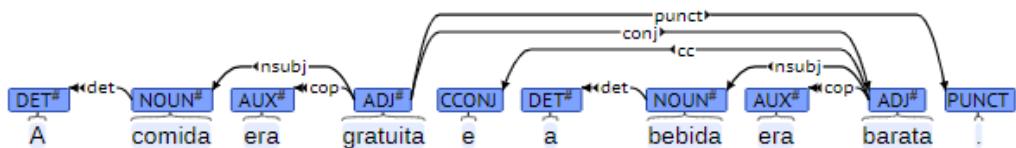


Figura 34. Não propagação de sujeito: dependente de **conj** já possui sujeito

3.1.2 Inversão de voz entre *tokens* coordenados e adequação do sujeito

Quando a relação **conj** liga duas orações, pode haver inversão de voz ativa para passiva ou vice-versa. Esse tipo de situação exige que eventuais sujeitos sofram adequação (**nsubj** para voz ativa e **nsubj:pass** para voz passiva), conforme ilustra a Figura 35.

Ele foi **avisado**, mas não **tomou** precauções.

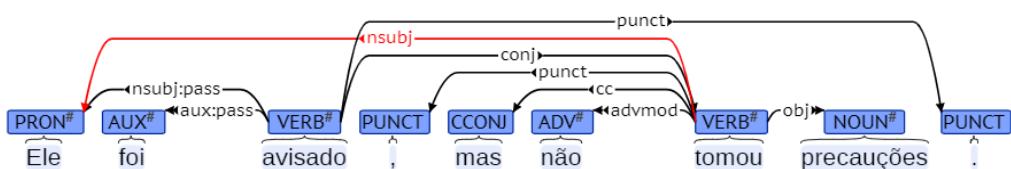


Figura 35. Propagação do sujeito de **conj** com inversão de voz

Na Figura 35, a primeira oração é uma passiva ("foi avisado") e por isso o sujeito de "avisado" é **nsubj:pass**. A segunda oração coordenada, contudo, é uma construção de voz ativa e por isso o sujeito "Ele" foi anotado como **nsubj** de "tomou", portanto, sofrendo adequação de voz.

3.1.3 Propagação de sujeito em cascata

É possível “aproveitar” o sujeito atribuído numa *enhanced dependency* para fazer outra *enhanced dependency*. Isso ocorre, por exemplo, quando duas orações coordenadas, sem

sujeito, ocorrem depois de uma oração **xcomp** que teve seu sujeito mapeado em uma *enhanced dependency*.

Ele pretende **estudar** engenharia e **trabalhar** na construção de navios.

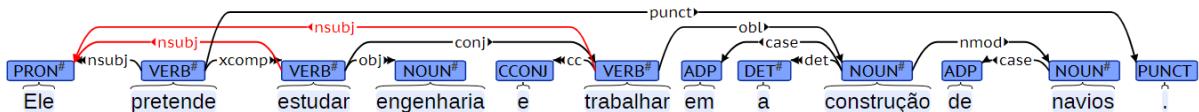


Figura 36. Propagação de sujeito em cascata, de **xcomp** para *head* e dependente de **conj**

Na Figura 36, o *head* e o dependente de **conj** não possuem sujeito nas *basic dependencies* e, a princípio, não haveria o que ser propagado. No entanto, como o *head* de **conj** - “estudar”, é ao mesmo tempo dependente de **xcomp**, herda o sujeito “Ele” por meio da *enhanced dependency* de sujeitos de **xcomp**. Esse sujeito *enhanced*, por sua vez, passa a ser o sujeito propagável para o dependente de **conj** - “trabalhar”.

3.2 Propagação de **case** e **mark**

As relações **case** e **mark** são importantíssimas para a atribuição da semântica dos modificadores adverbiais (locativo, temporal, causal, etc.), além de terem outras utilidades para a pesquisa linguística. Por isso, sua propagação enriquece muito o córpus. A regra é simples: desde que o dependente de **conj** não tenha seu próprio **case** ou **mark**, o **case** ou **mark** do *head* de **conj** é propagado, conforme pode ser observado nas Figuras 37, 38, 39 e 40:

Ele se arrependeu de voltar e confessar a culpa.

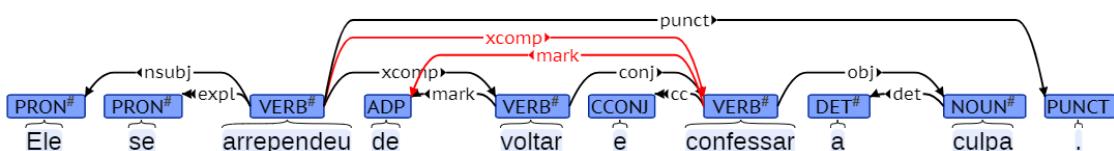


Figura 37. Propagação de **mark** do *head* para o dependente de **conj**

Só viemos **porque** fomos **convidados** e **incentivados** a comparecer.

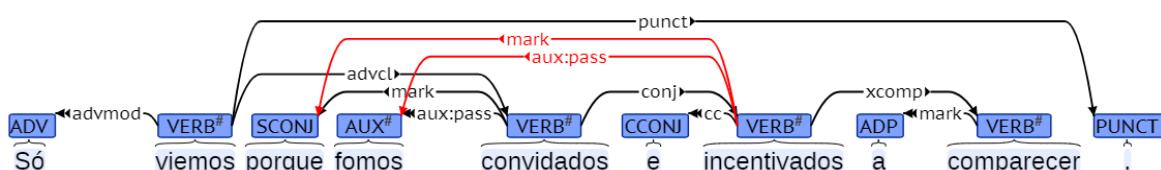


Figura 38. Propagação de **mark** do *head* para o dependente de **conj**

O governador disse que está ciente e vai tomar providências.

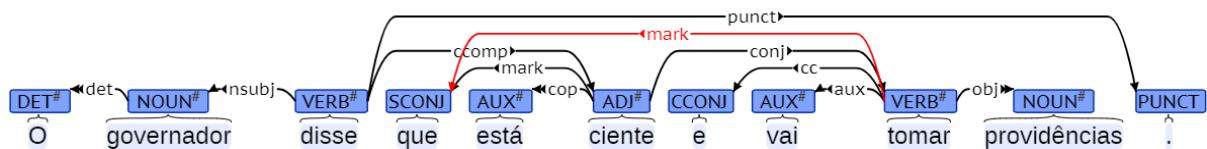


Figura 39. Propagação de **mark** do *head* para o dependente de **conj**

Ele a viu em carne e osso só uma vez.

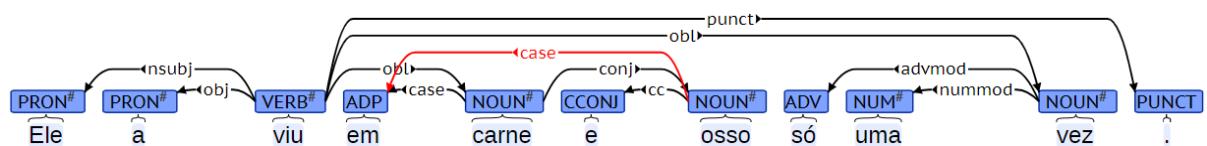


Figura 40. Propagação de **case** do *head* para o dependente de **conj**

3.3 Relações que não são propagadas

Entre as relações não propagadas, há aquelas que não faz sentido propagar (3.2.1) e aquelas cuja propagação faz sentido na maior parte das vezes, mas a forma do *token* nem sempre possui a flexão adequada (3.2.2).

3.3.1 Relações que não faz sentido propagar

O *head* de **conj** pode ser *head* de diversas relações, algumas das quais nem constituem relações sintáticas, como **punct**, **goeswith** e **reparandum**. Como regra geral, se não estiver explícito que um tipo de relação deva ser propagado nas *enhanced dependencies*, ele simplesmente não deverá ser propagado.

Uma relação modificadora não propagada, por exemplo, é **nummod**, porque o quantificador do *head* de **conj** geralmente não modifica o dependente de **conj**. Exemplos:

Compramos 8 pães e queijo suficiente para recheá-los.

Há 20 cadeiras e algumas mesas disponíveis.

3.3.2 Relações que faz sentido propagar, mas os *tokens* nem sempre se adequam

As relações **det**, **aux**, **cop** e **expl** não são propagadas porque nem sempre têm a mesma forma para os dois *tokens* ligados por **conj**, pois são formas flexionadas. Embora em alguns casos a propagação dessas relações possa ser feita, por segurança optou-se por não

fazê-la automaticamente¹⁹, mesmo porque são relações que não têm informação tão semanticamente relevante quanto as demais que são propagadas automaticamente.

Os exemplos a seguir e a Figura 41 mostram casos em que a propagação de **det**, **aux**, **cop** e **expl** não seria adequada. Todos esses casos estão relacionados à elipse de *tokens* (indicados em azul), mas só o caso da Figura 41 constitui elipse de predicado e, portanto, é tratada com a inserção de um *token* vazio (NOFORM).

Ele **estava** **vindo** e nós [estávamos] **chegando**.

=> o **aux** “estava” não propaga de “vindo” para “chegando” porque tem forma e *features* diferentes daquelas exigidas pelo **aux** elíptico “estávamos”.

Ele **estava feliz** e nós [estávamos], **tristes**.

=> o **cop** “estava” não propaga de “feliz” para “triste”, pois suas *features* não se adequam para representar o **cop** elíptico “estávamos”. Observe-se que aqui não há inserção de *token* vazio, pois os dois predicados nominais estão explícitos e o que está elíptico é apenas o **cop** da segunda oração.

Vou levar meus filhos e [suas] respectivas esposas.

=> o det “meus” não propaga de “filhos” para “esposas”, pois “esposas”, além de ser feminino, ao contrário de “filhos” não são “minhas”, ou seja, não é possessivo da primeira pessoa do singular.

Na Figura 41, a seguir, o **expl** “se” não é propagado de “tornou” para o *token* vazio NOFORM inserido para representar o predicado elíptico “tornei”, pois o “se” teria que ser flexionado para “me” e, como a inserção de *token* elíptico na UD só contempla predicados, não há possibilidade de inserir o *token* “me”, partícula integrante do verbo pronominal na primeira pessoa do singular.

Ele **se tornou** professor e eu, [me] [tornei] pesquisador.

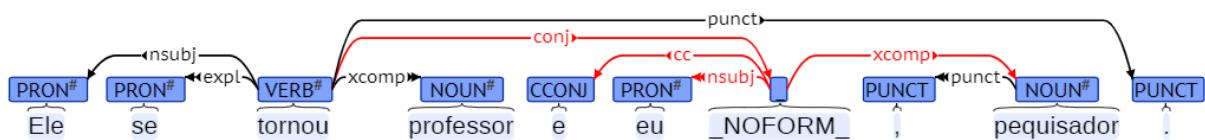


Figura 41. Não propagação de **expl** para o *token* vazio NOFORM dependente de **conj**

¹⁹ Em aplicações que visem explicitar todos os tokens elípticos inferíveis a partir do contexto, porém, isso poderia ser feito. É uma possibilidade a ser explorada, por exemplo, por sistemas de simplificação textual, visando diminuir um número de inferências exigidas do leitor.

3.4 Propagação condicional de dependentes de conj

As relações **obj**, **iobj**, **obl**, **nmod**, **amod**, **advmod**, **xcomp**, **ccomp**, **advcl**, **acl**, **acl:relcl** podem ou não ser propagadas, dependendo de sua posição em relação ao *head* e ao dependente de **conj**.

Na verdade, utiliza-se a posição do dependente de cada uma dessas relações mantidas pelo *head* de **conj** para inferir se o escopo dessas relações comprehende ou não o dependente de **conj**.

Há três configurações em que pode ocorrer o dependente de uma dessas relações cujo *head* é também *head* de **conj**. A seguir, em vermelho, destaca-se esse dependente nas três configurações:

- A. **dependente** - HEAD de **conj** - DEPENDENTE de **conj**
- B. HEAD de **conj** - **dependente** - DEPENDENTE de **conj**
- C. HEAD de **conj** - DEPENDENTE de **conj** - **dependente**

3.4.1 Configuração A: dependente anteposto ao *head* de **conj**

Na configuração A, o dependente da relação ocorre antes do *head* de **conj** e seu escopo pode ou não abranger o dependente de **conj**. As relações que mais ocorrem nesta configuração são **advmod** e **advcl**.

Ele **raramente fuma**, mas **bebe** demais. => não propaga

Ele **raramente fuma e bebe** em casa. => propaga

Ele **raramente fuma e bebe** só nos finais de semana. => não propaga

Quando **viajamos**, **relaxamos** e nos **divertimos**. => propaga

Por **morarmos** aqui, **sofremos** com o calor e **enfrentamos** enchentes. => propaga

Enquanto a propagação de **advcl** na configuração A parece ser segura, a propagação de **advmod** exige muitas especificações. O **advmod** é uma das relações cujo escopo é mais complexo de ser inferido.

O **advmod** anteposto ao *head* de **conj** não deve ser propagado se a polaridade de duas orações coordenadas for contrária, o que ocorre em duas situações:

- quando a conjunção (**cc**) é adversativa (“mas”, “porém”, “contudo”, etc.). Ex:

Eu **dificilmente ouço**, mas **leio** relatos de golpes. => não propaga

- quando o advérbio é de negação (“não”, “nem”, “nunca”). Ex:

Ele **nem chegou** e já **foi** embora. => não propaga

Eu **não entendi** e **confesso** isso. => não propaga

Ela **nunca estuda** e **passa**. => não propaga

Mas há exceções para os os advérbios negativos, como:

Eu **nunca almoço** e **deito** em seguida. => propaga

Além disso, se a conjunção é alternativa (“ou”), o escopo do advérbio negativo parece alcançar o dependente de **conj**:

Eu **não ouvi** ou **li**. => propaga

Nas coordenações aditivas cujo head de **conj** é modificado por outros advérbios, que não os de negação, o escopo do advérbio parece alcançar o dependente de **conj**, desde que o dependente de **conj** não tenha seu próprio **advmod**:

As pessoas **sempre votam** nele e se **arrependeram**. => propaga

As pessoas **sempre votam** e raramente se **arrependeram**. => não propaga

As crianças **quase se afogaram** e **morreram**. => propaga

3.4.2 Configuração B: dependente entre o *head* e o dependente de **conj**

Na configuração B, o dependente da relação ocorre imediatamente após o *head* de **conj** e antes do dependente de **conj**, e essa é uma posição que torna o dependente exclusivo do *head* de **conj**, motivo pelo qual não é propagado para o dependente de **conj**. Exemplos:

Vamos sempre para a praia e **nadamos** no mar. => não propaga **advmod**

Relaxamos quando **viajamos** e nos **divertimos** muito. => não propaga **advcl**

Temos **alergia a pó** e **intolerância à lactose**. => não propaga **nmod**

Tentamos conseguir e **falhamos**. => não propaga **xcomp**

Disse que **viria** e **faltou**. => não propaga **ccomp**

O centro acolhe **mulheres jovens** e seus **bebês**. => não propaga **amod**

Pretendo **estudar** engenharia e **trabalhar** no exterior. =>não propaga **obj**

3.4.3 Configuração C: dependente após o dependente de **conj**

Na configuração C, o dependente da relação ocorre após o dependente de **conj** e, portanto, seu arco “abraça” os dois *tokens* coordenados. Qualquer dúvida quanto a qual seria o *head* desse dependente posposto já foi decidida no momento da anotação das *basic dependencies*, conforme ilustram as Figuras 42, 43, 44, 45 e 46.

Relaxamos e nos divertimos muito quando viajamos.

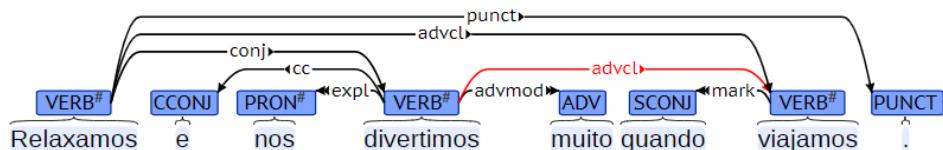


Figura 42. Propagação de **advcl** para dependente de **conj**

O advogado **defende** e o promotor **acusá** os **réus**.

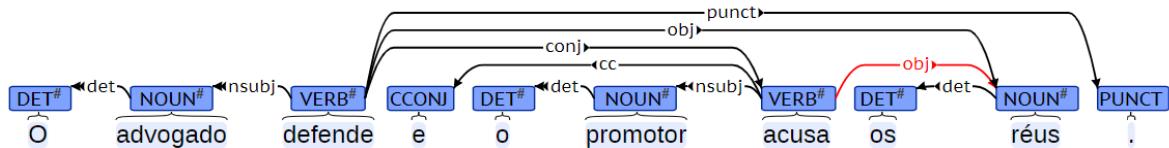


Figura 43. Propagação de **obj** para dependente de **conj**

Ele me **ensinou** e **incentivou** a **pescar**.

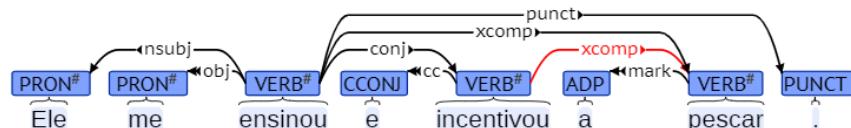


Figura 44. Propagação de **xcomp** para dependente de **conj**

A medida visa amparar **homens e mulheres afro-descendentes**.

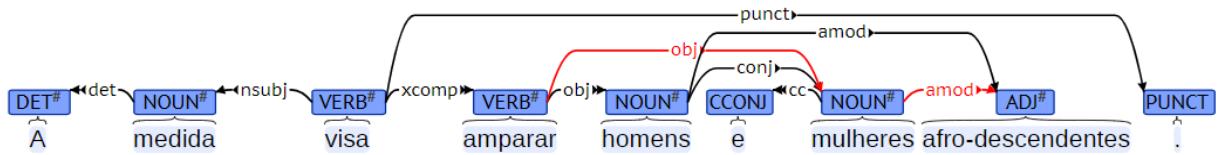


Figura 45. Propagação de **amod** para dependente de **conj**

Arecio os **ritmos** e as **danças** do **Brasil**.

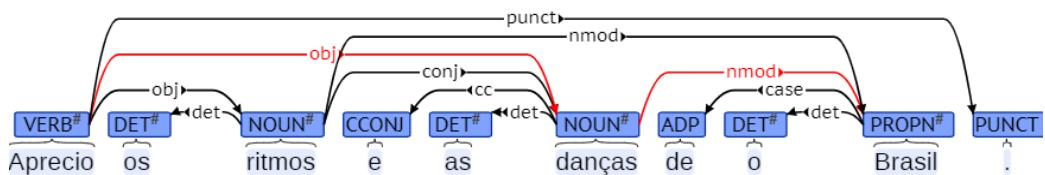


Figura 46. Propagação de **nmod** para o dependente de **conj**

3.5 Conclusões sobre a propagação de dependentes do **head** de conj

Pelo que foi exposto nesta seção, a propagação de dependentes do *head* para o dependente de **conj** é seletiva, pois apenas algumas relações podem ser propagadas.

Mesmo nas relações que podem ser propagadas, deve-se observar se o dependente de **conj** já não tem um dependente próprio do mesmo tipo (um sujeito, um objeto, etc.), caso em que a propagação não deverá ser feita.

Além disso, a propagação condicional de algumas relações (dependendo da configuração A, B ou C) indica que a propagação dos **advmod** antepostos ao *head* de **conj** devem ser revisados com maior atenção, pois são muitas as variações observadas e pode ser que nem todas elas correspondam a regras do programa de anotação de *enhanced dependencies*.

4. Atribuição do sujeito de xcomp

A relação **xcomp** tem como característica o fato de que seu dependente não admite um sujeito explícito. Isso é mais do que ter o sujeito elíptico, pois um sujeito elíptico poderia ser eventualmente preenchido, o que não é o caso da oração dependente de **xcomp**. Essa ausência de sujeito é chamada *null subject* ou “sujeito nulo”. Apesar de o dependente de **xcomp** não ter um sujeito sintático, é possível inferir qual é o sujeito, pois o sujeito nulo é “controlado” por um *token* presente na oração *head* de **xcomp**. Nas *enhanced dependencies*, esse *token* “controlador” do sujeito nulo é identificado e anotado como sujeito do dependente de **xcomp**. Para simplificar, no restante desta seção esse sujeito anotado nas *enhanced dependencies* será referenciado como “sujeito do dependente de **xcomp**”. A Figura 47 ilustra, em vermelho, um caso de sujeito atribuído ao dependente de **xcomp**.

Eles conseguiram perceber a diferença.

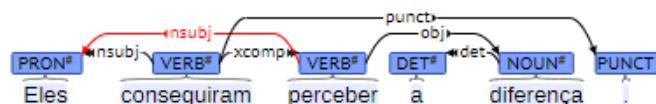


Figura 47. Sujeito do dependente de xcomp = sujeito do head de xcomp

O sujeito do dependente de **xcomp** é um dos diferentes *tokens* da estrutura argumental do *head* de **xcomp**: o sujeito (**nsubj**, **nsubj:pass**) (Figura 47) ou o objeto direto (**obj**) (Figura 48) ou, em casos mais raros, o objeto indireto, seja ele expresso por um pronome (**iobj**) (Figura 49) ou por um sintagma nominal (**obl**) (Figura 50). Nas figuras, a relação *enhanced* de sujeito do dependente de **xcomp** está destacada em vermelho.

Ele viu a situação mudar.

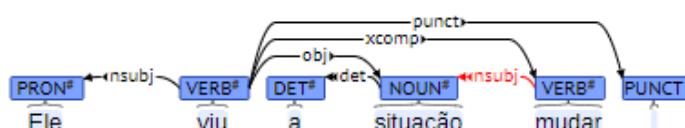


Figura 48. Sujeito do dependente de xcomp = obj do head de xcomp

Eles não lhe permitiram falar.

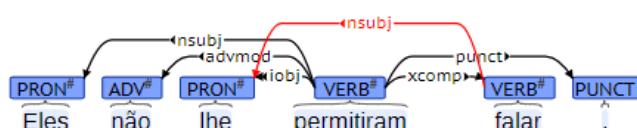


Figura 49. Sujeito do dependente de **xcomp** = **iobj** do *head* de **xcomp**

A lei faculta aos **desabrigados** **pedir** indenização.

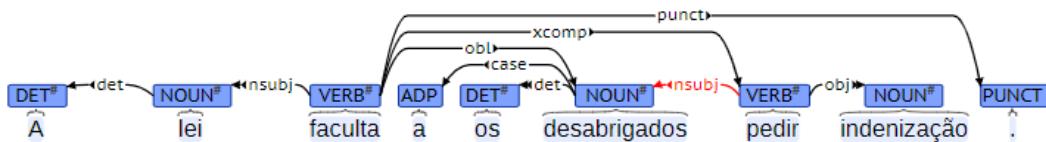


Figura 50. Sujeito do dependente de **xcomp** = **obl** do *head* de **xcomp**

Esses casos apresentam relação direta com o léxico dos verbos, como será visto nas próximas seções.

4.1 Sujeito do dependente de **xcomp** controlado pelo sujeito do *head* de **xcomp**

Os verbos modais (ex: poder, dever), aspectuais (ex: começar, acabar, continuar), volitivos (ex: querer, pretender) e outros com complemento oracional no infinitivo são clássicos exemplos em que o sujeito do *head* de **xcomp** pode ser identificado como sujeito do dependente de **xcomp**. A restrição é para os casos em que o *head* de **xcomp** também é *head* de um objeto direto (**obj**) ou de um objeto indireto (**iobj**, **obl**), como será visto nas seções 4.2 e 4.3, respectivamente.

Nos exemplos a seguir, é possível perceber que “Ele” é sujeito sintático da primeira oração e sujeito controlador (inferível) da segunda oração, “fazer”.

Ele deve **fazer**.

Ele começou a **fazer**.

Ele pretende **fazer**.

Ele sabe **fazer**.

Ele ousou **fazer**.

Ele decidiu **fazer**.

Ele desistiu de **fazer**.

Se o *head* da **xcomp** está na voz ativa e o dependente da **xcomp** está na voz passiva, é necessário adequar o nome da relação de dependência, de **nsubj** para **nsubj:pass**, a fim de refletir isso, como mostra a Figura 51.

Ele cansou de ser acusado injustamente.

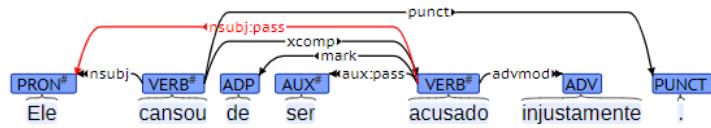


Figura 51. Sujeito do *head* de **xcomp** na voz ativa e sujeito do dependente na passiva

O inverso também é verdadeiro, ou seja, se o *head* de **xcomp** está na voz passiva e o dependente na voz ativa, altera-se a relação de **nsbj:pass** para **nsbj**, como ilustra a Figura 52.

Ele foi obrigado a confessar.

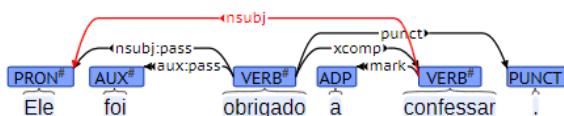


Figura 52. Sujeito do *head* de **xcomp** na voz passiva e sujeito do dependente na ativa

Mesmo quando o dependente de **xcomp** é um predicado nominal, a *enhanced dependency* opera normalmente, como pode ser observado nas Figuras 53, 54 e 55.

Ele decidiu ser advogado.

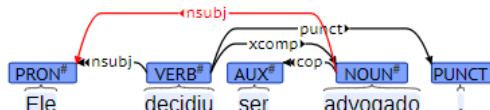


Figura 53. Dependente de **xcomp** predicado nominal (NOUN)

Ele pretende estar por perto.

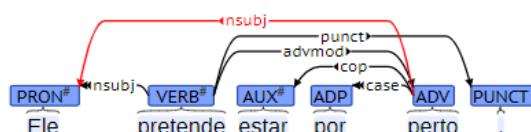


Figura 54. Dependente de **xcomp** predicado nominal (ADV)

Ele afirma estar disponível.

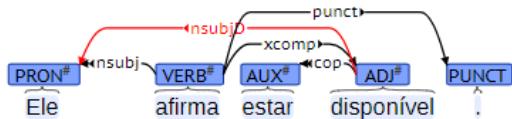


Figura 55. Dependente de **xcomp** predicado nominal (ADJ)

Se, no entanto, o *head* de **xcomp** não tiver sujeito explícito²⁰, o dependente de **xcomp** consequentemente não terá um sujeito atribuído por meio de *enhanced dependency*, como nas Figuras 56, 57 e 58.

Comecei a trabalhar cedinho.

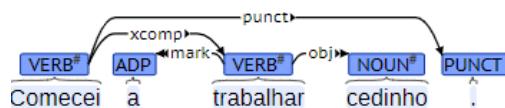


Figura 56. Relação de **xcomp** com sujeito do *head* elíptico.

Pediram para serem avisados.

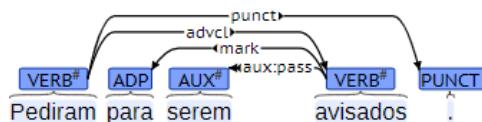


Figura 57. Relação de **xcomp** com sujeito do *head* elíptico

Pretende-se chegar até amanhã.

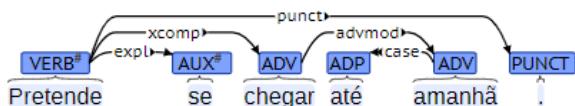


Figura 58. Relação de **xcomp** com sujeito do *head* indeterminado

Embora nas sentenças das Figuras 56 e 57 o sujeito esteja apenas oculto, pois está marcado na pessoa do verbo, não existe um *token* que possa receber a relação de sujeito

²⁰ Como os *tokens* são as unidades mínimas para anotação na UD, o fato de a pessoa estar marcada na desinência verbal não é relevante nesse caso, pois não há um *token* para receber a anotação do sujeito.

nas *basic dependencies* e, por consequência, não há um *token* que possa receber a relação de sujeito nas *enhanced dependencies* também.

4.2 Sujeito do dependente de **xcomp** controlado pelo **obj** do *head* de **xcomp**

Sempre que a oração *head* de **xcomp** também for *head* de um **obj**, esse **obj** será identificado como sujeito do dependente de **xcomp**. Há algumas categorias de verbos que apresentam esse tipo de estrutura, as quais serão comentadas e exemplificadas nas próximas subseções.

4.2.1 Verbos causativos, sensitivos e resultativos

Há menções, nas gramáticas e sites que discutem a língua portuguesa, a “sujeitos acusativos”, ou seja, um *token* que desempenha simultaneamente função de objeto direto da oração matriz e de sujeito da oração subordinada. (v. Bechara, E., 2019, p. 556; Mateus, M. H. M. et al., 2003, p. 642)

Em construções desse tipo, enquadram-se os verbos causativos (fazer, deixar, mandar), sensitivos (ver, ouvir, sentir) e o verbo usado em português para construções resultativas (ter). Optou-se, no projeto Porttinari, por anotar o *token* com dupla função como **obj**, pois é a função do nó mais próximo da raiz da árvore sintática e anotar o complemento oracional do verbo como **xcomp**, já antevendo que as *enhanced dependencies* permitiriam anotar as duas funções do *token* simultaneamente (**obj** do *head* de **xcomp** e sujeito do dependente da **xcomp**).

Verbos causativos

Nos verbos causativos: “fazer”, “deixar” e “mandar”, o **obj** pode se apresentar em forma de substantivo, próprio ou comum, pronome oblíquo, pronome indefinido (“alguém”, “quem”) e, agramaticalmente, mas cada vez com mais frequência, em forma de pronome do caso reto (eu, ele, você). As diferentes paráfrases a seguir ilustram esse tipo de construção (construções fora da norma culta assinaladas com asterisco):

Ele fez **Maria** chorar.

Ele fez a **mãe** chorar.

Ele a fez chorar.

*Ele fez **ela** chorar.

Ele fez as **crianças** chorarem.

Ele fez as **crianças** chorar.

Ele fez **chorar** as **crianças**.

As **crianças**, ele fez **chorar**.

Ele **as** fez **chorar**.

*Ele **as** fez **chorarem**.

*Ele fez **elas** **chorarem**.

Com esses verbos, a flexão do infinitivo (presente na oração dependente de **xcomp**) é opcional para **obj** em forma de substantivo e vetada para **obj** em forma de pronome acusativo. Decidimos, no projeto Porttinari, não adotar anotação diferente quando o infinitivo está flexionado. Casos desses verbos causativos, com a respectiva *enhanced dependency* de sujeito de dependente de **xcomp**, são ilustrados nas Figuras 59, 60 e 61.

Ele nos deixou pegar qualquer coisa.

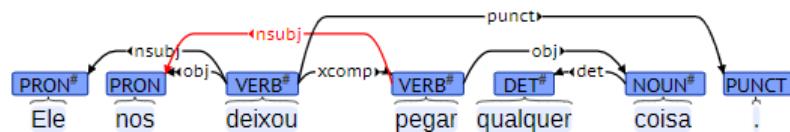


Figura 59. O **obj** do *head* de **xcomp** atribuído como sujeito do dependente de **xcomp**

Ele me fez reagir.

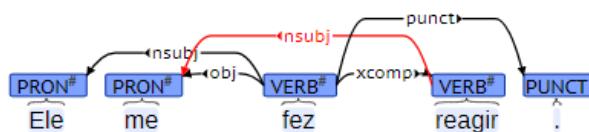


Figura 60. O **obj** do *head* de **xcomp** atribuído como sujeito do dependente de **xcomp**

Ele mandou você fazer isso.

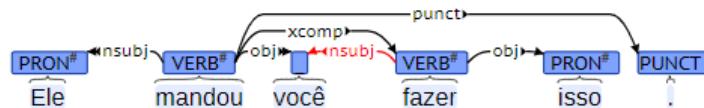


Figura 61. O **obj** do *head* de **xcomp** atribuído como sujeito do dependente de **xcomp**

Quando o sujeito do dependente da **xcomp** é um pronome acusativo (**obj** do *head* da **xcomp**), é possível, a partir da pessoa e número do pronome, descobrir o pronome nominativo correspondente: me (eu), te (tu), o (ele), a (ela), nos (nós), vos (vós), os (eles), as (elas). Assim, se para alguma aplicação for interessante separar o objeto e o sujeito amalgamados num só pronome, isso pode ser feito: “Ele me mandou [eu] fazer isso.”

4.2.1 Verbos sensitivos

Verbos sensitivos (ou de percepção), como “ver”, “ouvir”, e “sentir” também apresentam construções em que um *token* tem dupla função. Isso pode ser observado na Figura 62 e nos exemplos apresentados na sequência.

Ele ouviu o pássaro cantar.

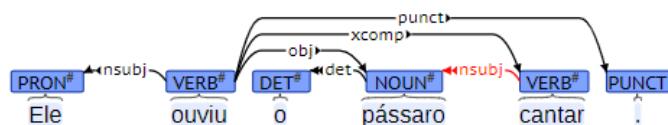


Figura 62. O **obj** do *head* de **xcomp** atribuído como sujeito do dependente de **xcomp**

Nos exemplos a seguir, o dependente de **xcomp** está em negrito e o **obj** do *head* de **xcomp**, controlador do sujeito do dependente de **xcomp**, em vermelho:

Ele viu a **chuva cair**.

Ele sentiu a **terra tremer**.

Verbo “ter” em construções resultativas

Por fim, *tokens* com dupla função são encontrados nas construções resultativas²¹, as quais utilizam o verbo “ter”, acompanhado de **obj** e de **xcomp** no particípio. Nessas construções, o sujeito do dependente da **xcomp** é sempre de voz passiva (**nsubj:pass**), como ilustra a Figura 63 e os exemplos na sequência.

Ele teve a carteira roubada.

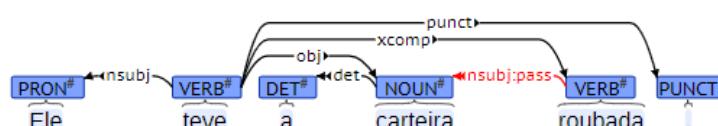


Figura 63. O objeto da construção resultativa atribuído como sujeito da passiva da **xcomp**

²¹ Cf: “Ele teve a carteira roubada” X “Ele teve roubada a carteira”. Em ambos os casos, o sentido é “A carteira dele foi roubada”.

O soldado teve a **perna amputada**.

Ele teve o **visto aprovado**.

O **visto**, ele teve **aprovado** só agora.

Ele teve, **aprovados**, o **visto** dele e do filho.

Note-se que, nas construções resultativas, o dependente de **xcomp** concorda com o objeto do *head* de **xcomp**, pois esse **obj** também é marcadamente seu sujeito (**nsubj:pass**).

4.2.2 Verbos que preveem predicativo do objeto

Outra categoria que tem o **obj** do *head* de **xcomp** como sujeito do dependente de **xcomp** é a dos verbos que requerem predicativo do objeto (também chamados de verbos de dois lugares): “considerar”, “achar”, “eleger”, “julgar”, “tornar”, “manter”, etc. Nesses verbos, o dependente de **xcomp** é sempre um predicado nominal. Essa estrutura pode ser observada na Figura 64 e nos exemplos na sequência.

Ele considerou o salário justo.

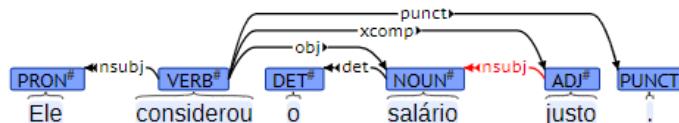


Figura 64. O **obj** do *head* de **xcomp** atribuído como sujeito do dependente de **xcomp**

Nos exemplos a seguir, o dependente de **xcomp** está em negrito e o **obj** do *head* de **xcomp**, controlador do sujeito do dependente de **xcomp**, em vermelho:

Ele achou a **prova difícil**.

Nós **o elegemos** **prefeito**.

O juiz julgou os **réus culpados**.

A vida **os** tornou **descentes**.

Mantenham as **mãos ocupadas**.

Pode-se observar, nesses exemplos, que o dependente de **xcomp** apresenta concordância com o **obj** do *head* de **xcomp**, pois é, simultaneamente, seu sujeito.

4.2.3 Verbos Transitivos Diretos e Indiretos com oração objetiva indireta.

Também apresentam um **obj**, que controla o sujeito da **xcomp**, os verbos transitivos diretos e indiretos (VTDI) em que a **xcomp** faz o papel de objeto indireto e é introduzida por preposição. Há diversos verbos nesta categoria: “motivar (a)”, “incentivar (a)”, “aconselhar (a)”, “autorizar (a)”, “ajudar (a)”, “impedir (de)”, “privar (de)”, “proibir (de)”, etc. Essa estrutura pode ser observada na Figura 65 e nos exemplos na sequência.

Ele proibiu os funcionários de trabalharem sem uniforme.

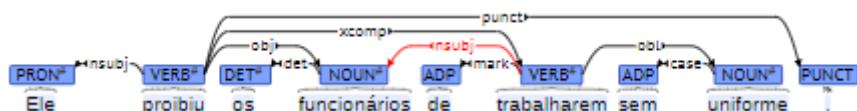


Figura 65. O **obj** do *head* de **xcomp** atribuído como **nsubj** do dependente de **xcomp**

Nos exemplos a seguir, o dependente de **xcomp** está em negrito e o **obj** do *head* de **xcomp**, controlador do sujeito do dependente de **xcomp**, em vermelho:

Ele motivou os **alunos** a **visitarem** o museu.

Ele incentivou os **filhos** a **estudarem**.

Ele **me** aconselhou a **ter** calma.

Ele autorizou os **fãs** a **entrarem** no camarim.

Ele ajudou as **vítimas** a **saírem** do local.

Ele impedi-nos de **entrar**.

Ele privou-os de **conhecer** a mãe.

4.2.4 Verbos transitivos que admitem objetos elípticos

Quando o **obj** do *head* da **xcomp** está elíptico, não haverá atribuição de sujeito para o dependente da **xcomp**.

A UD propõe anotar a relação que sofre elipse de **obj** como **ccomp**, a fim de que não haja erro na *enhanced dependency* de **xcomp**, ou seja, para que, na ausência de **obj**, o sujeito do *head* de **xcomp** não seja atribuído equivocadamente ao dependente de **xcomp**. As sentenças a seguir destacam em azul o lugar de elipse do **obj**.

Ele deixa **[obj elíptico]** fazer qualquer coisa.

Ele mandou [obj elíptico] fazer um terno para o casamento.

A prefeitura não permite [obj elíptico] estacionar aqui.

Eles aconselham [obj elíptico] esperar dois dias.

Eu não vi (**obj** elíptico) fazerem isso.

Contudo, numa língua *pro drop* (que admite elipse do sujeito) como o português, talvez essa solução não seja a ideal e o melhor seja tratar o problema no nível do léxico, marcando os verbos das categorias que preveem um **obj** que possa sofrer elipse. Essa situação ocorre em português com mais frequência com os verbos causativos (“fazer”, “mandar”, “deixar”), eventualmente com verbos sensitivos (“sentir”, “ver”, “ouvir”), mas potencialmente com qualquer outro verbo bitransitivo que tenha um dos complementos em forma oracional (como “permitir” e “aconselhar” nos exemplos anteriores).

No projeto Porttinari-base, decidiu-se manter a anotação de **xcomp** nesses casos e, por isso, ao anotar as *enhanced dependencies*, é preciso atenção para não atribuir o sujeito erroneamente em casos de **obj** elíptico. Em outras palavras, é preciso observar se o verbo pertence às categorias que preveem um **obj** que é simultaneamente sujeito do dependente da **xcomp**. Se pertencer, o sujeito do *head* da **xcomp** nunca pode ser atribuído ao dependente da **xcomp**, mesmo que o **obj** previsto esteja elíptico.

4.3 Sujeito do dependente de xcomp controlado por iobj ou obl do *head* de xcomp.

Embora mais raros, há casos de Verbos Transitivos Diretos e Indiretos (VTDI) em que a **xcomp** desempenha a função de objeto direto e um pronome (**iobj**) ou sintagma nominal preposto (**obl**) desempenha a função de objeto indireto. Nesses casos, o sujeito do dependente de **xcomp** é o **iobj** ou o **obl** com função de objeto indireto. Esse tipo de estrutura é ilustrado nas Figuras 66, 67 e 68.

A lei faculta aos desabrigados pedir indenização.

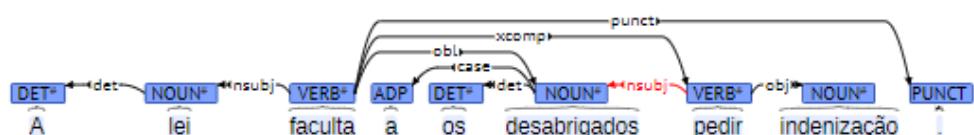


Figura 66. O **obl** do *head* de **xcomp** atribuído como **nsubj** do dependente de **xcomp**

Eles não nos permitiram entrar.

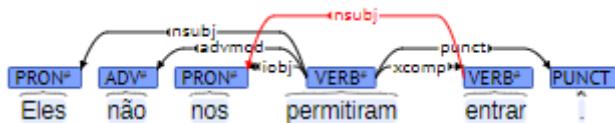


Figura 67. O **iobj** do *head* de **xcomp** atribuído como **nsubj** do dependente de **xcomp**

Isso possibilitou-**Ihe** receber restituição.

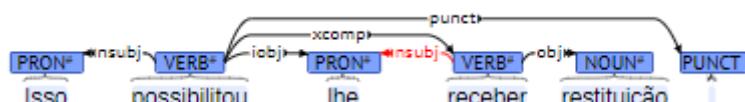


Figura 68. O **iobj** do *head* de **xcomp** atribuído como **nsubj** do dependente de **xcomp**

4.4 Objetos diretos e indiretos antepostos ao *head* de **xcomp**

Como vimos, se o *head* de **xcomp** é também *head* de **obj**, esse **obj** é sujeito do dependente de **xcomp**. Contudo, se o **obj** for um sintagma nominal e estiver à esquerda do *head* de **xcomp**, é bem provável que não seja um **obj** próprio do *head* de **xcomp** e, nesse caso, não poderá ser atribuído como sujeito do dependente de **xcomp**. Um **obj** é impróprio quando foi anotado indevidamente para não acarretar cruzamento de arcos. Por exemplo, na sentença a seguir, “plantão” é anotado como **obj** de “queria” (para evitar cruzamento de arcos), mas é, na verdade, **obj** de “pegar”.

O plantão do domingo ninguém **queria pegar**.

Nesse caso, a regra de anotar o **obj** do *head* da **xcomp** como sujeito do dependente da **xcomp** não deve valer. Em outras palavras, o **obj** impróprio não deve ser sujeito do dependente de **xcomp**.

A Figura 69a mostra “plantão” anotado nas *basic dependencies* como **obj** de “queria”.

O plantão do domingo ninguém queria pegar.

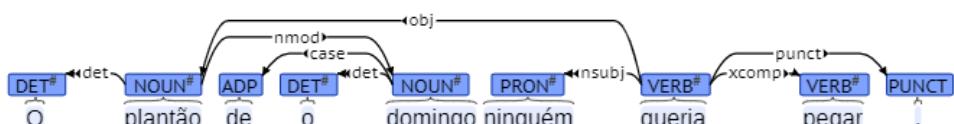


Figura 69a. **obj** não pronominal à esquerda do *head* de **xcomp** anotado nas *basic dep*.

Já a Figura 69b mostra “plantão” anotado como **obj** de “pegar”, com cruzamento de arcos. Nessa configuração, o *head* da **xcomp** deixaria de ter um **obj** e o sujeito do dependente **xcomp** passaria a ser “ninguém”.

O plantão do domingo ninguém queria pegar.

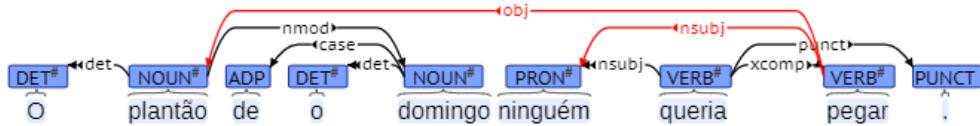


Figura 69b. **obj** não pronominal à esquerda do *head* de **xcomp** com o *head* correto

Mas é possível que o **obj**, mesmo à esquerda, seja um **obj** próprio do *head* da **xcomp**. Nesse caso, a regra deve valer, ou seja, o **obj** do *head* é sujeito do dependente da **xcomp**. A Figura 70 ilustra um caso desse tipo.

Uma brisa sentimos soprar.

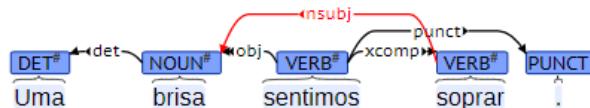


Figura 70. **obj** do *head* de **xcomp** anteposto.

Dante da incerteza quanto ao **obj** ser próprio ou impróprio, é importante conhecer as categorias de verbos que são *head* de **xcomp** e que apresentam **obj** ou **obl** ou **iobj** como sujeito do dependente de **xcomp**²²

De forma similar, há **obl** argumental²³ que pode ocorrer à esquerda do *head* de **xcomp** e é, na verdade, um **obl** do dependente de **xcomp** e não seu sujeito. A correção desse tipo de situação abre oportunidade para outro tipo de *enhanced*: a realocação, para o *token* correto, do *head* de componentes sintáticos antepostos, como mostrado na Figura 71, com o **obl** “nesses ativos” ligado ao verbo “investir”, dependente de **xcomp**.

²² Basicamente: verbos causativos (fazer, deixar, mandar); verbo resultativo (ter); verbos sensitivos (sentir, ver, ouvir) e verbos VTDI que apresentam o objeto direto oracional e o objeto indireto sintagmático (ex: permitir a alguém fazer alguma coisa).

²³ Na UD apenas pronomes são anotados como **iobj** e os objetos indiretos sob forma de sintagma nominal são anotados como **obl**. Como **obl** é uma relação também usada para anotar os adjuntos adverbiais sob forma de sintagma nominal, os **obl** que se referem a objetos indiretos são chamados de **obl** argumentais (mas não anotados diferentemente no Projeto Porttinari)

Nesses ativos pretendo investir muito pouco.

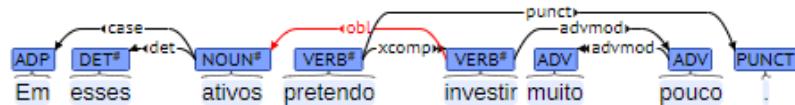


Figura 71. **obl** argumental à esquerda do *head* de **xcomp** com cruzamento de arcos

Ressalta-se que o problema da anteposição só afeta os casos de objetos diretos ou indiretos sob forma de sintagma nominal (**obj** e **obl**). Casos de objetos preenchidos com pronomes (**obj** pronominal e **iobj**), até onde se observou em córpus, não geram ambiguidade.

4.5 Orações infinitivas que não são xcomp

Há verbos que frequentemente apresentam sujeito posposto, como “acontecer”, “existir”, “faltar”, etc. Esses verbos, contudo, quando constituem complemento de outros verbos, admitem alternâncias em que o sujeito é anteposto ao sujeito da oração matriz e anotado como **obj** da oração matriz, para evitar cruzamento de arcos, conforme observado nas Figuras 72, 73 e 74.

Isso ninguém previa acontecer.



Figura 72. Caso de “falsa” **xcomp** com verbo “acontecer”

Essa probabilidade eles diziam existir desde o início.

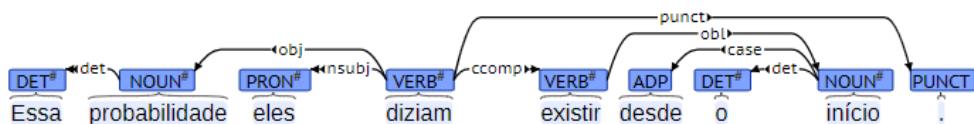


Figura 73. Caso de “falsa” **xcomp** com verbo “existir”

Tais requisitos eles acreditam faltar no programa.

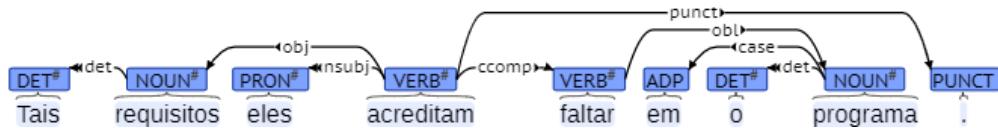


Figura 74. Caso de “falso” xcomp com verbo “faltar”

Nesses casos, como o sujeito é explícito (não nulo), a relação entre a oração matriz e a oração subordinada não pode ser **xcomp** e, por consequência, deve ser **ccomp**. Nas *enhanced dependencies*, o *head* desses *tokens* antepostos deve ser deslocado do primeiro verbo e alocado no segundo verbo, com a relação de sujeito, conforme ilustra a Figura 75.

Isso ninguém previa acontecer.

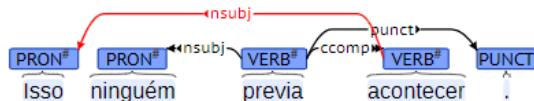


Figura 75. Anotação *enhanced* de “falsa” xcomp com verbo “acontecer”

Caso semelhante ocorre com os verbos impessoais que possuem **obj**, como “haver” e “ter”. Quando constituem locução verbal, o **obj** desses verbos pode ocorrer anteposto ao *head* de **xcomp**, como sujeito (Figuras 76 e 77).

Muitas nuances de cor pode haver no arco-íris.

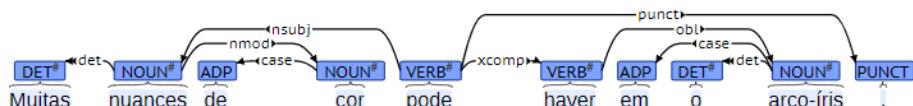


Figura 76. Anotação de xcomp com verbo “haver” como dependente

Essas variedades de planta não costumamos ter no Brasil.

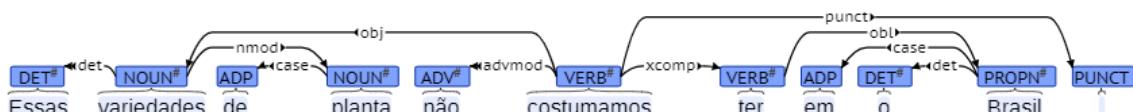


Figura 77. Anotação de xcomp com verbo “ter”, existencial, como dependente

Nesses casos, como o sujeito de “haver” e “ter” é inexistente, não há *enhanced dependency* de atribuição de sujeito ao dependente de **xcomp**. Nas *enhanced dependencies*, eventual

elemento anteposto ao *head* de **xcomp** deve ter seu *head* realocado para o dependente de **xcomp** com a função de **obj**, como ilustrado nas Figuras 78 e 79.

Muitas nuances de cor pode haver no arco-íris.

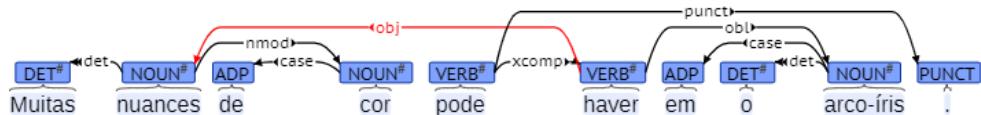


Figura 78. Anotação *enhanced* de **xcomp** com verbo “haver” como dependente

Essas variedades de planta não costumamos ter no Brasil.

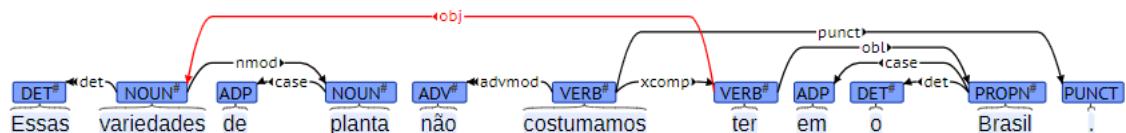


Figura 79. Anotação *enhanced* de **xcomp** com verbo “ter”, existencial, como dependente

O tipo de operação ilustrado nas Figuras 75, 78 e 79, de realocar o *head* de um *token*, mudando ou não a respectiva relação, será desenvolvido na Seção 7.5 (Realocação de elementos antepostos).

4.6 Conclusões sobre a atribuição de sujeito de **xcomp**

A *enhanced dependency* de atribuição de sujeito de **xcomp** tem uma regra geral muito clara que cobre a maioria dos casos: se o *head* da **xcomp** tem **obj**, o sujeito do dependente da **xcomp** é o **obj**, caso contrário, o sujeito do dependente da **xcomp** é o sujeito do *head* da **xcomp**.

O detalhamento de casos fornecido nesta seção teve por objetivo compartilhar diferentes cenários encontrados em córpus, mesmo que pouco frequentes, a fim de alargar a compreensão da atribuição da *enhanced* em foco.

5. Correferência nas acl:relcl

A *enhanced dependency* que trata o fenômeno da correferência nas orações relativas (**acl:relcl**) tem por objetivo ligar, por meio da relação **ref**, o pronome relativo (que, o qual, etc.) a seu antecedente, e ao mesmo tempo transferir, para o antecedente, a relação que chegava no pronome relativo.

A direção da relação **ref** é sempre da esquerda para a direita, ou seja, o antecedente do pronome relativo é o *head* de **ref** e o pronome relativo é o dependente de **ref**. A exceção é o pronome relativo “cujo” e suas flexões, tratada no item 5.4.

Uma vez que essa *enhanced dependency* altera, na coluna 9, algumas das relações anotadas nas colunas das *basic dependencies*, nesta seção os gráficos das *basic dependencies* e das *enhanced dependencies* serão apresentados em figuras separadas, com destaque, em vermelho, para as relações não compartilhadas entre as duas anotações (ou seja, aquelas alteradas nas *enhanced dependencies*).

A Figura 80a ilustra as *basic dependencies* de uma sentença contendo uma **acl:relcl**. O *head* da **acl:relcl** é o token “árvore” e o dependente é “vi” (verbo “ver”); o pronome “que”, por sua vez, se liga a “vi” como dependente da relação **obj**.

A árvore que eu vi estava sem folhas.

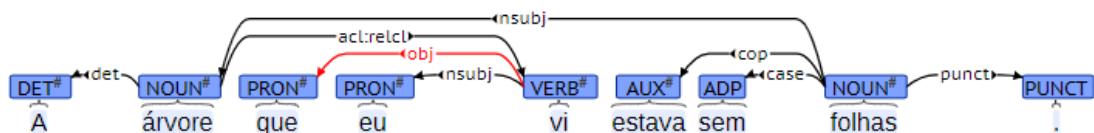


Figura 90a. Anotação de pronome relativo com função de **obj** nas *basic dependencies*.

A árvore que eu vi estava sem folhas.

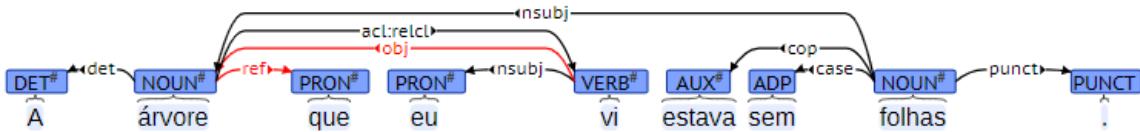


Figura 80b. Anotação da relação **ref** e realocação do dependente de **obj** nas *enhanced dependencies*.

Na Figura 80b, a sentença é a mesma da Figura anterior, porém o gráfico ilustrado mostra as *enhanced dependencies*. O token “árvore” liga-se ao pronome relativo “que” pela relação **ref**. Ao mesmo tempo, a relação **obj**, da qual o pronome “que” era dependente, transfere-se para “árvore”, o *head* de **ref**. Com essa operação, o pronome relativo é “isolado” e o token

“árvore” passa a ser simultaneamente **nsubj** de “folhas” (o que já era nas *basic dependencies*) e **obj** de “vi”.

Com essa *enhanced*, se uma aplicação baseada em sintaxe procura resposta para a pergunta “Viu o quê?” dirigida à sentença ilustrada na Figura 80b, não terá mais um simples “que” como resposta, mas sim uma palavra de conteúdo: “árvore”.

As Figuras 81a e 81b, a seguir, ilustram a anotação nas *basic* e nas *enhanced dependencies*, respectivamente, de uma sentença contendo oração relativa (**acl:relcl**) na qual o pronome relativo tem função de sujeito (**nsubj**). Como pode ser observado, nas *enhanced dependencies* a palavra plena “ano” passa a ser simultaneamente **nsubj** de “passou” e **nsubj** de “ficou”.

O ano que passou ficou na história.

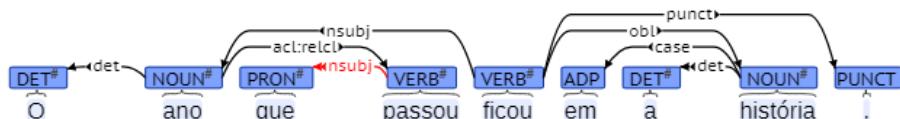


Figura 81a. Anotação de pronome relativo com função de **nsubj** nas *basic dependencies*.

O ano que passou ficou na história.

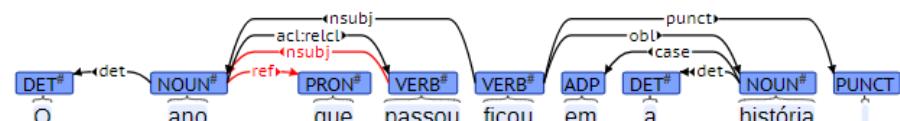


Figura 81b. Anotação da relação **ref** e realocação do dependente de **nsubj** nas *enhanced dependencies*.

As Figuras 82a e 82b, ilustram a anotação nas *basic* e nas *enhanced dependencies*, respectivamente, de uma sentença contendo oração relativa (**acl:relcl**) na qual o pronome relativo “o qual” tem função de objeto indireto não pronominal (**obl**). Como pode ser observado, nas *enhanced dependencies* a palavra plena “problema” passa a ser simultaneamente **nsubj:pass** de “resolvido” e **obl** de “reclamei”.

O problema do qual reclamei foi resolvido.

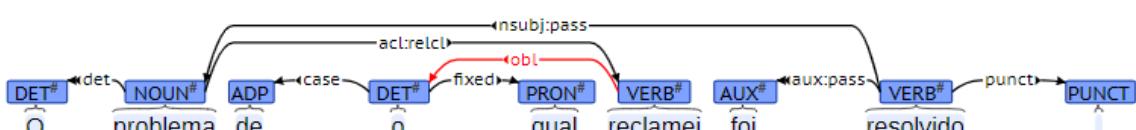


Figura 82a. Anotação de pronome relativo “o qual”, que é **obl** nas *basic dependencies*.

O problema do qual reclamei foi resolvido.

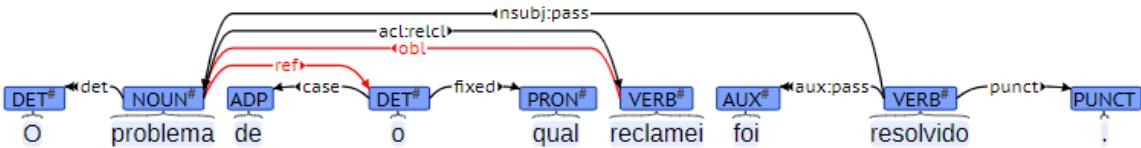


Figura 82b. Anotação da relação **ref** e realocação do dependente de **obl** nas *enhanced dependencies*.

Nas próximas subseções são exibidos casos em que a anotação da *enhanced dependency* aqui discutida apresenta alguma especificidade.

5.1 Pronome relativo exercendo função de advmod

Quando um advérbio (“onde”, “como” e “quando”) é usado no lugar de um pronome relativo, é preciso mudar a função de **advmod** para **obl**, ao se transferir, para o antecedente, a relação que chegava a esse advérbio, pois a função deixa de ser exercida por um advérbio e passa a ser exercida por um substantivo. Isso pode ser observado nas Figuras 83a e 83b, que ilustram as *basic* e as *enhanced dependencies* de uma sentença em que “onde” ocorre com função de pronome relativo.

A casa onde nasci não existe mais.

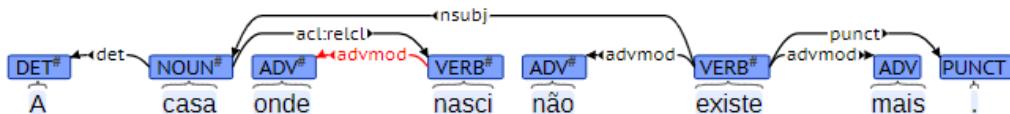


Figura 83a. Anotação do advérbio “onde” como **advmod** nas *basic dependencies*.

A casa onde nasci não existe mais.

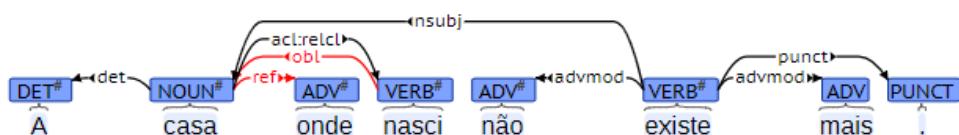


Figura 83b. Realocação do dependente e mudança de **advmod** para **obl** nas *enhanced dependencies*.

5.2 Pronomes relativos que são seu próprio antecedente

Alguns pronomes relativos podem amalgamar duas funções sintáticas: antecedente, *head* de **acl:relcl**, e pronome relativo integrante da oração dependente de **acl:relcl**. Como um mesmo *token* não pode ser dependente de mais de uma relação, prioriza-se a relação de mais alto nível, ou seja, aquela cujo *head* está mais próximo à raiz da árvore sintática.

Na sentença ilustrada na Figura 84a, o pronome relativo “Quem” foi anotado apenas como **nsubj** da oração matriz (“conseguiu”) embora devesse ser também **nsubj** da oração subordinada (“chegou”). A **acl:relcl**, nesses casos, não possui em seu interior um pronome relativo.

Quem chegou agora não conseguiu lugar.



Figura 84.a. Pronome relativo com dupla função anotado só com uma delas nas *basic dep.*

Quem chegou agora não conseguiu lugar.



Figura 84.b. Pronome relativo com dupla função anotado com as duas funções nas *enhanced dependencies*

Como nesse caso não há um pronome relativo dentro da **acl:relcl**, não há oportunidade de anotar a relação **ref** nas *enhanced dependencies*, pois o pronome “Quem” é seu próprio referente.

Há oportunidade, contudo, para anotar a função **nsubj** que o pronome relativo exerce na oração subordinada. Sendo assim, o pronome “Quem” passa a ser dependente **nsubj** de “conseguiu”, dependente **nsubj** de “chegou”, além de **head** de **acl:relcl**.

Para exemplificar a complexidade desses casos, apresentam-se duas sentenças nas quais os pronomes relativos têm duas funções. Figuras ilustrando suas respectivas anotações nas *basic* e nas *enhanced dependencies* são fornecidas na sequência (Figuras 85a e 85b, Figuras 86a e 86b).

O **medo** de **quem** foi **ferido** é sofrer de novo. (quem=aquele que)

=> “quem” é **nmod** de “medo” e **nsubj:pass** de “ferido”

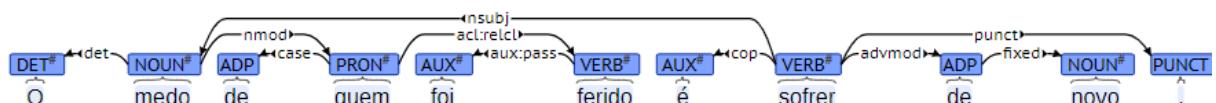


Figura 85a. Pronome relativo com dupla função anotado só com uma delas nas *basic dep.*

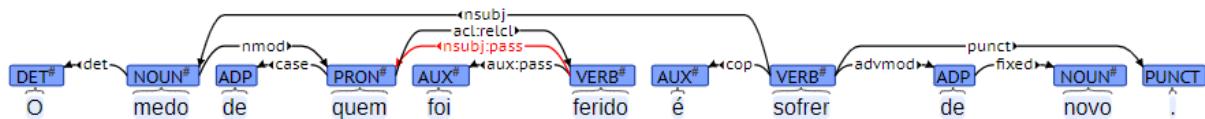


Figura 85b. Pronome relativo com dupla função anotado com as duas nas *enhanced dep.*

As pessoas de **onde** eu **venho** são solidárias. (**onde**=o lugar que)

=> “onde” é **nmod** de “pessoas” e **advmmod** de “venho”

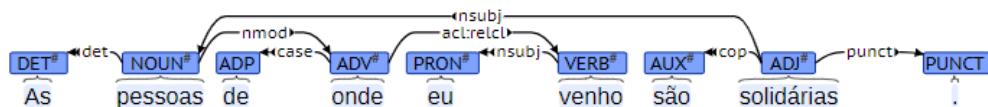


Figura 86a. Pronome relativo com dupla função anotado só com uma delas nas *basic dep.*

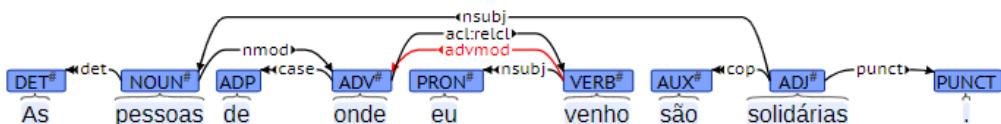


Figura 86b. Pronome relativo com dupla função anotado com as duas nas *enhanced dep.*

5.3 Pronomes relativos que são o predicado da **acl:relcl**

Nas orações **acl:relcl** cujo predicado é nominal, pode acontecer de o próprio pronome relativo ser o núcleo do predicado. Nesses casos, além de anotar a relação **ref** entre o pronome relativo e seu antecedente, desloca-se o sujeito da **acl:relcl** (se estiver explícito) para o antecedente.

As Figuras 87a e 87b ilustram essa situação, que no português é bem rara. Observa-se que, embora “onde” tenha como dependentes um **nsubj** (“ele”) e um **advmmod** (“hoje”), por determinação da UD apenas o **nsubj** é transferido para o token “interior”, antecedente de “onde”.

Ele mudou-se para o interior, onde ele está até hoje.

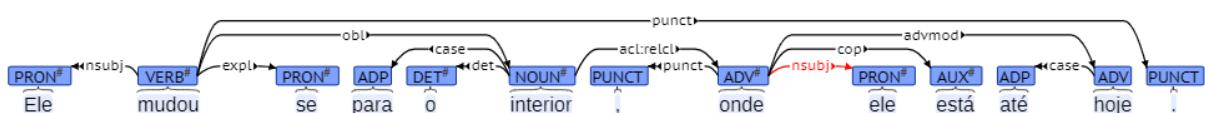


Figura 87a. Pronome relativo, predicado da **acl:relcl**, anotado nas *basic dependencies*

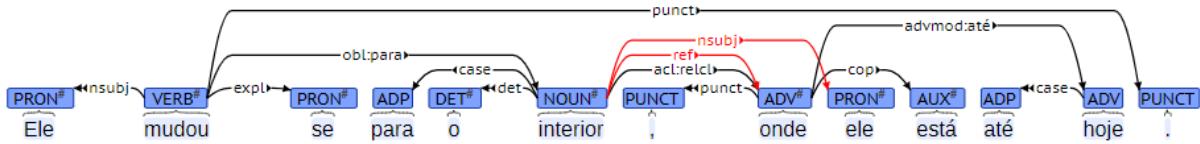


Figura 87b. Pronome relativo, predicado da **acl:relcl**, anotado nas *enhanced dependencies*

5.4. Pronome relativo não ligado diretamente ao predicado da **acl:relcl**

O pronome “cujo” e suas flexões desempenham a função de **det**, ligando-se ao substantivo que representa a “coisa” possuída pelo substantivo *head* da **acl:relcl**.

Nas *enhanced dependencies*, a relação **det** é eliminada e o substantivo ao qual “cujo” se ligava passa a ser anotado como *head* de **nmod**, relação cujo dependente é o *head* da **acl:relcl**. Os dois exemplos a seguir mostram como essa relação **nmod** é inferida:

a pessoa cujo nome = o nome da pessoa

=> “pessoa” é **nmod** de “nome”

os livros cujo preço = o preço dos livros

=> “livros” é **nmod** de “preço”

As Figuras 88a e 88b ilustram a anotação, nas *basic* e nas *enhanced dependencies*, de uma sentença que contém o pronome “cujo”. Observa-se que o antecedente do pronome relativo, *head* de **acl:relcl**, é dependente do substantivo que era modificado por “cujo”.

Esse é o diretor cujo mandato expira em 2025.

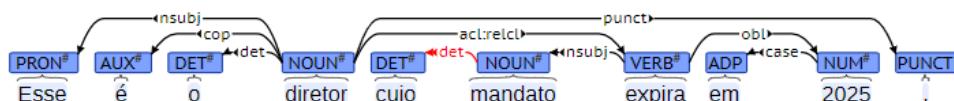


Figura 88a. Anotação do pronome “cujo” nas *basic dependencies*

Esse é o diretor cujo mandato expira em 2025.

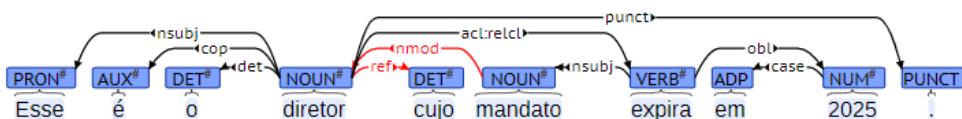


Figura 88b. Anotação do pronome “cujo” nas *enhanced dependencies*

5.5 Conclusões sobre a *enhanced dependency* de correferência nas acl:relcl

Conforme foi visto nesta seção, a *enhanced dependency* que introduz a relação **ref** na anotação não é intuitiva, apesar de ser bem clara a relação de correferência entre os pronomes relativos e seus antecedentes. No entanto, é possível depreender dos exemplos que a ideia básica é substituir, nas **acl:relcl**, os pronomes relativos pelos nomes que eles representam.

Pelo que se observou no projeto Porttinari, os casos mais complexos (ilustrados em 5.2 e 5.3) são também os mais raros.

6. Inserção de caso

Antes de mais nada, é preciso esclarecer o que significa o termo “caso”. Esse termo é usado nas línguas desinenciais, ou seja, línguas que “marcam” o papel semântico por meio de um sufixo e, por isso, a posição das palavras na oração não importa.

Nas línguas posicionais (não desinenciais), como a língua portuguesa, os papéis semânticos são deduzidos a partir da posição das palavras na oração (para sintagmas nominais não preposicionados, que são sujeito ou objeto) e pelas preposições que introduzem sintagmas nominais preposicionados e conjunções que introduzem orações coordenadas ou subordinadas. Por exemplo, orações adverbiais introduzidas pela preposição “por” ou pela conjunção “porque” recebem, normalmente, o papel de Causa:

Eles foram baleados **porque resistiram à ordem de prisão**.

Eles foram indiciados **por estarem vendendo drogas**.

O termo “caso” usado em línguas posicionais refere-se, portanto, a uma “pista” sintática para deduzir o papel semântico²⁴. A *enhanced dependency* em foco nesta seção acrescenta, ao nome das relações cujo dependente elas introduzem, o token das preposições e das conjunções, dependentes das relações **case**, **mark** e **cc**. A Figura 89 ilustra esse acréscimo em cinco relações (em vermelho): três do tipo **obl** (**obl:em**, **obl:em**, **obl:de**), uma do tipo **nmod** (**nmod:de**) e uma do tipo **advcl** (**advcl:apesar_de**). Destaca-se o fato de que a conjunção subordinativa (dependente de **mark**) que introduz a **advcl** do exemplo é uma expressão **fixed**: “apesar de”.

Votamos nele nas eleições de 2022 apesar de sabermos de seus defeitos.

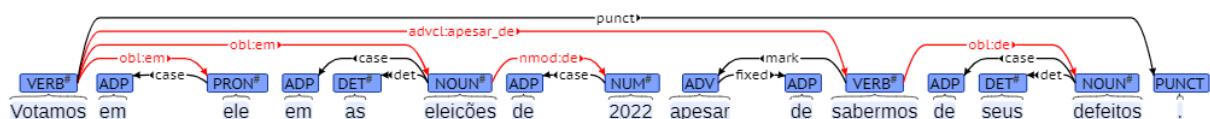


Figura 89. Inserção de caso nas relações **case** e **mark**

As relações cujas *enhanced* são acrescidas da informação de caso no português são: **acl**, **advcl**, **advmod**, **ccomp**, **nmod**, **obl** e **xcomp**. São exemplos as Figuras 90 e 91.

²⁴ Essa equivalência entre as preposições, nas línguas posicionais, e as desinências, nas línguas desinenciais foi discutida no artigo seminal “The case for case” (Fillmore, 1967).

Embora assustado, disse que estava calmo.

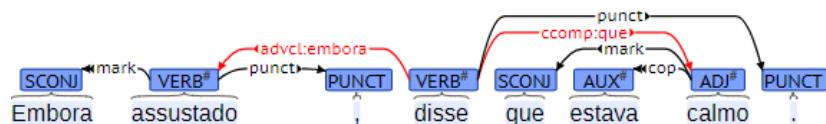


Figura 90. Inserção de caso nas relações **advcl** e **ccomp**

Todos gostam de relaxar ao chegar em casa.

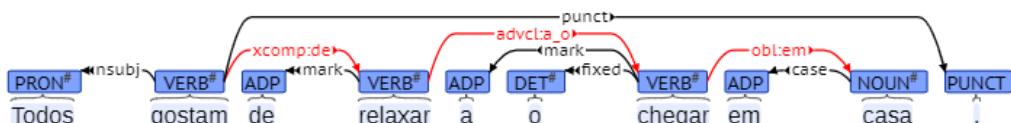


Figura 91. Inserção de caso nas relações **xcomp**, **advcl** e **obl**

6.1 Inserção de caso propagada em coordenações aditivas e alternativas

Quando há várias relações **conj** partindo de um mesmo *head* e apenas a última possui uma conjunção (dependente da relação **cc**), a inserção da conjunção pode ser propagada para todas as **conj** anteriores, como ilustra a Figura 92. Essa propagação, contudo, só pode ser feita quando a conjunção é aditiva ou alternativa ("e" e "ou" respectivamente).

Ele veio, almoçou, conversou e voltou para o trabalho.

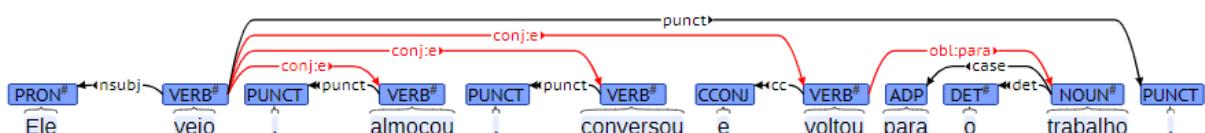


Figura 92. Inserção de caso nas relações **conj** (aditiva) e **obl**

Quando há várias relações **conj** partindo de um mesmo *head* e uma delas for adversativa, a conjunção adversativa não propaga para as **conj** anteriores, conforme ilustra a Figura 93.

Ele veio, almoçou, mas não tocou no assunto.

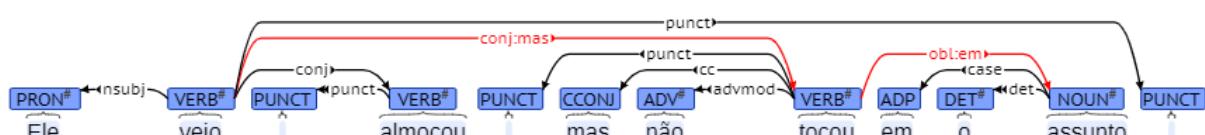


Figura 93. Inserção de caso nas relações **conj** (adversativa) e **obl**

Importante:

O acréscimo de caso utiliza a forma das preposições e conjunções e não o lema. Essa informação é importante para os casos em que **case**, **cc** e **mark** são constituídos por mais de um *token*, ou seja, por uma expressão **fixed**. Assim, a expressão **fixed** “salvo se”, com função de **mark**, composta por VERB e SCONJ, não deve ser anotada com os lemas das respectivas palavras, “salvar_se”, mas sim com as formas, “salvo_se”.

6.2 Conclusões sobre a *enhanced dependency* de inserção de caso

O propósito da *enhanced dependency* de inserção de caso é auxiliar na desambiguação de papéis semânticos. Embora o próprio léxico das preposições e conjunções possa apresentar alguma ambiguidade (pois não são exclusivas de um único papel semântico), informar o caso reduz o número de possíveis papéis.

Por exemplo, a preposição “com” está relacionada aos papéis semânticos de companhia e instrumento, ao passo que a preposição “em” está mais fortemente relacionada aos papéis semânticos de tempo e local, mas ambas são também usadas para introduzir sintagmas ou orações que complementam o sentido de alguns verbos, como é o caso de “preocupar-se **com** / importar-se **com**” e “insistir **em** / falar **em** / acreditar **em**”.

Essa é a *enhanced dependency* mais numerosa em qualquer córpus e sua anotação manual seria dispendiosa e altamente sujeita a falhas de esquecimento. Devido a sua baixa complexidade, é a mais simples de ser tratada computacionalmente.

7. Novas enhanced dependencies, não descritas na UD

As diretrizes de *enhanced dependencies* descrevem os seis tipos de operações discutidas nas seis primeiras seções deste relatório. Esses tipos de *enhanced dependencies* não são dependentes de língua, embora possam apresentar particularidades dependendo de cada língua.

Durante os estudos para implementação de *enhanced dependencies* em português, contudo, foram detectadas oportunidades para novas *enhanced*, aumentando a anotação de relações inferíveis entre os *tokens*.

A seguir serão discutidas essas possibilidades, embora não seja aceita a anotação de novos tipos de *enhanced* nos círculos disponibilizados no site da UD. Isso é compreensível, pois restringir os tipos de *enhanced dependencies* é importante para garantir comparatividade entre os círculos das diversas línguas.

Isso não significa, contudo, que uma versão mais enriquecida não possa ser produzida e disponibilizada fora do site da UD.

Esta seção tem por objetivo registrar essas novas *enhanced dependencies* concebidas, abrindo caminho para sua implementação.

7.1 Propagação do sujeito de ccomp

Nas línguas como o português, que admitem a elipse do sujeito (chamadas de “línguas *pro drop*²⁵”), as orações subordinadas do tipo **ccomp** podem ter o sujeito elíptico e esse sujeito pode coincidir com o sujeito da oração *head* de **ccomp**. Nesses casos, é possível criar uma relação *enhanced* propagando o sujeito do *head* de **ccomp** para o dependente de **ccomp**. Essa propagação de sujeito de **ccomp** colocaria as línguas *pro drop* em condições de igualdade com as línguas que não admitem a elipse de sujeito. A Figura 94 ilustra a propagação de um sujeito de **ccomp** no português e a Figura 95 ilustra a sentença equivalente em inglês, língua na qual o sujeito do dependente de **ccomp** não pode sofrer elipse e, portanto, não precisa da anotação *enhanced* nesse caso.

Ele disse que **[ele]** vai aposentar.

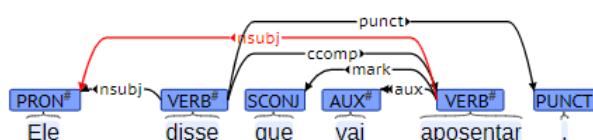


Figura 94. Propagação de sujeito de **ccomp** no português

²⁵ Das línguas românicas, apenas o francês não é *pro-drop*.

He said he will retire.

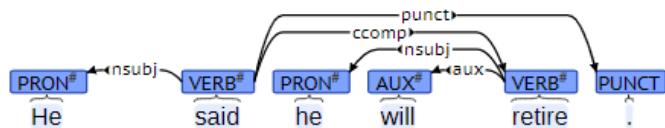


Figura 95. Sentença equivalente, em inglês, na qual **ccomp** não pode ter sujeito elíptico

Como se pode observar, as duas orações de cada figura teriam seus sujeitos anotados (“disse” e “aposentar”, na Figura 94 e “said” e “retire” na Figura 95).

Além de facilitar a equivalência entre as línguas, a propagação do sujeito para preencher o sujeito elíptico do dependente de **ccomp** teria outra vantagem: não interromperia a anotação de cadeias de sujeitos implícitos. Isso é demonstrado ao se contrastar a anotação sem propagação do sujeito de **ccomp** (Figura 96) e a anotação com a propagação do sujeito. Na Figura 97, a anotação do sujeito do dependente de **ccomp** (“aposentar”) torna possível propagar o mesmo sujeito para a oração dependente de **conj** (“pretende”) e anotar o sujeito controlador da oração dependente de **xcomp** (“viajar”).

Ele disse que vai aposentar e pretende viajar durante um ano.

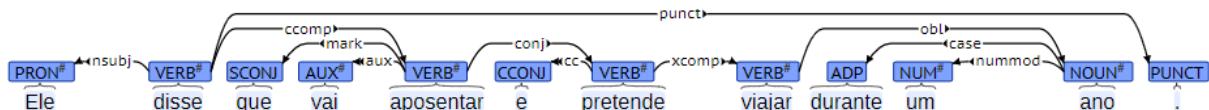


Figura 96. Sentença com **ccomp** sem propagação do sujeito.

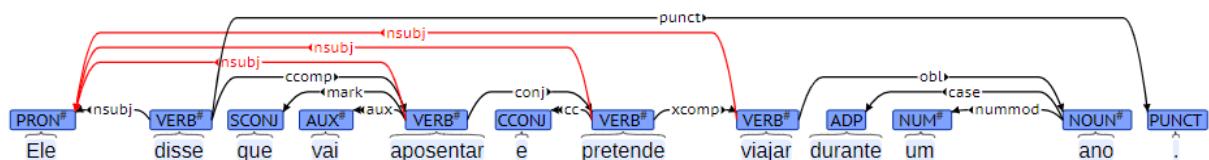


Figura 97. Sentença com **ccomp** com propagação do sujeito.

Novamente, isso aumentaria a equivalência entre os sujeitos anotados no português e os sujeitos anotados em inglês. Como no inglês o sujeito do dependente de **ccomp** é sempre explícito, a cadeia de sujeitos não fica prejudicada, como mostra a Figura 98:

He said he's going to retire and intends to travel for a year.

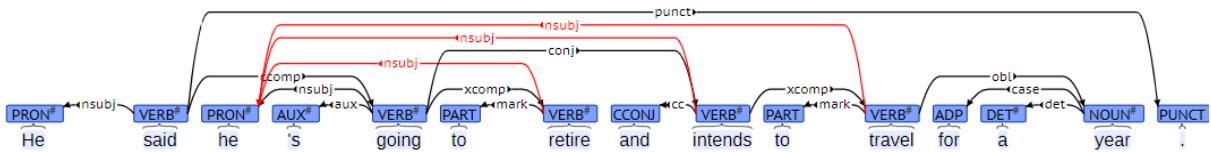


Figura 98. Anotação de sentença em inglês, sem necessidade de *enhanced* de **ccomp**.

A propagação do sujeito do *head* para o dependente de **ccomp** requer que algumas condições sejam atendidas:

1. A oração dependente de **ccomp** não pode ter seu próprio sujeito; caso contrário, não há elipse a ser preenchida. Ex:

Perceberam que o clima está realmente **mudando**.

=> “clima” é sujeito de “mudando”, então não há o que propagar.

2. A oração *head* de **ccomp** precisa ter um sujeito explícito ou atribuído por outra *enhanced*; caso contrário, não há o que propagar. Ex:

Perceberam que **erraram**.

=> “perceberam” não tem sujeito explícito nem *enhanced*, então não há sujeito a propagar

Eles sentiram que foram **prejudicados**.

=> o sujeito “eles” propaga de “sentiram” para “prejudicados”

Eles continuam dizendo que não **erraram**.

=> o sujeito “eles” propaga de “continuam” para “dizendo” e de “dizendo” para “prejudicados”

3. As *features* Person e Number dos verbos *head* e dependente de **ccomp** (verbo ou respectivos **aux**, **aux:pass** ou **cop**) precisam apresentar valores idênticos aos apresentados pelo dependente de **ccomp**, como nos exemplos a seguir:

Ele falou que **errou**.

(Person=3|Number=Sing no *head* e Person=3|Number=Sing no dependente)

Ele falou que **vai ser curado**.

(Person=3|Number=Sing e Person=3|Number=Sing no **aux:pass** do dependente)

Eu falei que **sou teimoso**.

(Person=1|Number=Sing no *head* e Person=1|Number=Sing no **cop** do dependente)

Eles **falaram** que **têm obtido** sucesso.

(Person=3|Number=Pl no head e Person=3|Number=Pl no **aux** do dependente)

Caso contrário, não deverá haver propagação, como nos exemplos a seguir:

Ele **falou** que **erramos**.

(Person=3|Number=Sing no *head* e Person=1|Number=Pl no dependente)

Ele **falou** que **vou ser curado**.

(Person=3|Number=Sing e Person=1|Number=Sing no **aux:pass** do dependente)

Eu **falei** que **são teimosos**.

(Person=1|Number=Sing no head e Person=3|Number=Pl no **cop** do dependente)

Eles **falaram** que **tenho obtido** sucesso.

(Person=3|Number=Pl no head e Person=1|Number=Sing no **aux** do dependente)

4. O dependente de **ccomp** não pode ser um verbo impessoal, pois, obviamente não há um lugar de sujeito nesse verbo (“haver” é o caso mais comum, mas há os verbos meteorológicos). Os exemplos a seguir mostram isso:

Ele **disse** que **há** indícios de fraude.

Ele **disse** que **choveu** em São Paulo.

Atendidas essas exigências, a propagação pode ser realizada, observando que, se a *feature* Voice aparece em apenas um dos verbos unidos pela relação **ccomp**, o tipo de sujeito precisa ser invertido. Isso significa que, se um dos predicados tiver Voice=Pass e o outro não, adapta-se a relação de **nsubj:pass** para **nsubj** ou **nsubj** para **nsubj:pass**, de forma que o sujeito do dependente da **ccomp** tenha um sujeito compatível com sua *feature* Voice. Os exemplos a seguir mostram essas duas situações.

Ele soube que será condecorado. (**nsubj** e **nsubj:pass**)

Eles foram comunicados que ganharam imunidade. (**nsubj:pass** e **nsubj**)

7.2 Propagação²⁶ do sujeito de **ccomp:speech**

A relação **ccomp:speech** requer as mesmas condições exigidas para propagação do sujeito de **ccomp**, exceto por uma: os valores das *features* Person e Number do *head* não precisam coincidir com os do dependente.

²⁶ Inspirada em discussão levantada por Claudia Freitas com relação à anotação de papéis semânticos.

A condição para propagação do sujeito do *head* para o dependente de **ccomp:speech** é feita sempre que o dependente (verbo ou respectivo **aux** ou **aux:pass** ou **cop**) estiver na primeira pessoa do singular ou do plural e não tiver seu próprio sujeito.

Não **vou desistir**, **disse** eu. (“vou” => Person=1)
=> “eu” é sujeito de “desistir”

Não **vou desistir**, **disse** João. (“vou” => Person=1)
=> “João” é sujeito de “desistir”

Não **vamos desistir**, **disse** eu. (“vamos” => Person=1)
=> “eu” é sujeito de “desistir”

Não **desisti**, **disse** João. (“desisti” => Person=1)
=> “João” é sujeito de “desistir”

Não **sou covarde**, disse João. (“desisti” => Person=1)
=> “João” é sujeito de “covarde”

Não **somos covardes**, disse João. (“somos” => Person=1)
=> “João” é sujeito de “covardes”

Não **somos covardes**, dissemos nós. (“somos” => Person=1)
=> “nós” é sujeito de “covardes”

A Figura 99 também ilustra a propagação do sujeito de **ccomp:speech**.

Produzimos experiências sonoras, afirma Roberto Guimarães.

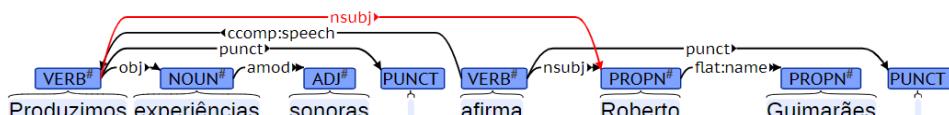


Figura 99. Propagação do sujeito de **ccomp:speech**

Se contudo, o dependente de **ccomp:speech** tiver seu próprio sujeito, obviamente o sujeito no *head* não será propagado, conforme ilustra a Figura 100, na qual “Roberto Guimarães” é **nsbj** do *head* “afirma” e “Nós” é **nsbj** do dependente, “produzimos”.

Nós produzimos experiências sonoras, afirma Roberto Guimarães.

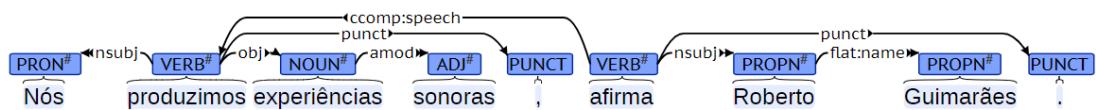


Figura 100. Head e dependente de **ccomp:speech** com seus próprios sujeitos

7.3 Propagação do sujeito de advcl

As orações subordinadas do tipo **advcl** também podem apresentar elipse de sujeito e, portanto, são candidatas a obter a propagação do sujeito de seus respectivos *heads*.

Contudo, como há vários tipos de orações adverbiais, muitas delas com várias elipses, as regras para propagação do sujeito do *head* para o dependente nem sempre alcançam uma boa precisão.

Julgou-se conveniente iniciar esta exposição mostrando exemplos de orações adverbiais que admitem e que não admitem propagação, para então falar em regras e, por fim, discutir particularidades desse tipo de *enhanced*. Nos exemplos a seguir, o predicado da oração matriz está em negrito, o predicado da subordinada adverbial em vermelho e o sujeito propagado em azul (tachado nos casos de não propagação).

7.3.1 Exemplos de **advcl** em que o sujeito pode ser propagado

As irmãs sempre **chamavam** a mãe quando **[as irmãs] queriam** conselhos.

Ela conta que os pais, embora **[os pais]** a tenham **apoiado**, nunca a **entenderam**.

O preto só não é **básico** se **[o preto] receber** algum brilho.

Furacão Nate **perde** força e vira tempestade tropical ao **[Furacão Nate] atingir** EUA.

O Brasil **anda** tão violento que **[o Brasil]** não é mais **Pátria** Amada, é Pátria Armada!

Gil **vaijava** para **[Gil] fazer** show.

O Conselho de Ética **arquivou** o caso sem **[o Conselho de Ética]** nem sequer **abrir** investigação.

Ferrer tinha **ido** a Quito **[Ferrer] jogar** futebol.

Che **esperava** em Guayaquil para **[Che] embarcar** rumo a o Panamá.

7.3.2 Exemplos de **advcl** em que o sujeito não deve ser propagado

*Quando **[e-veículo] acordou**, o veículo não estava mais **lá**. (obviamente, “o veículo” não é sujeito de “acordou”)

Minha intenção é **contribuir** com isso, **[minha-intenção] trazendo** dados confiáveis.

Aqui, **[70%-da-carga] tirando** minério de ferro, 70% de a carga **transita** em rodovia. (aqui há indeterminação do sujeito não marcada pelo “se”: “tirando[-se] o minério de ferro”)

Esses resultados **classificam** os chilenos sem que [esses resultados] **tenham** de se preocupar com outros jogos. (quem não tem que se preocupar são os chilenos, objeto direto de “classificam”)

Essa tecnologia está sendo **implantada** em toda a rede, **começando** [essa tecnologia] pela capital. (não é a tecnologia que começa, mas sim a implantação)

A ideia era **facilitar** a divulgação de alertas para a população, [a ideia] **evitando** jargões científicos. (o sujeito da adverbial é indeterminado: “evitando[-se] jargões científicos”)

7.3.3 Condições básicas para propagação de sujeito de **advcl**

De modo geral, as condições para propagação são as mesmas exigidas para propagação do sujeito de **ccomp**. Embora essas condições não garantam que o sujeito possa ser propagado, elas pelo menos garantem a exclusão de casos que, com certeza, não devem ser propagados.

1. O *head* de **advcl** precisa ter um sujeito;
2. O dependente de **advcl** não pode ter sujeito próprio;
3. O verbo do dependente de **advcl** não pode ser impessoal²⁷ ;
4. O verbo do dependente de **advcl** (ou seu auxiliar ou seu cípula) devem ter os mesmos valores nas *features* Person e Number que o verbo do *head* de **advcl**. Essa condição faz sentido para orações finitas, mas não faz sentido se a oração adverbial for uma reduzida de infinitivo, particípio ou gerúndio;
5. O verbo do dependente de **advcl** não pode ser simultaneamente *head* de **expl:impers**, isto é, pronome “se” marcador de indeterminação de sujeito. Se o corpus não diferenciar os tipos de **expl**, é mais seguro não propagar o sujeito sempre que o dependente de **advcl** for head de **expl** e o *token* dependente for “se”, como nos casos a seguir:

Como **se sabe**, ele **escreveu** o livro após voltar de viagem. (não propaga)

O dólar **subiu** muito, como já **se sinalizava**. (não propaga).

Ela só **aceitou** o emprego porque lá **se paga** bem. (não propaga)

Mais muros serão construídos, porque **se trata** de uma resposta fácil. (não propaga)

²⁷ Verbos impessoais não admitem sujeito. São impessoais os verbos meteorológicos: amanhecer, anoitecer, entardecer, chover, ventar, trovejar, nevar, etc. e os verbos existenciais que só admitem objeto: haver, ter (no sentido de haver), fazer (no sentido de tempo decorrido).

Contudo, ignorando todos os **expl** “se”, perdem-se casos que deveriam propagar, como:

O policial **atirou** porque [o policial] **se sentiu** ameaçado. (propaga)

Os irmãos se **abraçaram** como se [os irmãos] não **se vissem** há anos. (propaga)

Atenção:

Se a anotação da *feature Person* não estiver correta, a regra 4 pode não funcionar. Há muitas formas verbais que são iguais para a primeira e a terceira pessoa do singular e se a anotação da *feature Person* foi feita automaticamente, pode haver erro de anotação. Exemplos de formas ambíguas entre a primeira e a terceira pessoas do singular: eu/ele [tinha, estava, andava, corria, cantava, etc.], eu/ele [teria, estaria, andaria, correria, cantaria, etc.].

Eu não **desisti** do cinema porque [eu] **estava** cansado. (“estava” Person=1)

Eu não **desisti** do cinema porque [cinema] **estava** lotado. (“estava” Person=3)

Além disso, o verbo “ter” pode gerar confusão, pois se estiver em seu sentido existencial, sinônimo de “haver”, não admite sujeito. Como a forma pode ser ambígua, dificulta a aplicação da regra 3.

Ele não **foi** ao cinema porque [ele] não **tinha** companhia.

Ele não **foi** ao cinema porque não **tinha** cinema aberto naquela hora.

7.3.4 As elipses nas **advcl** e suas consequências na propagação do sujeito

Uma característica das orações adverbiais é ter muitas elipses. No exemplo a seguir, as elipses estão indicadas em azul:

O resultado seria uma melhora, embora [o resultado] fosse [uma melhora] menor que a [melhora] esperada.

Caso só a elipse do sujeito fosse resolvida na *enhanced*, o sentido poderia ficar corrompido:

*O resultado seria uma melhora, embora o resultado fosse menor que a esperada.

Sendo assim, parece ser melhor não propagar o sujeito nesses casos a fim de não produzir um resultado confuso.

Esse problema é mais frequente nas orações adverbiais comparativas, nas quais quase tudo está elíptico, exceto o que distingue os termos comparados. Se o próprio predicado

está elíptico, fica impossível verificar se a oração tem um sujeito próprio, pois o sujeito só seria revelado caso houvesse a inserção de um *token* vazio para representar o predicado elíptico. A sentença a seguir ilustra essa situação.

Ele é mais **pernicioso** para os policiais que o Marcola [é pernicioso para os policiais].

“Marcola” é sujeito da **advcl** comparativa e, portanto, não deveria haver propagação de “Ele”, mas esse sujeito só se revela se houver a inserção de um *token* vazio para representar o predicado nominal elíptico: “pernicioso” (o que não está previsto na UD ainda). Como as adverbiais comparativas são muito complexas²⁸, recomenda-se que não sejam submetidas à propagação automática de sujeito, exceto se a propagação for verificada caso a caso, manualmente.

7.3.5 Inversão de voz do sujeito

Nos casos em que o dependente de **advcl** está no particípio e tem a *feature* Voice=Pass, pode ser necessário adequar o tipo de sujeito. No exemplo a seguir (Figura 101), “STF” é **nsubj** de “pode” na oração matriz, mas é **nsubj:pass** de “exposto”

Até onde pode ir o STF quando [STF] [é] exposto a constrangimento?

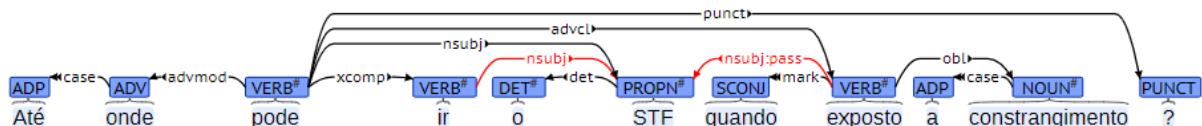


Figura 101. Sujeito da ativa propagado como sujeito da passiva em **advcl** temporal

A oração na forma de particípio pode ser uma reduzida, como o exemplo anterior, ou uma oração desenvolvida, como o exemplo a seguir:

O prazo **encerrou** ontem, depois de [o prazo] ter sido **prorrogado** por 2 meses.

7.3.6 A relação entre tipos de **advcl** e propagação “correta” de sujeitos

Como os tipos de **advcl** (causais, comparativas, concessivas, condicionais, conformativas, consecutivas, finais, modais, proporcionais, temporais) podem ser previstos com razoável segurança a partir dos *tokens* dos **mark** que as introduzem, pode-se tratar cada tipo separadamente ao implementar a propagação do sujeito.

²⁸ Um estudo sobre as elipses nas orações adverbiais comparativas foi relatado em Duran et. al. (2024).

Ao se estudar as **advcl** em círculos a fim de buscar regularidades que possam ser transformadas em regras, observou-se que alguns tipos de **advcl** são mais “comportados” e possibilitam uma propagação automática mais segura do sujeito. É o caso das **advcl** temporais e causais.

Por outro lado, há dois tipos de **advcl** em que a propagação quase não acontece: as comparativas e as conformativas. Por esse motivo, uma boa medida seria evitar a propagação automática de sujeito nesses tipos de **advcl**.

O problema das orações comparativas foi explicado na subseção 7.3.4, que discute as elipses nas **advcl**. Conforme mostra o exemplo a seguir, o sujeito da **advcl** pode estar presente e não estar anotado como tal, devido à elipse do predicado (indicado em azul).

Os brasileiros são mais **resistentes** ao frio do que os europeus [são resistentes] ao calor.

O problema das orações conformativas, porém, parece envolver um viés de anotação: orações iniciadas pelas conjunções “conforme”, “como” e “segundo”, foram classificadas como **advcl**, porém, em alguns casos essas orações assemelham-se mais a **parataxis**, pois não modificam a oração matriz, mas sim constituem um comentário paralelo a ela, como pode ser visto nos exemplos a seguir:

Esse fato o abalou, conforme vem sendo revelado. => Vem sendo revelado que esse fato o abalou.

Segundo consta no processo, o imóvel foi vendido por R\$ 4,5 milhões. => Consta no processo que o imóvel foi vendido por R\$ 4,5 milhões.

Na paráfrase apresentada ao lado de cada sentença, sugere-se que, adicionando elementos de coesão, a relação de subordinação poderia, inclusive, ser invertida: a matriz viraria subordinada completiva e a subordinada viraria matriz, o que prova que essas orações não são nem adverbiais e nem subordinadas.

As *enhanced*, nesse caso, levam ao questionamento de decisões de anotação nas *basic dependencies*, mostrando diferenças sutis de sentido que não foram observadas.

7.3.7 Orações adverbiais reduzidas de infinitivo, gerúndio e particípio

Nas orações adverbiais reduzidas de infinitivo, gerúndio e particípio, a propagação apresenta singularidades.

As **advcl** reduzidas de infinitivo são quase sempre finais, causais ou temporais, e aceitam bem a propagação de sujeito. Exceto pelo caso com o verbo “ir”, todas as **advcl** reduzidas de infinitivo são marcadas, pois apresentam conjunção subordinativa (“para”, “por”, “ao”).

As colegas **foram** juntas à praia [as colegas] **nadar**.

Eu já **gastei** mais de R\$ 2 mil para [eu] **participar** do concurso.

O titular Diego Alves não **pode** atuar por [O titular Diego Alves] não estar **inscrito** no torneio.

Ao [Mário] **perceber** que estava enganado, Mário **pediu** desculpas.

As **advcl** reduzidas de particípio, quando o particípio é um verbo, quase sempre apresenta a *feature Voice=Pass* e isso exige que o sujeito propagado seja **nsubj:pass**:

Nenhum deles **saiu** [nenhum deles] **baleado**.

Quando o particípio é um adjetivo, contudo, o sujeito não é **nsubj:pass**

A vítima foi **encontrada** [a vítima] **desnutrida**.

Eles **foram** ao cinema, apesar de [eles] **cansados**.

As **advcl** reduzidas de gerúndio são muito frequentes e podem ser de diversos tipos. Quando são modais, quase sempre aceitam a propagação do sujeito:

O próprio Marcelo **chegou** à PF [Marcelo] **pxuando** a valise.

Regina **cresceu** [Regina] **observando** a mãe cozinhar.

O vendedor **respondia** minhas perguntas [o vendedor] **olhando** para o meu marido.

[Rubem Braga] **Trabalhando** como jornalista, Rubem Braga **produziu** cerca de 15 mil textos.

Porém, quando as **advcl** reduzidas de gerúndio têm outras funções, muitas vezes não aceitam a propagação de sujeito, como nos exemplos a seguir:

Eles estão **bem**, [eles] **tirando** o problema da falta de dinheiro.

[e—problema] **Querendo** ou não, o problema do aquecimento global tem que ser **discutido**.

[você] **Falando** em filme polêmico, o que você **achou** do vendedor do Oscar?

7.3.8 Sujeitos propagáveis após a anotação de outras *enhanced*

Muitas vezes, o sujeito da oração *head* de **advcl** é simplesmente um pronome relativo e só com a anotação da *enhanced* de correferência (v. Seção 7) é que a palavra de conteúdo é propagada. Essa situação é ilustrada nas Figuras 102a e 102b, que mostram a anotação nas *basic* e nas *enhanced dependencies*, respectivamente, de uma **advcl** subordinada a uma **acl:relcl**. O sujeito da adverbial comparativa, introduzida por “como”, é “fato”, mas o sujeito da oração matriz é “que”. Com a anotação de **ref** entre o “que” e seu antecedente “fato”, o sujeito propagado para a **advcl** passa a ser “fato”.

Foi um fato isolado, que **aconteceu**, como [um fato isolado] **acontece** às vezes até em outros países.

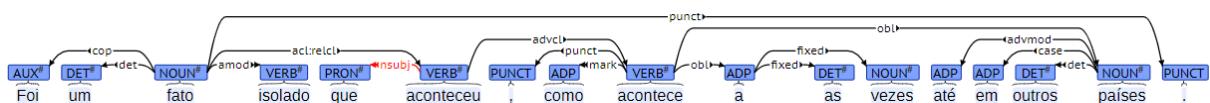


Figura 102a. Anotação de uma **advcl** subordinada a uma **acl:relcl** - *basic dependencies*

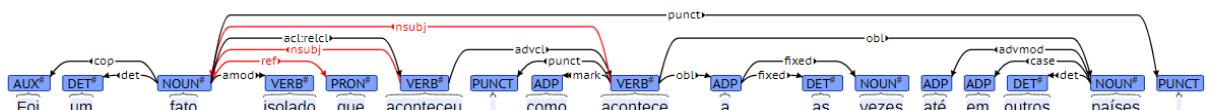


Figura 102b. Anotação de uma **advcl** subordinada a uma **acl:relcl** - *enhanced dependencies*

7.3.9 Conclusão sobre a propagação de sujeitos de **advcl**

Embora os casos de não propagação de sujeitos de **advcl** sejam menos frequentes que os de propagação, é aconselhável revisar manualmente as *enhanced* das **advcl**. Caso isso não seja possível, a decisão conservadora da UD, de não propagar sujeitos de **advcl**, pode ser mais acertada. Também é possível tomar uma decisão de anotação parcial, com bom grau de segurança, propagando apenas sujeitos de **advcl** temporais e causais.

7.4 Atribuição de sujeito de **acl**

Os dependentes de **acl** são orações adjetivas ou completivas nominais e sempre apresentam forma de orações reduzidas, isto é, estão no particípio, gerúndio ou infinitivo. A relação **acl** une um *token* substantivo a uma oração. Em algumas condições, o *token head* da **acl**, nas *basic dependencies*, é também o sujeito do dependente da **acl** nas *enhanced dependencies*.

7.4.1 Dependente de **acl** no particípio

As **acl** cujo dependente é um verbo no particípio são orações adjetivas reduzidas de voz passiva. Por isso, o *token head* da relação **acl** será **nsubj:pass** do dependente. Isso é exemplificado nas sentenças a seguir, nas quais o *head* da **acl** está em negrito e o dependente da **acl** está destacado em vermelho.

Houve um **contingenciamento** de despesas, **ordenado** no início do ano.

= o contingenciamento foi ordenado no início do ano

Um exemplo de sucesso é o **plástico feito** com etanol de a cana-de-açúcar.

= o plástico foi feito com etanol de a cana-de-açúcar

Esta é uma estratégia **adotada** pelo delator.

= a estratégia foi adotado pelo delator

O que houve foi histeria, **alimentada** por doses inacreditáveis de preconceito.

= histeria foi alimentada por doses de preconceito

A Figura 103 ilustra a anotação *enhanced* do sujeito **nsubj:pass** de uma **acl** com dependente no particípio.

A **certeza** de vitória **demonstrada** pelos advogados de defesa é surpreendente.

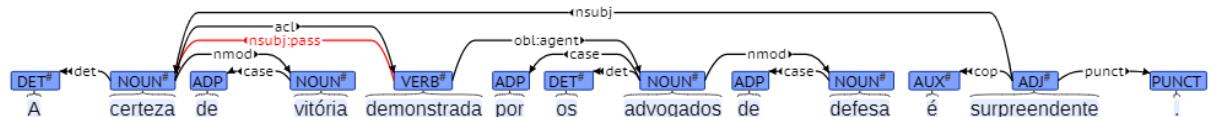


Figura 103. Anotação de **nsubj:pass** de dependente de **acl** no particípio

7.4.2 Dependente de **acl** no gerúndio

As **acl** reduzidas de gerúndio são sempre adjetivas e podem, seguramente, receber atribuição de sujeito, como mostra o exemplo a seguir. Contudo, o sujeito é **nsubj**, pois a oração é de voz ativa. Um exemplo dessa anotação é ilustrado na Figura 104.

Recebi uma **mensagem dizendo** que a carga tombou.

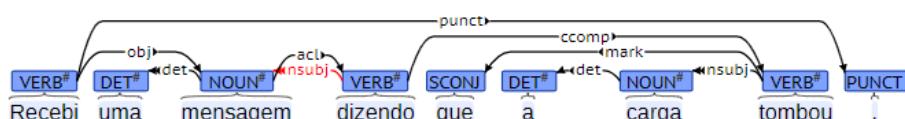


Figura 104. Anotação de **nsubj** de dependente de **acl** no gerúndio

7.4.3 Dependente de **acl** no infinitivo

O infinitivo pode constituir **acl** do tipo adjetivo (pouco frequente com infinitivo) e, nesses casos, é possível desenvolvê-la como uma oração relativa:

Raquel Dodge será a primeira **mulher** a **assumir** o cargo de PGR.

=> Raquel Dodge será a primeira **mulher** que **assumirá** o cargo de PGR.

A Folha foi **o único** de os principais jornais a não **utilizar** o verbo "denunciar" na manchete.

=> A Folha foi **o único** de os principais jornais **que não utilizou** o verbo "denunciar" na manchete.

Hernanes foi o primeiro **jogador** de o São Paulo a **falar**.

=> Hernanes foi o primeiro **jogador** de o São Paulo **que falou**.

Há casos também de orações adjetivas no infinitivo em que o *head* da **acl** poderia ser interpretado como objeto do dependente da **acl** ("preço a pagar" => "pagar o preço"). Esses casos admitem uma paráfrase por voz passiva e, por vezes, uma paráfrase por sujeito indeterminado:

Acredito que esse seja um **preço** razoável a **pagar** pelo serviço.

=> Acredito que esse seja um **preço** razoável a **ser pago** pelo serviço (passiva)

=> Acredito que esse seja um **preço** razoável a **se pagar** pelo serviço (indeterm.)

Há, porém, casos semelhantes de orações adjetivas que não admitem as mesmas paráfrases, como ilustra o seguinte exemplo

Você tem alguma **informação** a **acrescentar**?

=> *Você tem alguma **informação** a ser **acrescentada**?

=> *Você tem alguma **informação** a **se acrescentar**?

Mas, juntando um modalizador (verbo "poder"), novas paráfrases são possíveis:

=> Você tem alguma **informação** que possa ser **acrescentada**? (passiva)

=> Você tem alguma **informação** que **se** possa **acrescentar**? (indeterm.)

=> Você tem alguma **informação** que **[você]** possa **acrescentar**? (ativa)

No caso acima, uma opção seria propagar o sujeito da oração na qual o *head* da **acl** está inserido, como sugere a terceira paráfrase.

Na maioria das vezes, contudo, o infinitivo constitui uma **acl** do tipo complemento nominal, como exemplificado nas sentenças a seguir:

Propina é a **maneira** mais fácil de **atravessar** um muro na fronteira.

A proposta seria uma **forma** de **atingir** o equilíbrio fiscal.

Os lançamentos são um **convite** para **deixar** smartphones de lado.

Essas orações denotam uma indeterminação do sujeito, o que pode ser observado ao se adicionar um “se” indeterminador ao verbo:

Propina é a **maneira** mais fácil de **se atravessar** um muro na fronteira.

A proposta seria uma **forma** de **se atingir** o equilíbrio fiscal.

Os lançamentos são um **convite** para **se deixar** smartphones de lado.

Uma avenida a ser explorada é a de atribuir, à **acl** infinitiva, o sujeito da oração em que se encontra o substantivo *head* da **acl**. São, ao que parece, casos associados à expressão de sentimentos, como os estudados por Duran et al (2011), mostrados nas sentenças a seguir:

Os jovens estão com **dificuldade** de **lidar** com a ansiedade.

=> “jovens” sujeito de “lidar”

Entenda por que as pessoas têm **medo** de **tomar** vacinas.

=> “pessoas” sujeito de “tomar”

É natural criança sentir **vontade** de **brincar**.

=> “criança” sujeito de “brincar”

Uma vez que os casos de **acl** no infinitivo com sujeito indeterminado parecem ser muito mais frequentes do que os que têm sujeitos determinados, não é seguro automatizar a *enhanced* de atribuição de sujeito a elas, a não ser que se descubram regras para distinguir automaticamente **acl** adjetivas de **acl** completivas nominais e, entre as completivas nominais, quais aquelas que admitem um sujeito (por exemplo, quando o *head* da **acl** é uma palavra que expressa um sentimento).

7.5 Realocação de complementos separados de seus *heads* lógicos

A diretriz da UD de não anotar, nas dependências básicas, relações que produzam cruzamento de arcos (as chamadas “relações não projetivas”) acarreta erros de atribuição de *head* de algumas relações. Como nas *enhanced dependencies* o cruzamento de arcos é permitido, pode-se corrigir esses erros de *head*.

7.5.1 *Head* de **xcomp** usado como *head* de complementos ou modificadores do dependente de **xcomp**.

Um caso comum que requer realocação do *head* da relação são os complementos de um verbo dependente de **xcomp** que ocorre antes do *head* de **xcomp**, como nos exemplos a seguir:

Aliás, **isso** eu continuo sem **saber**, mas aprendi a instalar o aparelho.

Nesta sentença, “isso” deveria ser **obj** de “saber”, conforme ilustra a Figura 105²⁹, causando cruzamento de arcos.

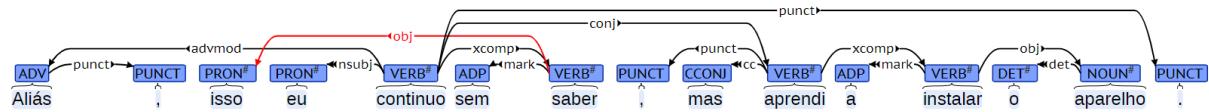


Figura 105. Anotação do **obj** anteposto com o *head* correto

Porém, para evitar cruzamento de arcos, “isso” foi anotado como **obj** de “continuo”, como ilustra a Figura 106.

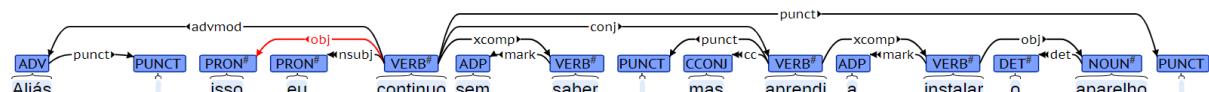


Figura 106. Anotação do **obj** anteposto com o *head* incorreto, para não produzir cruzamento de arcos nas *basic dependencies*

A correção, com substituição do *head*, poderia ser feita nas *enhanced dependencies*.

O mesmo ocorre na próxima sentença que, adicionalmente, apresenta uma oração relativa.

Eles pagam o **valor** que você querer **cobrar**.

A Figura 107a mostra a anotação desta sentença nas *basic dependencies*, com o pronome relativo “que” anotado como objeto de “querer”.

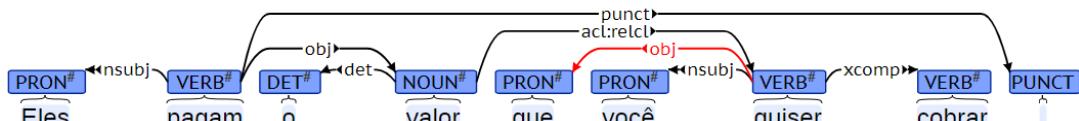


Figura 107a. Anotação de **obj** anteposto com *head* incorreto nas *basic dependencies*

A Figura 107b mostra o *token* “valor” anotado como **obj** de “querer” após aplicada a *enhanced* de correferência

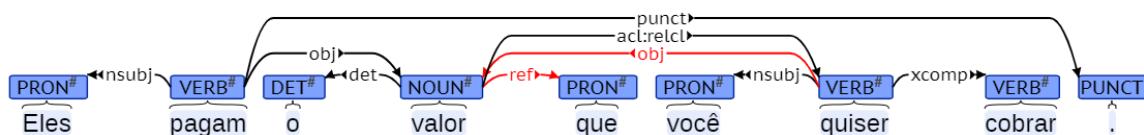


Figura 107b. Anotação de **obj** com *head* incorreto, porém com correferência

A Figura 107c, finalmente, apresenta a anotação correta, com o *token* “valor” anotado como **obj** de “cobrar”.

²⁹ Não estão exibidas as *enhanced* de propagação de sujeito de **xcomp** para dar destaque especial para o remanejamento do *head* da relação **obj**.

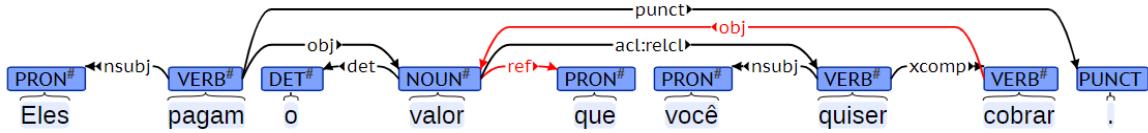


Figura 107c. Anotação de **obj** com *head* correto e correferência nas *enhanced*

Destaca-se o fato de que essa *enhanced* é do tipo que não reproduz a anotação das *basic dependencies*, mas sim apresenta uma anotação modificada, como fazem as *enhanced* de inserção de *token* elíptico e de inserção de correferência.

Casos como esses sempre ocorrem quando o **obj** está antes do *head* de **xcomp** e desde que o **obj** não seja um pronome acusativo (pronomes acusativos ocorrem antepostos normalmente, pois são simultaneamente objeto do *head* da **xcomp** e sujeito do dependente da **xcomp**). Ex:

Eu **o** deixei entrar. => deixei-o [=ele] entrar

Eu **os** considero suspeitos. => considero-os [=eles] [serem] suspeitos

Ele **me** fez chorar. => fez-me [=eu] chorar

Outra situação análoga ocorre com os **obl** antepostos. Alguns exemplos são listados a seguir, com o dependente de **obl** em vermelho, o *head* incorreto (anotado nas *basic dependencies*) sublinhado, e o *head* correto em negrito:

Não sei em que **lugar** você pretendia chegar.

Em que **hotel** você costuma ficar?

A que **pessoa** você gostaria de dirigir essa pergunta?

As Figuras 108a e 108b ilustram o antes e o depois da correção do *head* de um **obl** anteposto.

De quantas horas você pode dispor para essa atividade?

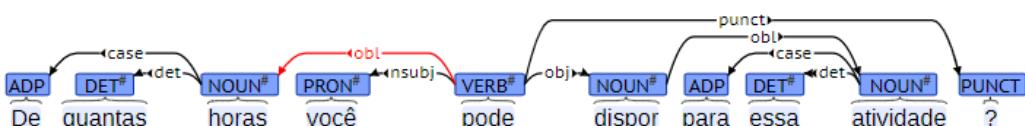


Figura 108a. **obl** anotado no *head* de **xcomp** para evitar cruzamento de arcos nas *basic dep.*

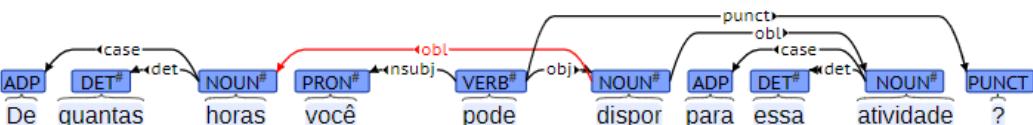


Figura 108b. **obl** com cruzamento, com *head* corrigido nas *enhanced dep.*

Porém, uma vez que um **obl** pode modificar o *head* de **xcomp**, não é tão simples identificar dependentes de **obl** com *heads* incorretos. As três sentenças a seguir apresentam **obl** que pertencem realmente ao *head* de **xcomp**.

Por que **razão** você não **tentou** avisá-lo?

Em que **ano** você **decidiu** mudar de vida?

Em **abril** eu **comecei** a trabalhar com vendas.

7.5.2 **nmod** complemento nominal anteposto, anotado como **obl** do verbo

Outro caso de anteposição é o **nmod** que complementa o sentido de um substantivo, anotado como **obl** do verbo, como nos casos a seguir:

Para essas **perguntas**, ninguém tem uma **resposta**.

Do **resultado**, nada vai nos dar uma **garantia**.

Disso você não precisa ter **vergonha**.

As Figuras 109a e 109b apresentam o antes e o depois da correção do *head* de um complemento nominal.

Do meu **amor** você pode ter **certeza**.

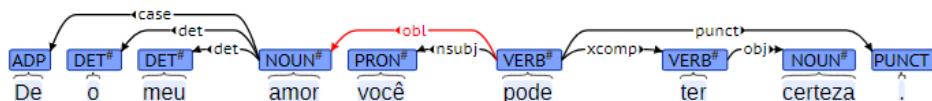


Figura 109a. **nmod** anotado como **obl** para evitar cruzamento de arcos nas *basic dep*.

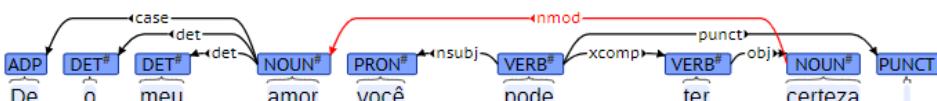


Figura 109b. **nmod** com cruzamento de arcos, com *head* corrigido nas *enhanced*

7.5.3 **nummod** com a parte inteira antes do NOUN e a parte fracionária após o NOUN

Outra situação em que o *head* da relação está incorreto são os **nummod** que ficam parte à esquerda e parte à direita de um NOUN, mas constituem parte de uma mesma quantidade. Ex:

Fiquei **uma hora** e **meia** esperando.

Nesses casos, o ideal seria coordenar “uma e meia”, pois esse é o modificador **nummod** de “hora”. Nas *basic dependencies*, para não cruzar arcos, o *token* “meia” está incorretamente coordenado a “hora”, como ilustra a Figura 110a. O correto seria corrigir isso na coluna das *enhanced*, produzindo a anotação ilustrada na Figura 110b.

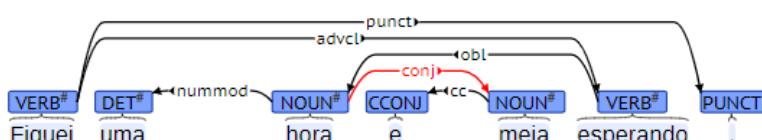


Figura 110a. Anotação com *head* incorreto de fração nas *basic dependencies*

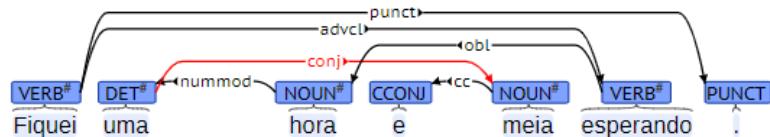


Figura 110b. Anotação com *head* correto de fração nas *enhanced dependencies*

Atenção: Esse tipo de “correção” de *head* de numerais fracionários só não se aplica a casos em que o *token* de uma fração está elíptico, como no exemplo a seguir, em que a palavra “minutos” foi suprimida.

Esperei uma hora e 15 [minutos] e não obtive resposta.

Se coordenássemos “uma” com “15”, teríamos “uma e 15 hora”, o que seria incorreto, pois “15” não quantifica “hora”.

7.5.4 Casos esporádicos de cruzamento de arcos evitado

Há casos em que a anotação correta produziria cruzamento de arcos, conforme ilustra a Figura 111a, que anota “no momento” como **obl** de “aplicar” e anota “contra gripe” como **nmod** de “vacinas”.

Pretendemos aplicar vacinas, no momento, apenas contra gripe.

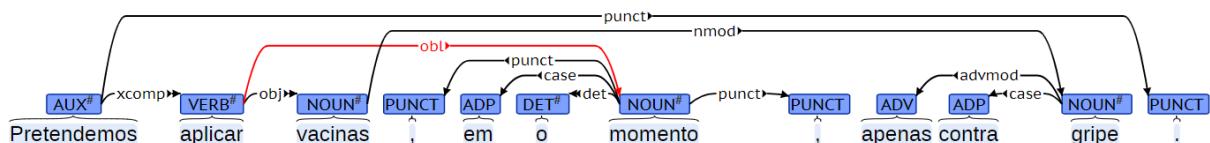


Figura 111a. Anotação de **nmod** e **obl** em seus *heads* corretos, com cruzamento de arcos

Para evitar cruzamento de arcos, optou-se por uma anotação que deixa o **obl** de “aplicar” anotado como **nmod** de “vacinas” (Figura 111b).

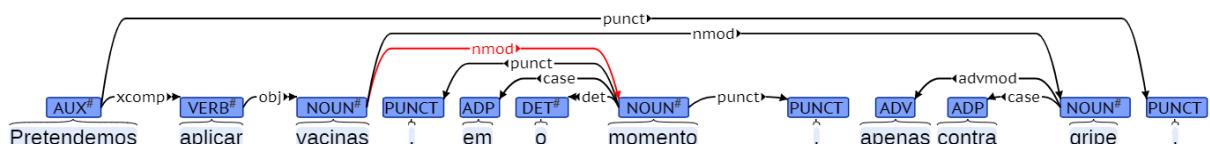


Figura 111b. Token anotado como **nmod** do NOUN, quando deveria ser **obl** do VERB

Como nas *enhanced* não existe restrição contra o cruzamento de arcos, casos assim poderiam ser corrigidos sempre que identificados.

7.5.5 Conclusão sobre *enhanced* de realocação de *heads* lógicos

Na coluna das *enhanced dependencies*, o cruzamento de arcos é permitido e isso abre oportunidade para serem corrigidas as anotações que atribuíram *heads* incorretos em algumas relações, a fim de evitar cruzamento de arcos nas *basic dependencies*. Em alguns

casos há heurísticas para detectar essas incorreções, porém, em casos esporádicos, não há como recuperá-las automaticamente.

Conclusões

Como pode ser observado ao longo deste relatório, as *enhanced dependencies*, desenvolvidas a partir da anotação das *universal dependencies*, são um tipo de anotação que representa as inferências do que está implícito no texto, mas não codificado na análise sintática.

Essas inferências utilizam os dados das colunas 7 e 8 do CoNLL-U, que codificam a análise sintática, além de, numa abordagem cíclica e recursiva, utilizar os próprios dados de uma *enhanced*, codificados na coluna 9 do CoNLL-U, para anotar outras *enhanced*.

Sendo assim, as *enhanced dependencies* multiplicam o número de informações anotadas sobre uma sentença, podendo beneficiar diversas aplicações que utilizam pré-processamento para melhorar seu desempenho.

Por isso, acredita-se que o conhecimento aqui descrito possa inspirar muitas ideias para todos que se dedicam ao PLN e ao PLN do português em especial.

Destaca-se que as *enhanced dependencies*, por se apoiarem na anotação sintática para gerar inferências, constituem uma ponte entre as funções sintáticas e os papéis semânticos.

Outro fato a ser observado é que esse tipo de anotação vem sendo implementado em várias línguas, aumentando o número de itens anotados com as mesmas etiquetas em várias línguas, o que pode funcionar como “interlíngua” em aplicações multilíngues.

As *enhanced dependencies* também podem ser utilizadas como base para o aumento do número de informações explícitas, uma possibilidade a ser explorada, por exemplo, por sistemas de simplificação textual, visando diminuir o número de inferências exigidas do leitor.

O importante, aqui, é salientar o potencial que as *enhanced dependencies* representam para o PLN e a importância de divulgar esse conhecimento, pois constituem um tipo de anotação não tão difundido quanto a anotação sintática e a anotação de papéis semânticos.

De forma imediata, este relatório traz o conhecimento necessário e quase sempre suficiente para a construção de programas de atribuição de *enhanced dependencies* em língua portuguesa. Os casos que não são totalmente resolvidos por regras, como o de inserção de *token* vazio para representar predicado elíptico e o de propagação de sujeito para as **advcl**, precisam ser revisados por humanos, o que pode dar um *feedback* para melhoria dos programas ou mesmo pistas para gerar um córpus que possa ser usado para a aprendizagem automática da tarefa.

Bibliografia

- Bechara, E. Moderna Gramática Portuguesa. Rio de Janeiro: Ed. Nova Fronteira, 2019.
- Duran, M. S.; Lopes, L.; Nunes, M.G.V.; Pardo, T.A.S. (2023). The Dawn of the Porttinari Multigenre Treebank: Introducing its Journalistic Portion. In the Proceedings of the 14th Symposium in Information and Human Language Technology (STIL), pp. 115-124. September, 25-29.
- Duran, M.S., & Ramisch, C. (2011). How do you feel? Investigating lexical-syntactic patterns in sentiment expression. In: Proceedings of Corpus Linguistics 2011: Discourse and Corpus Linguistics Conference. Disponível em: <https://www.semanticscholar.org/paper/How-do-you-feel-Investigating-lexical-syntactic>
- Duran, M. S.; Lopes, L.; Pardo, T. A. S.; Nunes, M. G. V. (2024). Construções comparativas em português e sua anotação usando a sintaxe de dependências. Revista Abralin (no prelo).
- Fillmore, Charles J. (1967) The case for case. DOI:10.1016/0141-9331(86)90222-x. Disponível em: <https://api.semanticscholar.org/CorpusID:61363578>
- Gosse Bouma, Djamel Seddah, and Daniel Zeman. (2020) Overview of the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies. In Proceedings of the 16th International Conference on Parsing Technologies and the IWPT 2020 Shared Task on Parsing into Enhanced Universal Dependencies, pages 151–161, Online. Association for Computational Linguistics.
- de Marneffe, M.-C.; Manning, C.D.; Nivre, J.; Zeman, D. (2021) Universal Dependencies. Computational Linguistics 47(2), 255-308.
- Mateus, M. H. M. et. al. Gramática da Língua Portuguesa. Lisboa, Ed. Caminho, 2003.
- Nivre, J.; Marongiu, P.; Ginter, F.; Kanerva, J.; Montemagni, S.; Schuster, S.; Simi, M. (2018) Enhancing Universal Dependency Treebanks: A Case Study. In Proceedings of the Second Workshop on Universal Dependencies, pages 102-107.
- Pagano, A. S.; Duran, M. S.; Pardo, T. A. S. (2023) *Enhanced dependencies* para o português brasileiro. In: Proceedings of the 2nd Edition of the Universal Dependencies Brazilian Festival, p. 461–470, Belo Horizonte, Brasil. Association for Computational Linguistics.
- Pardo, T. A. S.; Duran, M. S.; Lopes, L.; Felippo, A.; Roman, N. T. ; Nunes, M. G. V. (2021) Porttinari - a Large Multi-genre Treebank for Brazilian Portuguese. In: Simpósio Brasileiro de Tecnologia da Informação e da Linguagem Humana (STIL), 13. , 2021, Evento Online. Anais [...]. Porto Alegre: Sociedade Brasileira de Computação, 2021 . p. 1-10.
- Schuster, S.; Manning, C. D. (2016). Enhanced English Universal Dependencies: An Improved Representation for Natural Language Understanding Tasks. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), pages 2371–2378, Portorož, Slovenia. European Language Resources Association (ELRA).