

Documentação UD em português (e para língua portuguesa)

Elvis de Souza

PUC-Rio, Brasil

Tatiana Cavalcanti

Aline Silveira

Wograine Evelyn

Cláudia Freitas

O projeto Universal Dependencies ([mcdonald2013universal](#)) apresenta um tagset e uma gramática. Isso significa dizer que, para além de um conjunto de etiquetas que correspondem às classes da Gramática Tradicional (objeto, sujeito etc.), o UD também faz diversas escolhas que diferem da GT. Nesse documento, apresentamos a documentação detalhadas e as escolhas linguísticas relativas ao processo de revisão do material UD em Português. Considerando que UD funciona como uma espécie de segunda língua gramatical, partimos, sempre que possível, das categorias e análises de GT, e não de UD. Os exemplos de frases e as listas foram retirados do corpus [Bosque-UD](#) ([rademaker2017universal](#)) versão 2.5.

Conteúdo

Documentação UD em português

(e para língua portuguesa)

Elvis de Souza, Tatiana Cavalcanti, Aline Silveira, Wograinne Evelyn,
Cláudia Freitas

1

2 Formato UD

5

1	Princípios	5
2	Colunas/anotações	6
3	Manipulação em Python	6

3 Classes gramaticais (upos)

7

1	Verbos auxiliares	7
1.1	Verbos de ligação	8
1.2	Verbo <i>ser</i> como voz passiva	8
1.3	Locuções verbais de tempo composto	9
1.4	Locuções verbais aspectuais/modais não existem no UD	10
1.5	Proposta de anotação para as locuções verbais aspectu- ais/modais	10
1.6	Isso <i>foi</i> nos Estados Unidos: verbo <i>ser</i> como verbo pleno	12
1.7	Não são locuções verbais: dois verbos plenos	12
2	Numerais	13
2.1	Números compostos	13
2.2	<i>R\$ 50</i> : quem é o pai?	14
2.3	<i>50% das pessoas</i> : quem é o pai?	16
2.4	<i>Primeiro</i> lugar: adjetivo ou numeral?	16
3	Pronomes substantivos, pronomes adjetivos, interrogativos, arti- gos e determinantes	16

4 Atributos morfológicos (feats)

21

5 Dependências (dephead e deprel)

25

1	Locuções verbais	25
---	----------------------------	----

2	Estruturas comparativas	25
2.1	Frases do Working Group	25
2.2	Frases do Bosque-UD	26
3	Elementos discursivos	26
3.1	É que	26

2 Formato UD

[Ir para tabela de conteúdos](#)

Os treebanks adaptados para a gramática UD são disponibilizados no formato CoNLL-U, em que há um token por linha. Cada anotação de cada token, por sua vez, é disposta em uma coluna, sendo 10 colunas ao todo. Cada token tem a configuração conforme a [Tabela 1: Colunas do formato UD 2.0](#), com uma tabulação (*Tab*) separando as colunas. Colunas sem nenhum valor devem, necessariamente, ser preenchidas com *underline*, conforme diretivas da página do [formato CoNLL-U](#) (acesso em 28 de outubro de 2019).

Tabela 1: Colunas do formato UD 2.0

id	word	lemma	upos	xpos	feats	dephead	deprel	deps	misc
----	------	-------	------	------	-------	---------	--------	------	------

1 Princípios

De [Página do Universal Dependencies](#) (acesso em 28 de outubro de 2019).

What is needed for UD to be successful? The secret to understanding the design and current success of UD is to realize that the design is a very subtle compromise between approximately 6 things:

1. UD needs to be satisfactory on linguistic analysis grounds for individual languages.
2. UD needs to be good for linguistic typology, i.e., providing a suitable basis for bringing out cross-linguistic parallelism across languages and language families.
3. UD must be suitable for rapid, consistent annotation by a human annotator.
4. UD must be suitable for computer parsing with high accuracy.

5. UD must be easily comprehended and used by a non-linguist, whether a language learner or an engineer with prosaic needs for language processing. We refer to this as seeking a habitable design, and it leads us to favor traditional grammar notions and terminology.
6. UD must support well downstream language understanding tasks (relation extraction, reading comprehension, machine translation, ...).

2 Colunas/anotações

1. “id” corresponde ao número do token, em ordem crescente;
2. “word”, à palavra tal como aparece na frase (exceto no caso de contração, como “da”, em que a palavra será desmembrada nos tokens “de” e “a”);
3. “lemma” se refere à palavra tal como aparece no dicionário: em no singular e em masculino ou infinitivo;
4. “upos” (classe gramatical “universal”) se refere à classe gramatical;
5. No corpus Bosque-UD, a coluna “xpos” (classe gramatical específica) é preenchida com a saída do sistema PALAVRAS para a mesma frase;
6. “feats” (atributos morfológicos) é preenchida com as características morfológicas do token;
7. “dephead” (dependência sintática), com o id do token de quem é filho;
8. “deprel” (relação de dependência), com a relação sintática que o conecta ao seu pai;
9. “deps” (dependência específica) não é utilizado no Bosque-UD;
10. “misc” (miscelânea) se refere a quaisquer informações extras que desejemos adicionar ao token.

3 Manipulação em Python

Para manipular arquivos no formato UD em Python, com as classes `Corpus`, `Sentence` e `Token` (e suas respectivas anotações), desenvolvemos e utilizamos o `es-estrutura_ud.py`.

3 Classes gramaticais (upos)

Ir para tabela de conteúdos

Tabela 2: As classes gramaticais do UD em português

Tabela 2: As classes gramaticais do UD em português

upos	Observações
ADJ	adjetivos e numerais ordinais
ADP	preposições
PUNCT	pontuação
ADV	advérbio
AUX	Seção 1: Verbos auxiliares
SYM	símbolos
INTJ	interjeição
CCONJ	conjunção coordenativa
NOUN	substantivo
DET	determinante - artigos e pronomes adjetivos
PROPN	nomes próprios, apenas se com inicial maiúscula
NUM	numeral - exceto os ordinais, que são adjetivos
PART	partícula
VERB	verbo
PRON	apenas pronomes substantivos
SCONJ	conjunções subordinativas
X	no Bosque-UD, palavras estrangeiras

1 Verbos auxiliares

Verbos auxiliares são classificados como *AUX*. O que conta como um verbo auxiliar é alvo de discussão nas gramáticas do português (elvis2019locverbal). De modo geral, classificamos como verbos auxiliares os verbos de ligação (Subseção 1.1: Verbos de ligação), o verbo auxiliar na voz passiva (Subseção 1.2: Verbo

ser como voz passiva), as locuções de tempo composto (Subseção 1.3: Locuções verbais de tempo composto), e as locuções verbais aspectuais e modais. No UD, as locuções verbais aspectuais/modais não existem, ou seja, os verbos são anotados como se constituíssem duas orações (Subseção 1.4: Locuções verbais aspectuais/modais não existem no UD). No entanto, conforme apontado no trabalho supracitado, é possível propor uma anotação que leve em conta a auxiliaridade das locuções verbais aspectuais (Subseção 1.5: Proposta de anotação para as locuções verbais aspectuais/modais).

Via de regra, verbos auxiliares (*AUX*) não podem ter dependentes, e dependem de um verbo principal (*VERB*). O *deprel* pode ser *cop*, *aux* ou *aux:pass*, de acordo com a função sintática.

1.1 Verbos de ligação

Apenas os verbos “*ser*” e “*estar*” são considerados verbos de ligação, e portanto serão sempre anotados como *AUX*. Os demais verbos que a GT costuma elencar como verbo de ligação (“*parecer*”, “*permanecer*”, etc.) são anotados como *VERB*. Os verbos de ligação *AUX* terão relação sintática *cop*, e nunca poderão ser núcleo de uma oração (*X_s*) nem conter dependentes.

Figura 1: O preço é de US\$ 422

Tabela 4: Lista das 2 palavras que ocorrem 138 vezes como verbo de ligação (*AUX/cop*) no Bosque-UD

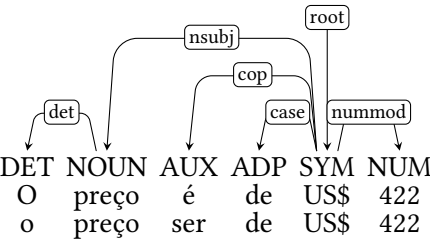


Figura 1: O preço é de US\$ 422

1.2 Verbo *ser* como voz passiva

A anotação de “*ser*” como voz passiva, além do upos *AUX*, deve receber *deprel aux:pass*.

Figura 2: A fotografia *foi* publicada na imprensa

Tabela 5: Lista das 2 palavras que ocorrem 1115 vezes como voz passiva (*AUX/aux:pass*) no Bosque-UD

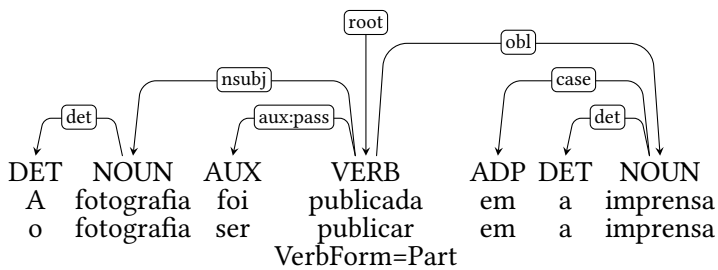


Figura 2: A fotografia *foi* publicada na imprensa

1.3 Locuções verbais de tempo composto

Segundo as gramáticas, são locuções verbais de tempo composto aquelas que têm como verbo auxiliar “ter”, “haver” e, para nós, também “ir”.

Figura 3: A Prefeitura não *havia retirado* o golfinho

Tabela 6: Lista das 2 palavras que ocorrem 1115 vezes como locução verbal de tempo composto (*AUX/aux*) no Bosque-UD

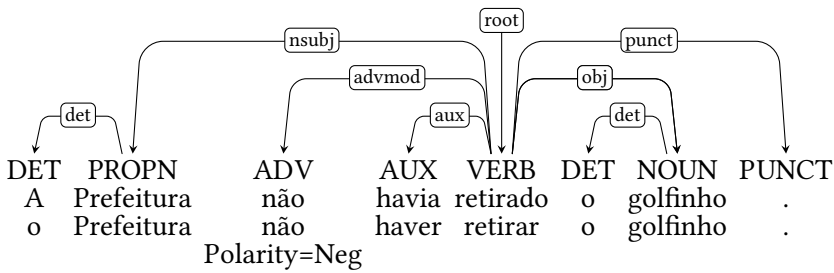


Figura 3: A Prefeitura não *havia retirado* o golfinho

1.4 Locuções verbais aspectuais/modais não existem no UD

Em UD, não existem locuções verbais aspectuais ou modais, de modo que as sentenças são anotadas como em **Figura 4: A seleção *deve* contar hoje com Giovane.**

Para entender melhor o fenômeno das locuções verbais e conhecer um modelo de anotação para o fenômeno, ver **Subseção 1.5: Proposta de anotação para as locuções verbais aspectuais/modais.**

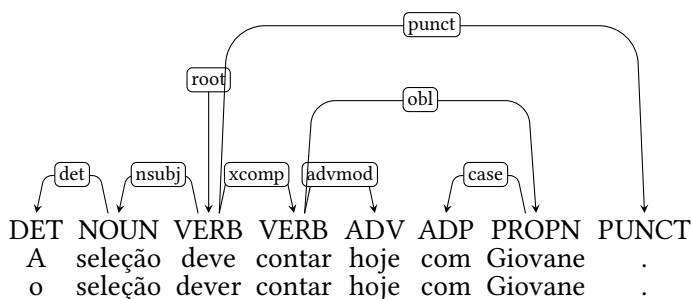


Figura 4: A seleção *deve* contar hoje com Giovane

1.5 Proposta de anotação para as locuções verbais aspectuais/modais

Em UD, não existem locuções verbais ou modais, mas estudamos uma alternativa para marcar a auxiliaridade dos verbos segundo alguns estudos gramaticais.

O que distingue a locução verbal aspectual da modal é que, na primeira, o verbo auxiliar perdeu o seu conteúdo lexical e serve apenas para caracterizar a temporalidade da ação do verbo principal, e na segunda, o verbo auxiliar (ou pleno, segundo algumas gramáticas) caracteriza um julgamento do enunciador sobre a ação do verbo principal.

Alguns gramáticos divergem sobre como a classificação se dá, portanto propomos apenas herdar, no Bosque-UD, o que já era a anotação originária do PALAVRAS (**bick2000parsing**), e não modificá-la em relação a quais devem ser os verbos auxiliares. Para o PALAVRAS (no Bosque-UD, até a versão 2.4), locuções verbais modais não são locuções, isto é, os dois verbos se configuram como verbos plenos (**Figura 5: A seleção *deve* contar hoje com Giovane.**).

Nos casos de locução verbal aspectual, sim, há uma locução, isto é, o primeiro verbo é auxiliar (*AUX/aux*), e o segundo, verbo pleno (*VERB*). Às vezes há uma partícula interveniente no meio da locução, como na **Figura 6: O Tribunal vai**

começar a ouvir as testemunhas, em que há um “a” entre “começar” e “ouvir”. Encaramos que estamos diante de um fenômeno de *phrasal verb*, uma MWE do tipo *AUX*, como sugerido em *elvis2019locverbal*. Na frase, *vai ouvir* é uma locução verbal de tempo composto (*Subseção 1.3: Locuções verbais de tempo composto*), e “começar a ouvir” é uma locução verbal aspectual, sendo “começar a” uma expressão multi-palavra (MWE) do tipo auxiliar.

Tabela 7: Lista das 31 MWEs que ocorreriam 570 vezes como locuções auxiliares (*AUX/aux + ADP/compound*) no Bosque-UD segundo nossa proposta de anotação

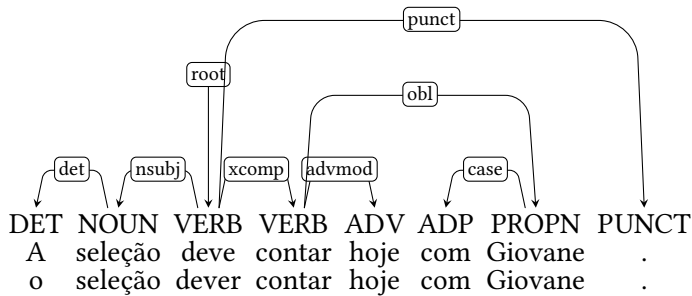


Figura 5: A seleção *deve contar* hoje com Giovane

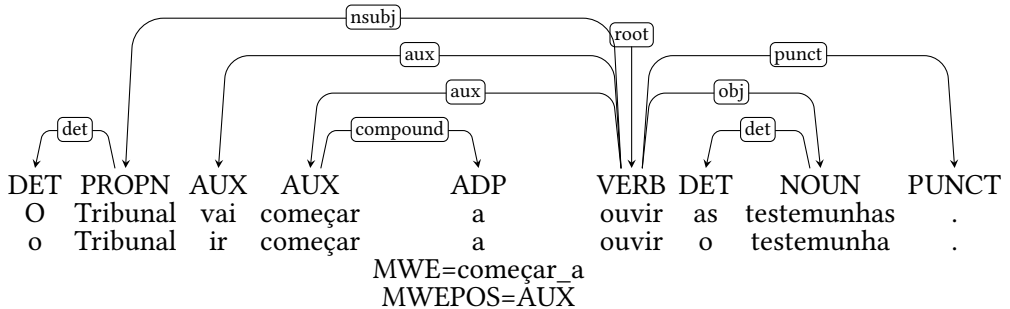


Figura 6: O Tribunal vai *começar a ouvir* as testemunhas

1.6 Isso foi nos Estados Unidos: verbo *ser* como verbo pleno

Atenção para 2 casos em que o “ser” não é verbo auxiliar e deve ser verbo pleno (upos *VERB*).

1) Como na **Figura 7: A expectativa era que chegasse a US\$7 milhões**, o “ser” deve ser núcleo da oração caso o predicado seja uma outra oração (*ccomp*, *xcomp*).

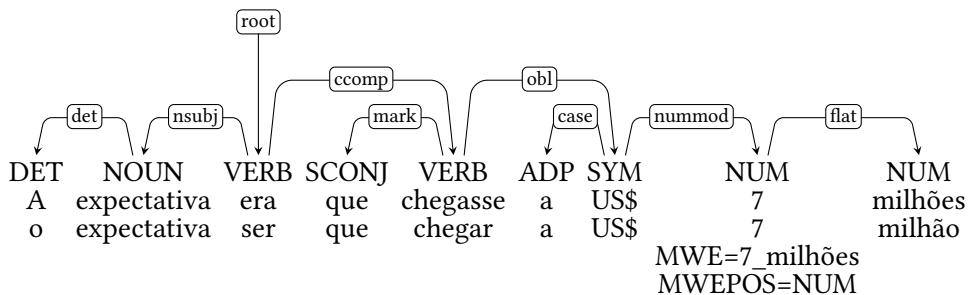


Figura 7: A expectativa era que chegasse a US\$7 milhões

2) “ser” verbo intransitivo (verbo pleno) também deve ter a anotação *VERB*, como na **Figura 8: Isso foi nos Estados Unidos**.

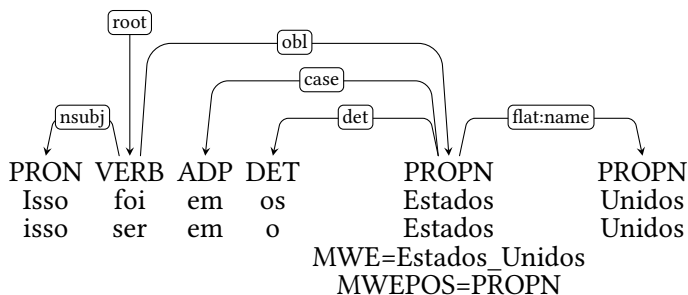


Figura 8: Isso foi nos Estados Unidos

1.7 Não são locuções verbais: dois verbos plenos

Ainda segundo as decisões do PALAVRAS (*bick2000parsing*), há outras tantas colocações de dois verbos que não são locuções verbais. Portanto, são todos ca-

sos de verbos plenos, de upos *VERB*, em que o segundo verbo é dependente do primeiro, como oração subordinada substantiva objetiva direta reduzida de infinitivo (Xx).

Figura 9: Muitos pintores *quiseram aprender a pintar* com ele

Tabela 8: Lista dos 196 verbos plenos que ocorrem 1150 vezes como *pais* de uma colocação verbal no Bosque-UD

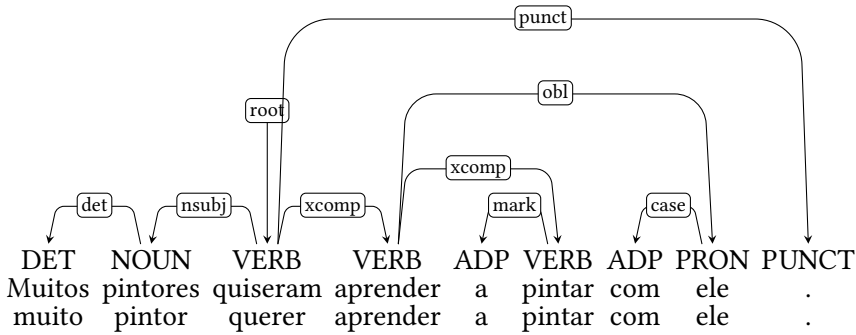


Figura 9: Muitos pintores *quiseram aprender a pintar* com ele

2 Numerais

Numerais cardinais (360) e fracionários (dois terços) devem ser anotados como de upos *NUM*, sejam por extenso ou em algarismos. Numerais multiplicativos (dobro, triplo), por outro lado, devem ser anotados como de upos *NUM*.

Numerais coletivos (como “dezenas”, “centenas”, “milhares”, etc.), no corpus Bosque-UD, podem estar anotados de duas formas diferentes: como *NUM*, caso seja um número exato (“duas centenas”, “dois bilhões”), ou como *NOUN*, quando é indefinido (“centenas de pessoas”).

Numerais ordinais devem ser anotados como *ADJ* (Subseção 2.4: *Primeiro lugar: adjetivo ou numeral?*).

Ver também: Subseção 2.1: *Números compostos*.

2.1 Números compostos

Números compostos podem conter upos *NUM*, *CCONJ* e *ADP*. Quando o número composto modifica um substantivo, o primeiro número deve se subordi-

nar ao substantivo como “nummod”; quando há um sintagma preposicionado, o primeiro número do número composto será o head, e o substantivo, *nmod* do número. Além disso, deve-se notar que o primeiro número deve receber no misc *MWE* e *MWEPOS*.

Figura 10: *Trinta e sete*

Figura 11: *Montante de 1300 milhões de dólares*

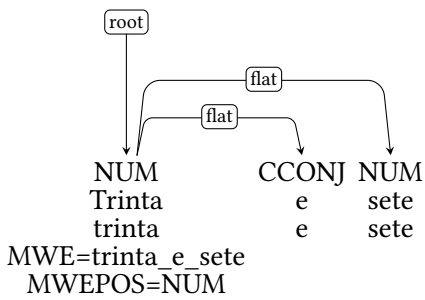


Figura 10: *Trinta e sete*

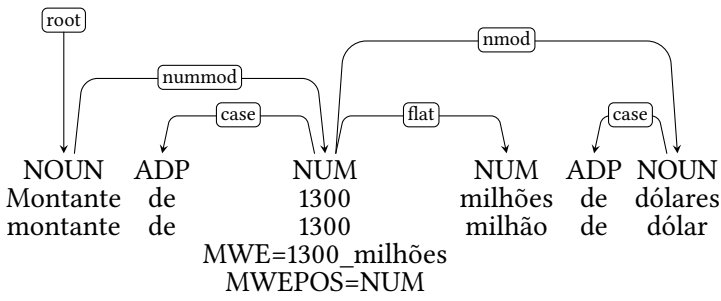


Figura 11: *Montante de 1300 milhões de dólares*

2.2 R\$ 50: quem é o pai?

Quando há a inserção de símbolos como *Figura 12: R\$ 50*, *Figura 13: US\$ 50* ou *Figura 14: 50%*, o símbolo deve ser o head. No caso de números coletivos (*Figura 15: 2 centenas*), o coletivo será o head.

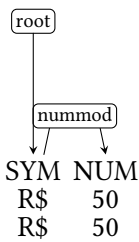


Figura 12: R\$ 50

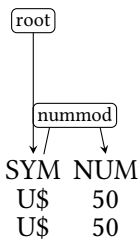


Figura 13: U\$ 50

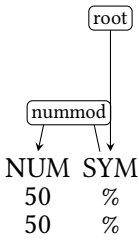


Figura 14: 50%

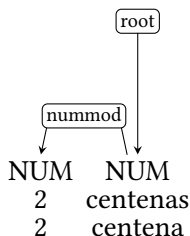


Figura 15: 2 centenas

2.3 50% das pessoas: quem é o pai?

Sempre que houver sintagma preposicionado, o sintagma anterior será o head, como em **Figura 16: 50% das pessoas** (onde o símbolo é o head, segundo **Subseção 2.2: R\$ 50: quem é o pai?**).

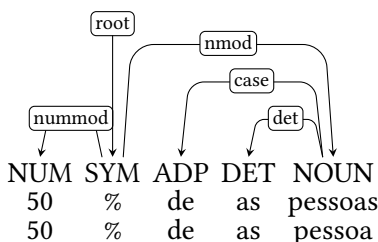


Figura 16: 50% das pessoas

2.4 Primeiro lugar: adjetivo ou numeral?

Numerais ordinais escritos por extenso devem ser anotados como *ADJ*, e recebem a feature *NumType=Ord*, como na **Figura 17: Primeira tentativa**.

3 Pronomes substantivos, pronomes adjetivos, interrogativos, artigos e determinantes

Pronomes substantivos são anotados como de upos *PRON* e são núcleo do sintagma de que fazem parte, recebendo a *deprel* do sintagma (**Figura 18: Talvez isto**

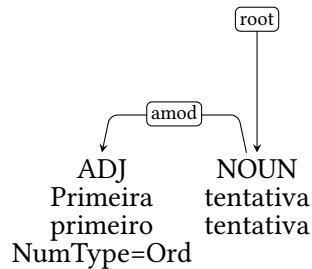


Figura 17: Primeira tentativa

seja muito barulho por nada).

Pronomes adjetivos são anotados como de upos *DET*, e recebem *deprel det*, sendo dependentes do token que modificam (Figura 19: *Aquele pensamento provocou-me um arrepio delicioso*).

Pronomes interrogativos, assim como os pronomes substantivos, têm upos *PRON* e recebem o *deprel* do sintagma (Figura 20: *E sabe quem eu ainda não vi?*). Não confundir com os advérbios interrogativos (*Xxx*), que podem inclusive se tornar conjunções quando introduzindo orações subordinadas (*Xxx*).

Artigos/determinantes têm anotação similar à dos pronomes adjetivos (*DET/det*), como na Figura 21: *Eram o retrato do cérebro deste partido*.

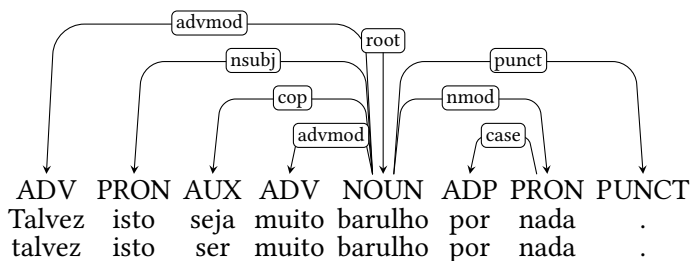


Figura 18: Talvez isto seja muito barulho por nada

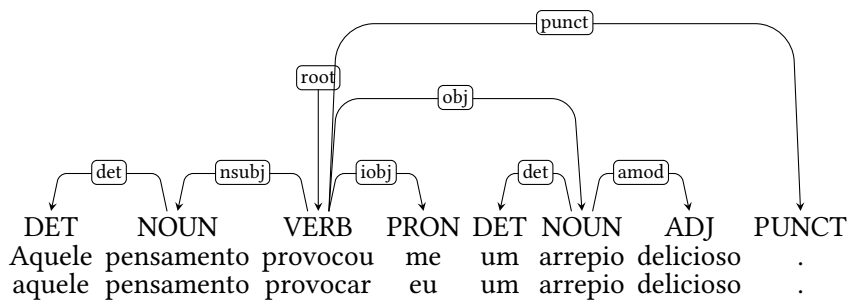


Figura 19: *Aquele* pensamento provocou-me um arrepio delicioso

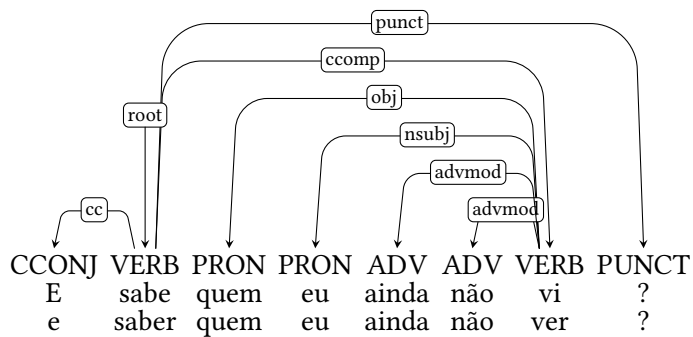


Figura 20: E sabe *quem* eu ainda não vi?

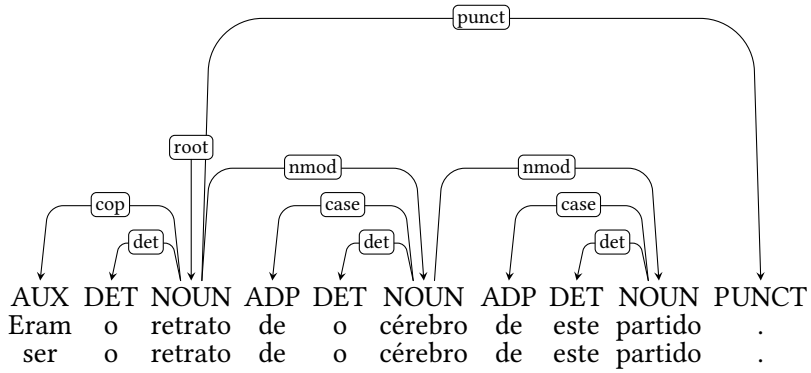


Figura 21: Eram o retrato do cérebro deste partido

4 Atributos morfológicos (feats)

Ir para tabela de conteúdos

Temos a seguinte distribuição de atributos morfológicos por classe gramatical (**Tabela 3: Atributos morfológicos (feats)**). É importante notar que os atributos morfológicos devem constar em ordem alfabética e são separados por uma barra reta.

upos	features
ADJ	Gender=[Fem, Masc, Unsp] NumType=[Ord] Number=[Plur, Sing]
ADP	–
ADV	Polarity=[Neg] –
AUX	Gender=[Fem, Masc] Mood=[Cnd, Imp, Ind, Sub] Number=[Plur, Sing] Person=[1, 2, 3] Tense=[Fut, Imp, Past, Pqp, Pres] VerbForm=[Fin, Ger, Inf, Part]
CCONJ	–
DET	Definite=[Def, Ind] Gender=[Fem, Masc, Unsp] Number=[Plur, Sing, Unsp] PronType=[Art, Dem, Emp, Ind, Int, Neg, Prs, Rel, Tot]

INTJ	–
NOUN	Foreign=[Yes] Gender=[Fem, Masc, Unsp] NumType=[Ord] Number=[Plur, Sing, Unsp]
NUM	Gender=[Fem, Masc, Unsp] NumType=[Card, Frac, Mult, Ord, Range, Sets] Number=[Plur, Sing]
PART	Gender=[Masc] Number=[Sing]
PRON	Case=[Acc, Dat, Nom] Definite=[Def, Ind] Gender=[Fem, Masc, Unsp] Number=[Plur, Sing, Unsp] Person=[1, 2, 3] PronType=[Art, Dem, Ind, Int, Neg, Prs, Rel, Tot] Reflex=[Yes] VerbForm=[Ger]
PROPN	Gender=[Fem, Masc, Unsp] Number=[Plur, Sing]
PUNCT	–
SCONJ	Gender=[Fem, Masc] Number=[Plur, Sing] PronType=[Ind, Rel]
SYM	–

VERB	Gender=[Fem, Masc] Mood=[Cnd, Imp, Ind, Sub] Number=[Plur, Sing] Person=[1, 2, 3] Tense=[Fut, Imp, Past, Pqp, Pres] VerbForm=[Fin, Ger, Inf, Part] Voice=[Pass]
------	---

X	—
---	---

5 Dependências (dephead e deprel)

Ir para tabela de conteúdos

1 Locuções verbais

Ver [Seção 1: Verbos auxiliares](#).

2 Estruturas comparativas

Estruturas comparativas são de anotação complexa, o que se verifica pela existência de um **working group (WG) em UD** dedicado especialmente a elas. A seguir, listamos as frases utilizadas no WG, traduzidas em português, e com a anotação adequada, além de algumas frases de anotação complexa no Bosque-UD.

2.1 Frases do Working Group

8	é	ser	AUX	-	Mood=Ind Number=Sing Person=3 Tense=Pres VerbForm=Fin	9	cop	-	-
9	essencial	essencial	ADJ	-	Gender=Masc Number=Sing	1	ccomp	-	-
10	para	para	ADP	-	11	case	-	-	-
11	funções	função	NOUN	-	Gender=Fem Number=Plur	9	obl	-	-
12	mentais	mental	ADJ	-	Gender=Fem Number=Plur	11	amod	-	-
13	de	de	ADP	-	15	case	-	-	-
14	altíssimo	alto	ADJ	-	Gender=Masc Number=Sing	15	amod	-	-
15	nível	nível	NOUN	-	Gender=Masc Number=Sing	11	nmod	-	SpaceAfter=No
16	,	,	PUNCT	-	20	punct	-	-	-
17	tais	tal	DET	-	Gender=Fem Number=Plur PronType=Dem	20	case	-	MWEPOS=ADP
18	como	como	ADP	-	17	fixed	-	-	-
19	o	o	DET	-	Definite=Def Gender=Masc Number=Sing PronType=Art	20	-	-	-
20	raciocínio	raciocínio	NOUN	-	Gender=Masc Number=Sing	11	appos:parataxis	-	SpaceAfter=No
21	,	,	PUNCT	-	23	punct	-	-	-
22	a	o	DET	-	Definite=Def Gender=Fem Number=Sing PronType=Art	23	det	-	-
23	memória	memória	NOUN	-	Gender=Fem Number=Sing	20	conj	-	SpaceAfter=No
24	,	,	PUNCT	-	26	punct	-	-	-
25	o	o	DET	-	Definite=Def Gender=Masc Number=Sing PronType=Art	26	det	-	-
26	reconhecimento	reconhecimento	NOUN	-	Gender=Masc Number=Sing	20	conj	-	-
27-28	dos	de	ADP	-	29	case	-	-	-
27	os	o	DET	-	Definite=Def Gender=Masc Number=Plur PronType=Art	29	det	-	-
28	rostos	rosto	NOUN	-	Gender=Masc Number=Plur	26	nmod	-	SpaceAfter=No
29	,	,	PUNCT	-	33	punct	-	-	-
30	das	de	ADP	-	33	case	-	-	-
31-32	de	o	DET	-	Definite=Def Gender=Fem Number=Plur PronType=Art	33	det	-	-
31	as	o	DET	-	Gender=Fem Number=Plur	29	conj	-	-
32	melodias	melodia	NOUN	-	Gender=Fem Number=Plur	33	amod	-	SpaceAfter=No
33	musicais	musical	ADJ	-	29	punct	-	-	-
34	,	,	PUNCT	-	20	conj	-	-	SpaceAfter=No
35	etc.	etc.	ADV	-	-	-	-	-	-
36				-	-	-	-	-	-

Figura 22: Eu coloquei *tanta farinha quanto* a receita pedia

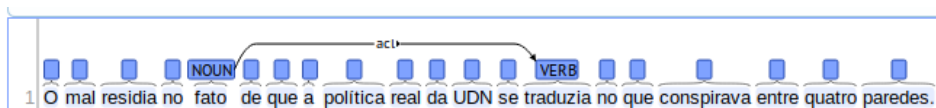


Figura 23: Martin é o cara *mais inteligente de todos*

2.2 Frases do Bosque-UD

3 Elementos discursivos

3.1 É que

A construção “é que” pode aparecer de duas maneiras no corpus: marcada ou não como MWE. Como MWE, “é que” é uma ocorrência típica do discurso oral, sem ligação clara com a estrutura da oração. Por outro lado, “é” e “que” são independentes entre si quando introduzem uma oração subordinada substantiva predicativa ou subjetiva. (ver Xx)

Quando “é que” tem esse efeito discursivo, anota-se da seguinte maneira:

É: tem upos *AUX* e deprel *discourse*, e aponta para a head da oração;

Que: tem upos *SCONJ* e deprel *fixed*, e aponta para “é”.

Figura 24: Só depois *é que* levanto a cabeça para fazer um lançamento

Figura 25: A melhor equipe *é que* deve jogar

Figura 26: *É que* são uns antipáticos que nunca oferecem copos

OBS.: Consideramos “Não é que” como MWE.

“*Não é que* o sábio das matrizes encontrou, no PSD, entusiastas seguidores?”

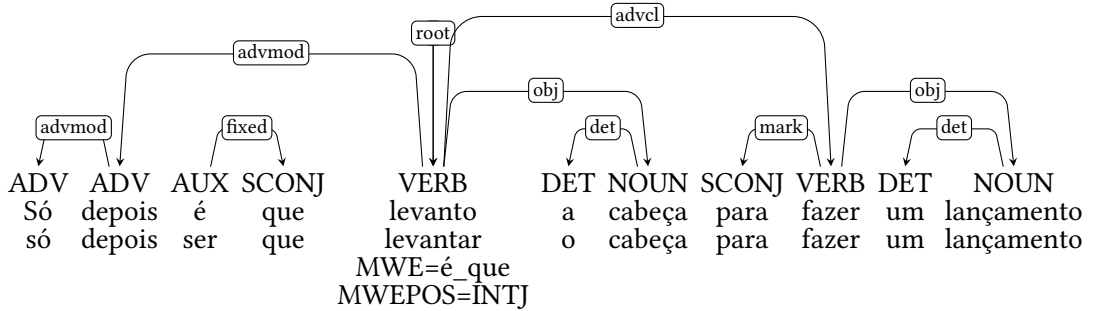


Figura 24: Só depois é que levanto a cabeça para fazer um lançamento

```
# text = A melhor equipe é que deve jogar.
1  A      o      DET      -      Definite=Def|Gender=Fem|Number=Sing|PronType=Art      3      det      -      -
2  melhor bom    ADJ      -      Gender=Fem|Number=Sing      3      amod      -      -
3  equipe equipe NOUN     -      Gender=Fem|Number=Sing      4      nsubj     -      -
4  { é }      é      AUX      -      Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin      7      discourse -      MWE=é que
5  { que }    que     SCONJ   -      -      4      fixed     -      -
6  deve      dever  AUX      -      Mood=Ind|Number=Sing|Person=3|Tense=Pres|VerbForm=Fin      7      aux      -      -
7  jogar     jogar  VERB     -      VerbForm=Inf      0      root      -      SpaceAfter=No
8  .          .      PUNCT    -      -      4      punct     -      SpaceAfter=No
```

Figura 25: A melhor equipe é que deve jogar

```
# text = É que são uns antipáticos que nunca oferecem copos.
1  { É }      é      AUX      -      5      discourse -      MWE=É que
2  { que }    que     SCONJ   -      Gender=Masc|Number=Sing      1      fixed     -      -
3  são      ser      AUX      -      Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin      5      cop      -      -
4  uns      um      DET      -      Gender=Masc|Number=Plur|PronType=Art      5      det      -      -
5  antipáticos antipático ADJ      -      Gender=Masc|Number=Plur      0      root      -      -
6  que      que     SCONJ   -      8      mark      -      -
7  nunca    nunca  ADV      -      8      advmod    -      -
8  oferecem oferecer  VERB     -      Mood=Ind|Number=Plur|Person=3|Tense=Pres|VerbForm=Fin      5      acl:relcl
9  copos    copo    NOUN     -      Gender=Masc|Number=Plur      8      obj      -      -
10 .         .      PUNCT    -      5      punct     -      SpaceAfter=No
```

Figura 26: É que são uns antipáticos que nunca oferecem copos

Apêndice

Ir para tabela de conteúdos

Todas as listas são retiradas do corpus Bosque-UD em sua versão 2.5.

Tabela 4: Lista das 2 palavras que ocorrem 138 vezes como verbo de ligação (*AUX/cop*) no Bosque-UD

Verbo de ligação	Frequência
estar	33
ser	105

Tabela 5: Lista das 2 palavras que ocorrem 1115 vezes como voz passiva (*AUX/aux:pass*) no Bosque-UD

Voz passiva	Frequência
ficar	1
ser	1114

Tabela 6: Lista das 2 palavras que ocorrem 1115 vezes como locução verbal de tempo composto (*AUX/aux*) no Bosque-UD

Voz passiva	Frequência
ficar	1
ser	1114

Primeiro verbo (ordem alfabética)	#	Primeiro verbo (ordem de frequência)	#
abster	2	querer	108
acabar	3	conseguir	83
aceder	1	tentar	64

aceitar	6	fazer	45
achar	5	pretender	39
aconselhar	5	permitir	37
acreditar	2	decidir	33
acusar	30	acusar	30
admitir	12	deixar	27
afirmar	13	procurar	21
aguentar	1	levar	20
ajudar	9	ver	19
alegar	3	dar	18
ambicionar	1	gostar	17
ameaçar	14	obrigar	17
andar	2	ser	17
anunciar	2	saber	16
aperceber	1	ameaçar	14
apetecer	1	precisar	14
aplicar	1	preferir	14
apostar	1	resolver	14
aprender	5	afirmar	13
apresentar	2	admitir	12
apressar	1	considerar	12
aprestar	1	encontrar	12
aproveitar	1	ter	12
arriscar	2	dizer	11
atender	1	destinar	10
atrever	1	limitar	10
autorizar	3	ajudar	9
avisar	1	ficar	9
bastar	3	preparar	9
caber	1	esperar	8
cansar	1	pensar	8
cessar	1	comprometer	7
chamar	4	consistir	7
citar	1	continuar	7
começar	1	convidar	7
comprometer	7	recusar	7

*Documentação UD em português
(e para língua portuguesa)*

concordar	3	aceitar	6
condenar	1	convencer	6
conduzir	1	estar	6
confidenciar	1	manter	6
confirmar	1	prometer	6
conseguir	83	achar	5
considerar	12	aconselhar	5
consistir	7	aprender	5
consultar	1	forçar	5
contar	1	mandar	5
continuar	7	parecer	5
contribuir	3	passar	5
convencer	6	propor	5
convencionar	1	tender	5
convidar	7	visar	5
credenciar	1	chamar	4
criticar	1	desejar	4
culpar	1	encarregar	4
dar	18	evitar	4
decidir	33	impedir	4
declarar	3	interessar	4
declinar	1	proibir	4
dedicar	2	reconhecer	4
deixar	27	sentir	4
depender	1	tencionar	4
desafiar	2	tratar	4
desejar	4	acabar	3
desistir	1	alegar	3
destinar	10	autorizar	3
dever	3	bastar	3
dispor	3	concordar	3
dizer	11	contribuir	3
duvidar	1	declarar	3
empenhar	2	dever	3
encarregar	4	dispor	3
encontrar	12	insistir	3

ensinar	1	ir	3
equivaler	1	optar	3
escolher	1	ousar	3
escusar	2	pôr	3
esperar	8	vir	3
estar	6	abster	2
estimar	1	acreditar	2
estimular	1	andar	2
estudar	1	anunciar	2
evitar	4	apresentar	2
exigir	1	arriscar	2
exortar	1	dedicar	2
experimentar	1	desafiar	2
falar	2	empenhar	2
fartar	1	escusar	2
fazer	45	falar	2
ficar	9	garantir	2
fingir	1	haver	2
foi	1	imaginar	2
forçar	5	impor	2
garantir	2	importar	2
gostar	17	incentivar	2
habituar	1	mostrar	2
haver	2	necessitar	2
imaginar	2	negar	2
impedir	4	ouvir	2
impor	2	parar	2
importar	2	queixar	2
incentivar	2	receber	2
incluir	1	referir	2
indagar	1	resistir	2
influenciar	1	sentar	2
informar	1	tornar	2
insinuar	1	aceder	1
insistir	3	aguentar	1
instar	1	ambicionar	1

*Documentação UD em português
(e para língua portuguesa)*

interessar	4	aperceber	1
ir	3	apetecer	1
lembrar	1	aplicar	1
levar	20	apostar	1
limitar	10	apressar	1
mandar	5	aprestar	1
manter	6	aproveitar	1
merecer	1	atender	1
mostrar	2	atrever	1
necessitar	2	avisar	1
negar	2	caber	1
obrigar	17	cansar	1
optar	3	cessar	1
orgulhar	1	citar	1
ousar	3	começar	1
ouvir	2	condenar	1
parar	2	conduzir	1
parecer	5	confidenciar	1
passado	1	confirmar	1
passar	5	consultar	1
pedir	1	contar	1
pensar	8	convencionar	1
perder	1	credenciar	1
permanecer	1	criticar	1
permitir	37	culpar	1
persuadir	1	declinar	1
planear	1	depende	1
poder	1	desistir	1
precisar	14	duvidar	1
preferir	14	ensinar	1
preparar	9	equivaler	1
pretender	39	escolher	1
prevenir	1	estimar	1
prever	1	estimular	1
procurar	21	estudar	1
proibir	4	exigir	1

projectar	1	exortar	1
prometer	6	experimentar	1
propiciar	1	fartar	1
propor	5	fingir	1
provocar	1	foi	1
pôr	3	habituar	1
queixar	2	incluir	1
querer	108	indagar	1
receber	2	influenciar	1
reclamar	1	informar	1
recomeçar	1	insinuar	1
recompensar	1	instar	1
reconhecer	4	lembrar	1
recordar	1	merecer	1
recusar	7	orgulhar	1
reduzir	1	passado	1
referir	2	pedir	1
resistir	2	perder	1
resolver	14	permanecer	1
respeitar	1	persuadir	1
restar	1	planear	1
retirar	1	poder	1
saber	16	prevenir	1
salientar	1	prever	1
sentar	2	projectar	1
sentir	4	propiciar	1
ser	17	provocar	1
soar	1	reclamar	1
sonhar	1	recomeçar	1
sublinhar	1	recompensar	1
sujeitar	1	recordar	1
suportar	1	reduzir	1
surgir	1	respeitar	1
suspeitar	1	restar	1
tardar	1	retirar	1
teimar	1	salientar	1

*Documentação UD em português
(e para língua portuguesa)*

temer	1	soar	1
tencionar	4	sonhar	1
tender	5	sublinhar	1
tentar	64	sujeitar	1
ter	12	suportar	1
tornar	2	surgir	1
tratar	4	suspeitar	1
trazer	1	tardar	1
ver	19	teimar	1
vir	3	temer	1
virar	1	trazer	1
visar	5	virar	1
voltar	1	voltar	1

Tabela 8: Lista dos 196 verbos plenos que ocorrem 1150 vezes como *pais* de uma colocação verbal no Bosque-UD

Tabela 7: Lista das 31 MWEs que ocorreriam 570 vezes como locuções auxiliares (*AUX/aux + ADP/compound*) no Bosque-UD segundo nossa proposta de anotação

MWE auxiliar	Frequência
acabar de	11
acabar por	30
andar a	3
chegar a	22
começar a	58
começar por	6
continuar a	57
continuar por	1
deixar de	30
dever a	1
estar a	124
estar para	1
estar por	1
ficar a	4
ficar de	1
haver a	1
haver de	2
haver que	1
ir a	3
ir de	1
parar de	3
passar a	43
poder a	3
ser de	3
tender a	1
ter a	9
ter de	62
ter que	3
tornar a	1
vir a	42
voltar a	42

MWE auxiliar	Frequência
estar a	124
ter de	62
começar a	58
continuar a	57
passar a	43
vir a	42
voltar a	42
acabar por	30
deixar de	30
chegar a	22
acabar de	11
ter a	9
começar por	6
ficar a	4
andar a	3
ir a	3
parar de	3
poder a	3
ser de	3
ter que	3
haver de	2
continuar por	1
dever a	1
estar para	1
estar por	1
ficar de	1
haver a	1
haver que	1
ir de	1
tender a	1
tornar a	1

Abreviações

[Ir para tabela de conteúdos](#)

Agradecimentos

[Ir para tabela de conteúdos](#)