
**DIRETRIZES DE ANOTAÇÃO DE RELAÇÕES DE DEPENDÊNCIA
EM *TWEETS* DO MERCADO FINANCEIRO**

ARIANI DI-FELIPPO
MARIA DAS GRAÇAS VOLPE NUNES
BRYAN KHELVEN DA SILVA BARBOSA

Nº 446

RELATÓRIOS TÉCNICOS



São Carlos – SP
Abr./2024

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

Diretrizes de Anotação de Relações de Dependência em *Tweets* do Mercado Financeiro

Orientações para anotação de relações de dependência sintática em *tweets* em língua portuguesa segundo a abordagem *Universal Dependencies* (UD)

Ariani Di-Felippo, Maria das Graças Volpe Nunes e Bryan Khelven da Silva
Barbosa

Abril/2024

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

Agradecimentos

À **Magali Sanchez Duran**, por todo o suporte, sempre atencioso, rigoroso e acurado, para a compreensão do nível sintático do modelo *Universal Dependencies* (UD) e para a aplicação dele no tratamento dos *tweets*. Esse suporte foi imperativo para a confecção desta primeira versão de um Manual de anotação das dependências previstas pelo modelo em *tweets* do mercado financeiro em língua português.

Este trabalho foi apoiado pelo Ministério da Ciência, Tecnologia e Inovações, com recursos da Lei N. 8.248, de 23 de outubro de 1991, no âmbito do PPI-Softex, coordenado pela Softex e publicado como Residência em TIC 13, DOU 01245.010222/2022-44.

Este trabalho foi executado no Centro de Inteligência Artificial (C4AI-USP) com apoio da Fundação de Apoio à Pesquisa do Estado de São Paulo (processo FAPESP #2019/07665-4) e da IBM Corporation.

SUMÁRIO

| | |
|---|----|
| Introdução | 5 |
| Diretrizes de anotação de dependências para <i>tweets</i> | 6 |
| PRIMEIRA PARTE – IDIOSSINCRASIAS | 8 |
| Fenômenos GCU | 8 |
| 1. RT: marca de <i>retweet</i> | 8 |
| 2. Menção | 10 |
| 2.1. Root: raiz | 10 |
| 2.2. Nmod: modificador nominal | 10 |
| 2.3. Vocative: vocativo | 11 |
| 2.4. Obl: nominal oblíquo | 13 |
| 2.5. Obj: objeto direto | 13 |
| 2.6. Nsubj: sujeito | 14 |
| 3. URL: endereço <i>web</i> | 16 |
| 4. Hashtag | 19 |
| 4.1. Root: raiz | 19 |
| 4.2. Nmod: modificador nominal | 19 |
| 4.3. Nsubj: sujeito | 20 |
| 4.4. Obj: objeto direto | 20 |
| 4.5. Obl: nominal oblíquo | 21 |
| 4.6. Appos: modificador apositivo | 21 |
| 4.7. Vocative: vocativo | 22 |
| 5. Cashtag | 24 |
| 5.1. Root: raiz | 24 |
| 5.2. Nmod: modificador nominal | 25 |
| 5.3. Nsubj: sujeito | 25 |
| 5.4. Obj: objeto direto | 25 |
| 2.6. Truncamento..... | 26 |
| 2.7. Índice de (des-)valorização das ações | 28 |
| 2.8. Emoticon e emoji | 29 |
| 2.9. Onomatopeias | 30 |
| 2.10. Repetição de sinal de pontuação..... | 31 |
| 2.11. Substituição lexical por símbolo (SYM) | 31 |
| Sintaxe não-padrão | 32 |
| Sujeito (nsubj) separado do predicado (<i>root</i>) por dois pontos | 32 |
| Advcl sem predicado..... | 33 |
| Coordenação (conj) introduzida por pontuação ou símbolo | 34 |

| | |
|--|----|
| Aposição (appos) sinalizada por PUNCT “/” (barra inclinada)..... | 36 |
| Root expresso por símbolo (SYM) | 36 |
| <i>Deprel</i> por inferência..... | 37 |
| Inferência de conjunção (conj)..... | 37 |
| Inferência de conjunção e verbo | 38 |
| Inferência de verbo e preposição..... | 39 |
| Inferências de símbolo como fórmula ou palavra de conteúdo | 40 |
| SEGUNDA PARTE - PADRÕES ESTRUTURAIS..... | 43 |
| Template 1 | 43 |
| Template 2 | 46 |
| Template 3 | 47 |
| Template 4 | 49 |
| Template 5 | 50 |
| Template 6 | 51 |
| Template 7 | 52 |
| Template 8 | 54 |
| Template 9 | 55 |
| Template 10 | 55 |
| Template 11 | 56 |
| Template 12 | 57 |
| Template 13 | 58 |
| Template 14 | 60 |
| Template 15 | 61 |
| Template 16 | 63 |
| Template 17 | 64 |
| Template 18 | 65 |
| Template 19 | 65 |
| Template 20 | 66 |
| Template 21 | 67 |
| Template 22 | 68 |
| Bibliografia | 69 |

Introdução

Apresenta-se neste relatório o Manual de Anotação de Relações de Dependência Sintática para *Tweets*¹ do Mercado Financeiro em Português desenvolvido no projeto POeTiSA (*PORtuguese processing - Towards Syntactic Analysis and parsing*), que faz parte da iniciativa de Processamento Automático 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 n.º. 2019/07665-4).

O POeTiSA é um projeto que visa aumentar os recursos baseados em sintaxe e desenvolver ferramentas e aplicações de PLN relacionadas à sintaxe para o português do Brasil e que alcancem o estado-da-arte. Esse objetivo inclui a produção de um *corpus* multigênero extenso e abrangente anotado segundo o modelo *Universal Dependencies* (UD) (NIVRE, 2015; NIVRE et al., 2020). Para tanto, dá-se continuidade a esforços anteriores, como os de construção da Floresta Sintá(c)tica (AFONSO et al., 2002), do MAC-MORPHO (ALUÍSIO et al., 2003) e do Bosque-UD (RADEMAKER et al., 2017). O *corpus* multigênero, chamado Portinari² (de *PORtuguese Treebank*) (PARDO et al., 2021), deve subsidiar estudos linguísticos e/ou o desenvolvimento de ferramentas de análise textual para o português, como *taggers* e *parsers*.

O modelo UD possui o formato de anotação CoNLL-U, constituído de 10 colunas, algumas das quais pedem decisões de anotação. Este Manual estabelece diretrizes de anotação de relações de dependência em *tweets* do mercado financeiro. Essa tarefa envolve decidir (i) os participantes da relação de dependência, (ii) o *head* e o dependente e (iii) o nome da relação que os liga. As colunas do CoNLL-U envolvidas pelas diretrizes são a 7ª, a 8ª e a 9ª. Este Manual complementa o Manual de Anotação de PoS Tags de PoS Tags em *Tweet*, publicado em março de 2022 na série de Relatórios Técnicos do ICMC sob número 438 e disponível na página do POeTiSA³. O Manual de *PoS tags* contempla a 4ª coluna do CoNLL-U. A divisão das diretrizes em 2 manuais se deve à decisão de revisar as colunas de anotação por etapas, a fim de que uma etapa pudesse abreviar o esforço requerido na outra, pois a tarefa de revisar *PoS tags* e relações de dependência, separadamente, já é bastante complexa.

Este manual reúne diretrizes para anotação de *tweets* do mercado financeiro, sendo que a anotação das estruturas gerais do português presentes nos exemplos segue o Manual de Duran (2022)⁴, proposto para a anotação de textos que possuem linguagem formal (ou padrão). Por isso, aliás, sugere-se o estudo do manual de Duran (2022) antes da leitura deste. Os exemplos deste documento foram extraídos de uma versão revisada do *corpus* de Silva et al. (2020) que já possui anotação de PoS segundo a UD (denominada DANTEStocks) e está disponível na página do POeTiSA⁵. O DANTEStocks contém *tweets* que mencionam ao menos um *ticker* de uma das 73 ações que compõem o Ibovespa (principal indicador da B3, que é a bolsa de valores oficial do Brasil), sendo que nele o *tweet* é a unidade básica de análise, uma vez que as postagens não passaram por qualquer segmentação, a não ser a tokenização. Os *tweets* do DANTEStocks também não passaram por um processo de normalização, contendo, por isso, todas as peculiaridades da linguagem informal dos *tweets* e do domínio do mercado financeiro (cf. Di-Felippo et al., 2021).

¹ Embora a plataforma tenha sido renomeada para “X” e as mensagens nela circulantes para “posts” após a aquisição da marca por Elon Musk e consequente reestruturação ocorrida em 2022, optou-se por utilizar as denominações originais (“Twitter” para plataforma e “tweet” para mensagem/postagem) em concordância com a época (2014) em que o *corpus* aqui utilizado foi compilado.

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

³ https://drive.google.com/file/d/1ka-GVNb8XgEJWmBOrNcd-Grfg10I_QEb/view?usp=sharing

⁴ <https://drive.google.com/file/d/1ile8Wfxu1qdrZOmLGqkvVuQ4fXvHgVMo/view?usp=sharing>

⁵ <https://drive.google.com/file/d/1wr9M4czkPgkUj1--U9GT9h8ncXc6rvz4/view?usp=sharing>

Diretrizes de anotação de dependências para tweets

O esquema de anotação do modelo *Universal Dependencies* (UD), em sua versão 2.0, disponibiliza 37 tags de relações de dependência ou *deprel* (do inglês, **dependency relation**).

Na Figura 1, extraída das *Guidelines* da UD⁶, exibem-se as 37 *deprel* da UD. A Figura destaca os argumentos principais (ou *core*) dos predicados, separando-os dos demais argumentos não-*core*. Na Figura 1, separam-se também os argumentos e modificadores de predicados dos modificadores de nominais. Ademais, a Figura 1 exhibe as etiquetas empregadas quando o dependente da relação está sob forma oracional (coluna *clauses*), que correspondem às orações subordinadas.

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

Figura 1 - Quadro de relações de dependência da UD.

Há três relações do tipo *core* em que os dependentes têm forma nominal⁷ (**subj**, **obj** e **iobj**) e três em que os dependentes têm forma oracional (**csubj**, **ccomp** e **xcomp**). Não são considerados argumentos *core* os complementos verbais preposicionados ou os adjuntos adverbiais.

Nas relações **obl**, **advmod**, **advcl**, **vocative**, **expl**, **dislocated**, **discourse**, **aux**, **cop** e **mark**, o predicado é o *head* e os dependentes são considerados seus modificadores. Apenas uma dessas relações tem forma oracional: **advcl**.

Já nas relações **nmod**, **appos**, **nummod**, **acl**, **amod**, **det** e **case**, o *head* é um nominal e o dependente é um modificador. O dependente nessas relações pode ser um nominal, uma palavra funcional (como as preposições, na *deprel case*, e os determinantes, na *deprel det*) ou um numeral (**nummod**). A *deprel acl* é a única que tem forma oracional, sendo usada para ligar nominais a orações adjetivas e completivas nominais.⁸

As *deprel cc* e *conj* são empregadas para tratar da coordenação de elementos da sentença/tweet.

⁶ <https://universaldependencies.org/u/dep/index.html>

⁷ Neste manual, o termo “nominal” se refere a palavras que podem exercer as funções típicas de substantivos (substantivos, pronomes, adjetivos e numerais com função substantiva).

⁸ Como destacado por Duran (2022), a relação de modificador nominal **clf** não é usada no português e, por isso, não está incluída neste manual.

Por fim, tem-se as 11 relações artificiais criadas pela UD, sendo **root** a mais importante delas. Essa *deprel* foi criada para marcar a raiz da árvore sintática de dependências. Estabelecer o **root** é o primeiro passo para se fazer a anotação de uma sentença ou *tweet*. As demais relações **fixed**, **flat**, **compound**, **parataxis**, **list**, **orphan**, **goeswith**, **reparandum**, **punct** e **dep** são usadas para anotar *tokens* que não apresentam relação sintática com outros *tokens*. Como essas relações não possuem sintaxe, a identificação do *head* da *deprel* é arbitrária, sendo esse papel normalmente atribuído ao primeiro dos dois *tokens* unidos pela relação.

Uma *deprel* liga dois elementos de uma sentença (ou *tweet*, no caso), tal que:

- um deles é denominado **head** (núcleo da relação) e o outro é denominado **dependente**;
- um *token* pode ser *head* de mais de uma relação;
- um *token* pode ser dependente de uma relação e *head* de outra;
- um *token* **não** pode ser dependente de mais de uma relação;
- o nome da relação está associado à função que o dependente realiza em relação ao *head*;
- a seta que representa uma *deprel* parte sempre do *head* em direção ao dependente da relação;
- o elemento apontado pela seta, no caso de dependente oracional, será o predicado da oração dependente;
- um *head* é sempre uma palavra de conteúdo (verbo, nome/substantivo, adjetivo, pronome, numeral e advérbio); exceções são símbolos que podem ser expressos por palavras, como R\$ (reais), % (por cento), etc.
- palavras funcionais (determinantes, preposições, conjunções) e sinais de pontuação, por sua vez, deverão ser sempre dependentes; no caso dos *tweets* do mercado financeiro, no entanto, há exceções, como ">" quando interpretado como verbo e anotado com **root**;

Além disso, a atribuição de *deprel* deve observar o princípio da projetividade, ou seja, os arcos das relações não podem se cruzar. A Figura 2 ilustra um *tweet* anotado com relações de dependência.

Antes de apresentar as diretrizes propriamente ditas, no entanto, cabem algumas observações importantes sobre a linguagem CGU ("conteúdo gerado por usuário"), especificamente dos *tweets* do mercado financeiro, uma vez que o corpus, por não ter sido normalizado, impõe desafios diversos à tarefa de anotação de *deprel*.

Os referidos *tweets* são comumente fragmentados, podendo ser compostos por sequências de sintagmas ou elementos simplesmente justapostos, apresentar truncamentos diversos, fenômenos típicos do *Twitter* (como URL, menção, *hashtag*, marca de *retweet* e outros), além de linguagem informal, com pontuação e ortografia que não seguem a norma padrão. Tudo isso acarreta elipses e ambiguidades, as quais dificultam a compreensão do conteúdo, sobretudo, a identificação do *root*. Além disso, a interpretação de um *tweet* sobre ações requer conhecimento do domínio financeiro devido à ocorrência de vocabulário terminológico (como as *cashtags*) e de estruturas de linguagem específicas. Dito isso, um *tweet* desse domínio pode ter mais de uma interpretação possível e, por conseguinte, mais de uma anotação sintática, não havendo, assim, uma escolha de *deprel* dita "correta". Assim, as diretrizes apresentadas aqui refletem a interpretação do anotador especialista.

Diferentemente do Manual de *deprel* para o português de Duran (2022), as diretrizes de anotação que compõem este Manual são apresentadas em duas partes.

Sobre as diretrizes, a primeira parte descreve as diretrizes de anotação das *deprel* do modelo UD para ocorrências idiossincráticas dos *tweets* do mercado financeiro. Tais ocorrências englobam:

- a) contextos com sintaxe não-padrão (e, por isso, não previstos no manual de anotação da língua portuguesa de Duran (2022), o que acarreta o emprego particular de certas *deprel*
- b) fenômenos que ocorrem nos *tweets* do referido corpus/domínio (cf. Di-Felippo et al. 2021), a saber: marca de *retweet*, menção, URL, *hashtag*, *cashtag*, truncamento, índice de valor das ações, *emoticon*, *emoji* e repetição de pontuação).

Na segunda parte, descrevem-se as diretrizes que são específicas para certos padrões estruturais que ocorrem com frequência nos *tweets* do mercado financeiro, sendo que algumas estruturas correspondem ao *tweet* como um todo.

Ademais, ressalta-se que, dada a fragmentação e, por vezes, a ocorrência de fenômenos típicos do domínio que dificultam a compreensão dos *tweets*, alguns exemplos estão associados a uma possível interpretação. Essa interpretação, aliás, é fundamental para compreender a anotação sintática (e escolha das *deprel*) ilustrada.

PRIMEIRA PARTE – IDIOSSINCRASIAS

Nesta segunda parte, descrevem-se as diretrizes de anotação das *deprel* do modelo UD para ocorrências idiossincráticas dos *tweets* do mercado financeiro. Tais ocorrências englobam:

- os fenômenos CGU (“conteúdo gerado por usuário”) que ocorrem nos *tweets* do referido *corpus*/domínio, a saber: *marca de retweet*, *URL*, *menção*, *hashtag*, *cashtag*, truncamento, índice de (des-)valorização das ações, *emoticon* e *emoji* e repetição de pontuação
- contextos com sintaxe não-padrão, o que acarreta o emprego particular de certas *deprel*

Fenômenos GCU

1. RT: marca de *retweet*

A marca de *retweet* (RT) sempre ocorre seguida de uma menção (@*mention*), sendo o *head* da relação de **nmod** que se estabelece com a menção (Figura 2).

Uma RT pode ocorrer integrada à sintaxe ou de forma isolada (em inglês, *standalone*).

Se integrada, deve ser anotada com a *deprel* relativa à sua função e PoS.

No caso dos *tweets* do DANTEStocks, a única ocorrência integrada à sintaxe foi anotada com **vocative**, como pode ser observado na Figura 2.

Nesse contexto, a RT é dependente do *head* “claro”, que é **root** do *tweet*.

Exemplo:

- (1) Mas é claro RT @garimpodeacoes : FITCH , cia de avaliação de risco diz que seca prolongada deve pressionar Sabesp (SBSP3) de forma mais severa

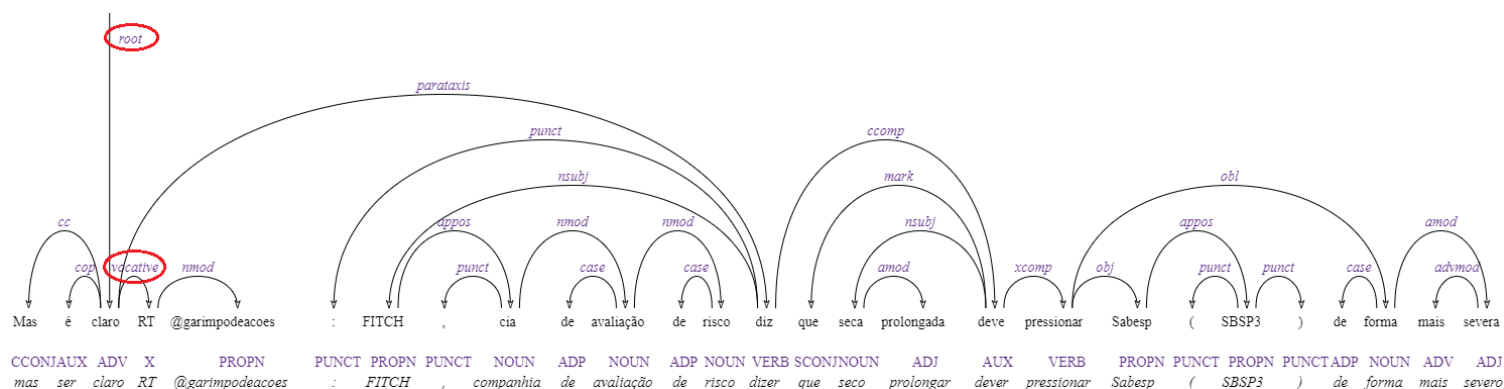


Figura 2 – Anotação de marca de *retweet* (RT) integrada à sintaxe com **vocative**.

Quando *standalone*, as marcas RTs devem ser conectadas ao **root** ou a outro *head* pela relação **parataxis**.

Exemplo:

- (2) **RT @daltonvieira : Ação ex-dividendos hoje : PCAR4 . As cotações históricas foram ajustadas . Saiba mais ! <http://t.co/C7k4DuDID2>**

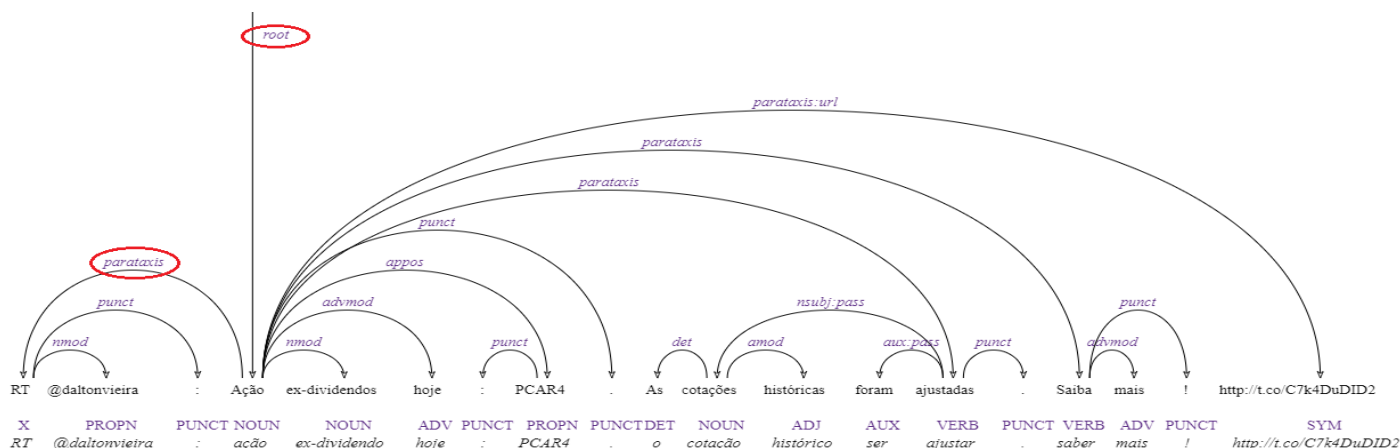


Figura 3 – Marca de retweet (RT) *standalone*, anotada como **parataxis** do **root**.

Um *retweet* (que se inicia com a marca de RT) pode apresentar conteúdo novo, inserido pelo usuário que realizou a repostagem.

Nesses casos, o **root** deve estar no trecho novo e a mensagem retuitada estará conectada ao **root** por **parataxis**.

Esse é o caso do *tweet* da Figura 4, em que o trecho sublinhado é a mensagem original (repostada) e o trecho em negrito foi inserido pelo usuário que retuitou a mensagem original.

Como se vê, o **root** está no trecho novo (“**piorando**”), sendo que a marca de RT está associada ao *head* do trecho original (“acompanhando”) por **parataxis**, e esse *head*, por sua vez, está ligado ao **root** também por **parataxis**.

Exemplo:

- (3) **RT @Pepez_Legal : @coroneldoblog Amigo , está acompanhando PETR4 ? ? > <http://t.co/WxnN4AOaKT> A Graça **está piorando** ainda mais o cenário .. rs ...**

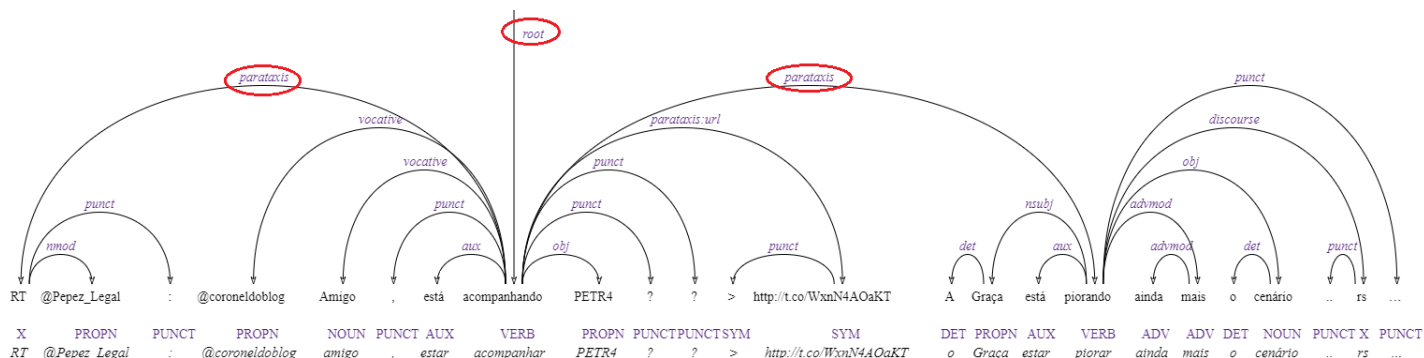


Figura 4 – *Retweet* com conteúdo novo, inserido pelo responsável pela repostagem.

2. Menção

As menções a usuários/perfis do *Twitter* foram anotadas com a PoS *tag* PROPN e podem ocorrer integradas à sintaxe da mensagem ou em contexto isolado (*standalone*).

Quando integradas, devem ser conectadas em nível sintático pela *deprel* que representa sua função/posição na mensagem, podendo ser **root** (Figura 5), **nmod** (Figura 6), **vocative** (Figura 7, 8, 9, 10, 11), **obl** (Figura 12), **obj** (Figura 13) e **nsubj** (Figura 14).

2.1. Root: raiz

A anotação de uma menção com **root** é dependente da interpretação do *tweet*. No caso da Figura 5, por exemplo, “@DepEduardoCunha” foi anotado como **root** por ser interpretado como predicado.

Exemplo:

- (4) @marisasacruz sim , sim . este é @DepEduardoCunha , contra o MC . E > @eduardocampos40 que fez PSB votar contra CPI #PETR4 <http://t.co/An5WtvvCYj>

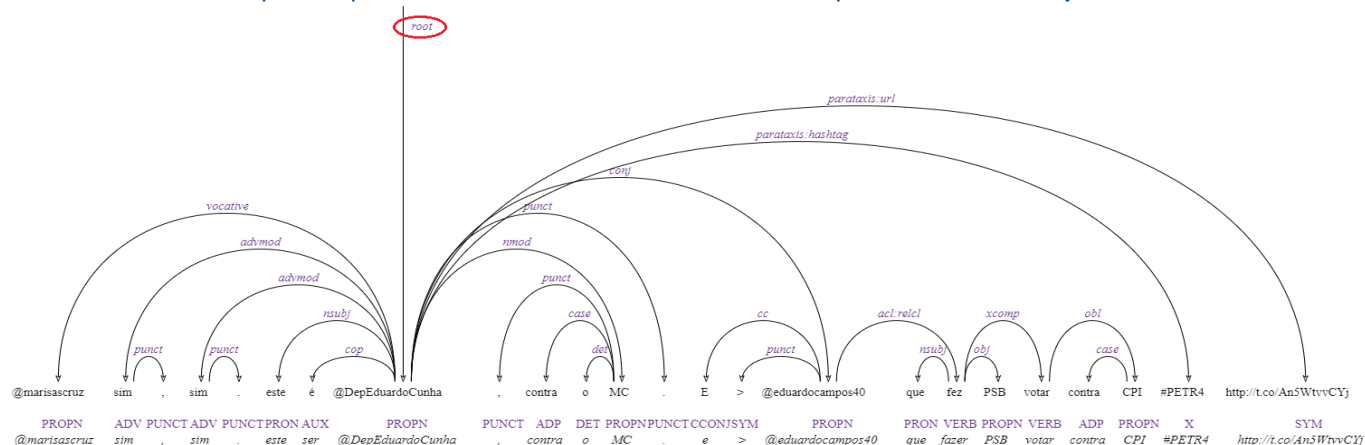


Figura 5 – Menção integrada à sintaxe anotada com **root**.

2.2. Nmod: modificador nominal

Quando uma menção funciona como modificador de outro substantivo, deve-se anotá-la com **nmod**, sendo que tal modificação ocorre primordialmente da esquerda para a direita.

Geralmente, a *deprel* **nmod** é intermediada por alguma preposição que, por sua vez, recebe a *deprel* **case**, a qual partirá do substantivo modificador (no caso, a menção) em direção à preposição.

Exemplo:

- (5) Cada vez que ouço a G. Foster defendendo o plano de investimento de a @petrobras , mais me certifico que devemos comprar PETR3 e 4 em a BOVESPA

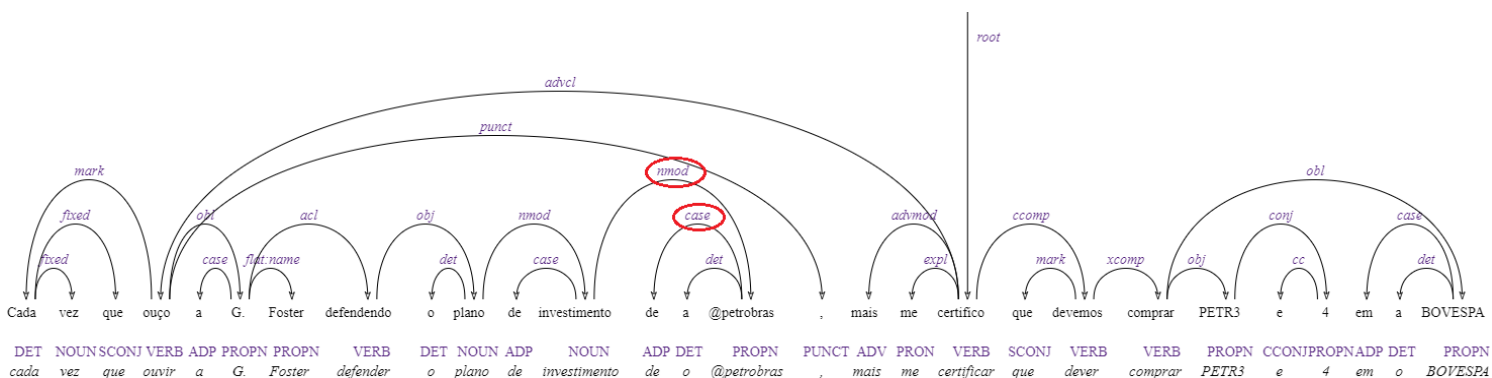


Figura 6 – Menção integrada à sintaxe anotada com **nmod**.

2.3. Vocative: vocativo

Nesses casos, interpreta-se a menção como uma referência ao participante do diálogo a quem se dirige a mensagem, sendo o predicado da oração principal o *head* da *deprel* **vocative**.

Nessa função, uma ou mais menções podem ocorrer no início ou no fim dos *tweets* (ou de uma sentença que os integra). Cada uma das menções deve ser conectada ao *head* por **vocative**. Quando ocorrem no final dos *tweets*, as menções podem ou não ser precedidas de pontuação.

Exemplo:

- (6) @alvarodias_ tem q ser contundente com essa petista q está ajudando a acabar com a Petr4 ...

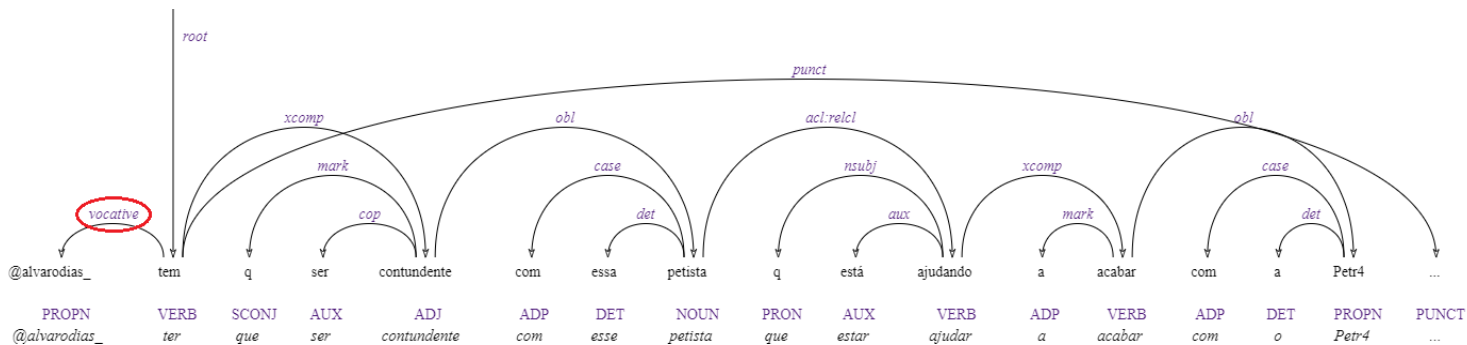


Figura 7 –Menção no início do *tweet* conectada ao *root* por **vocative** (exemplo 6).

Exemplo:

- (7) @ferriss @dfittarelli vendinha de itub4 em os 35,10 começando a dar frutos , VAI QUE VAI !

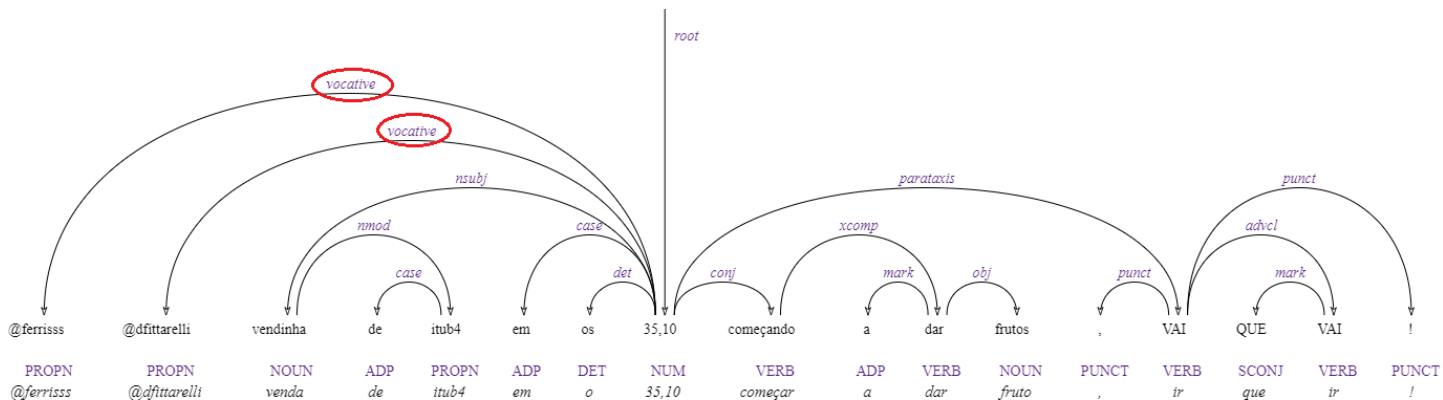


Figura 8 – Duas menções no início do *tweet* conectadas ao *root* por **vocative** (exemplo 7).

Exemplo:

- (8) Essa virada final em a #vale5 pagou meu japonês hoje . @ELISLEITAO

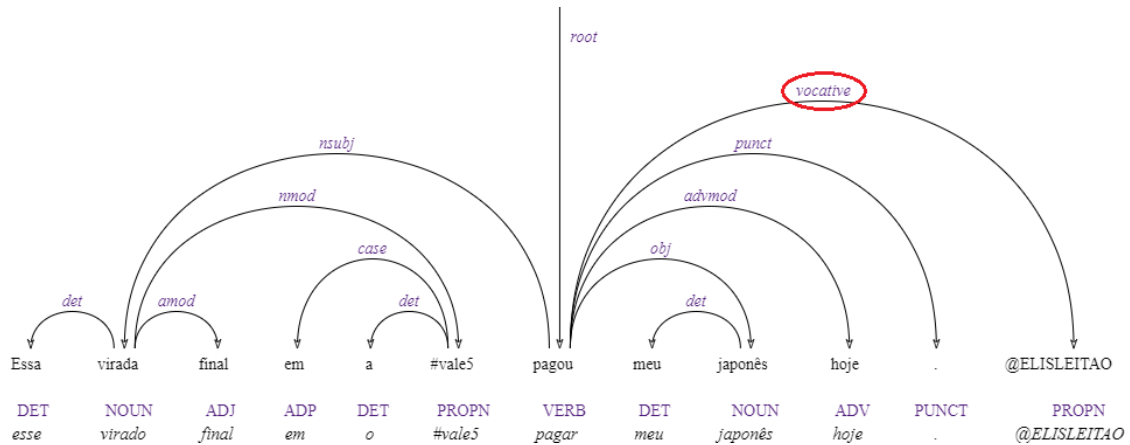


Figura 9 – Menção no final do *tweet* conectada ao **root** por **vocative** (exemplo 8).

Exemplo:

- (9) Aliás , essa p* de essa Pasadena tá funcionando ? 03/02/06 #PETR4 aprova compra de refinaria EUA <http://t.co/lfZWv799p0> @ivomarcelino @AryAntiPT

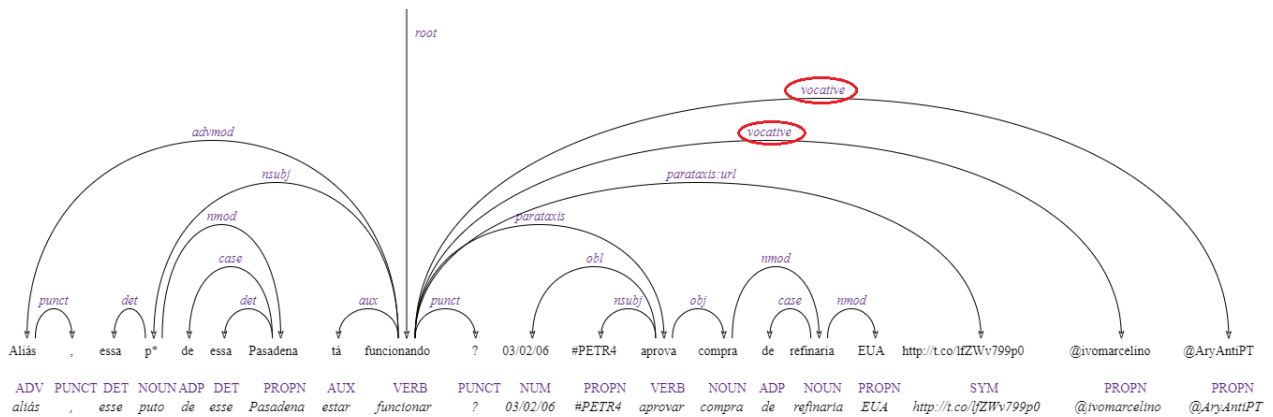


Figura 10 – Duas menções no final do *tweet* conectadas ao **root** por **vocative** (exemplo 9).

Exemplo:

- (10) Eita + um rojão em a Petr4 ! @clubedopairico Refinaria Pasadena – MP investigará Petrobras por evasão d divisas e peculato <http://t.co/egXYdJ7L8u>

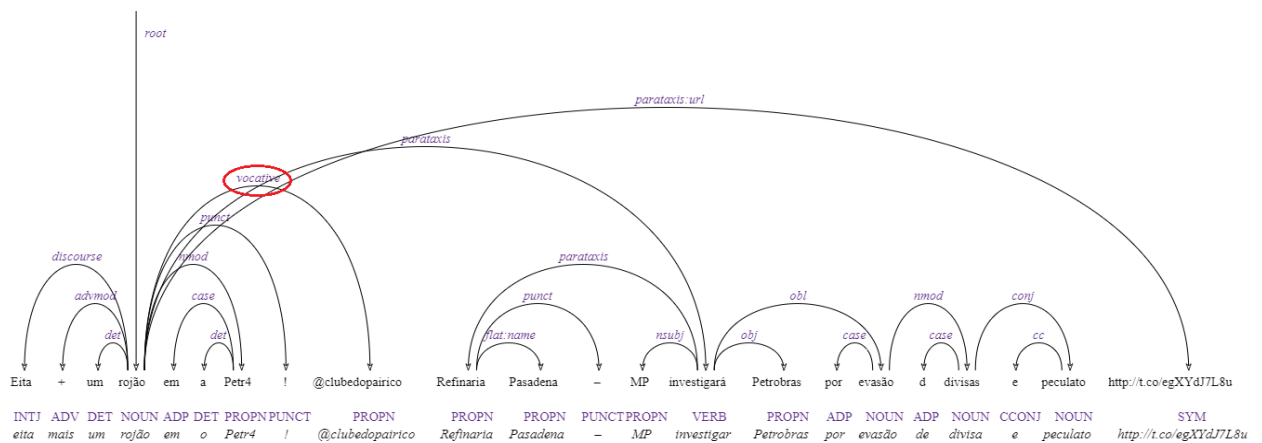


Figura 11 – Menção no final de uma sentença do *tweet* conectada ao **root** por **vocative** (exemplo 10).

2.4. Obl: nominal oblíquo

Uma menção deve ser anotada com a *deprel* **obl** quando for empregada para modificar um verbo (seu *head*). Nesse caso, a menção dependente de **obl** é introduzida por preposição.

Exemplo:

- (11) #PETR4 : EXCLUSIVO , em @VEJA : PF prende ex-diretor de a Petro em operação de lavagem de dinheiro <http://t.co/3wGEVANwmT> <http://t.co/YHocJpiEkv>

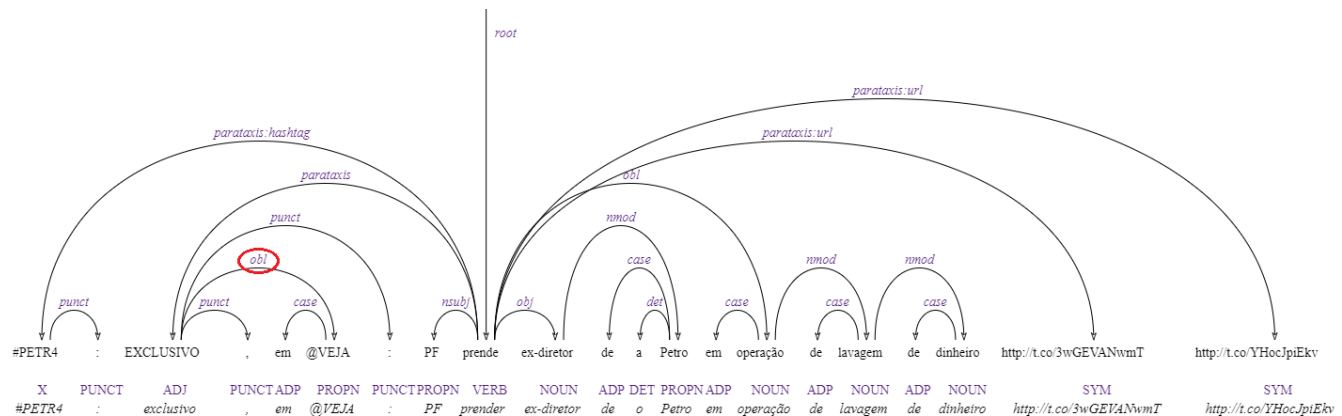


Figura 12 – Menção integrada à sintaxe anotada com **obl**.

2.5. Obj: objeto direto

Uma menção deve ser anotada com a *deprel* **obj** quando for empregada como segundo argumento *core* do verbo, sendo que o primeiro é **nsbj**. No caso da Figura 13, por exemplo, a menção “@petrobras” é dependente por **obj** do verbo “vincular”. Este, por sua vez, está conectado por **xcomp** a “tentam”, uma vez que o sujeito da oração dependente de **xcomp** (“vincular @petrobras a vergonha”) coincide com o sujeito da oração matriz, que é “coxinhas”.

Exemplo:

- (12) Os coxinhas aderem a o jogo sujo de o PIG e tentam vincular @petrobras a vergonha . Mas todo mundo quer compras ações de a PETR4

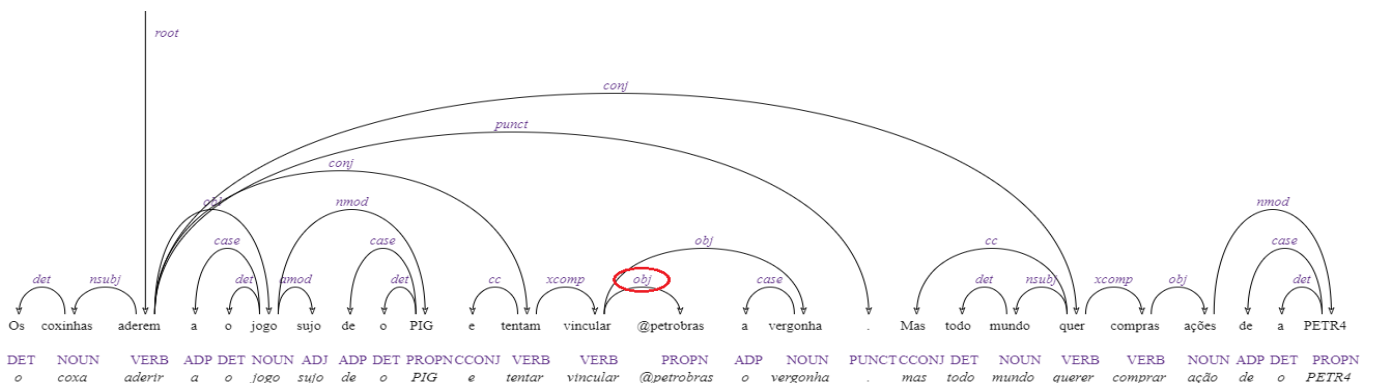


Figura 13 – Menção integrada à sintaxe anotada com **obj**.

2.6. Nsubj: sujeito

Deve-se anotar uma menção com a *deprel* **nsubj** quando ela ocorrer como primeiro complemento *core* de um predicado, isto é, como sujeito. No caso da Figura 14, por exemplo, a menção “@geraldalckmin_” foi conectada ao predicado verbal “diz” por meio de **nsubj**.

Exemplo:

- (13) último dia de o Gov FHC , a ação Petrobras (PETR3) > R\$ 3,3 @geraldalckmin_ diz q ações de a Petrobras “ viraram pó ” . <http://t.co/vAdFc65rmh>

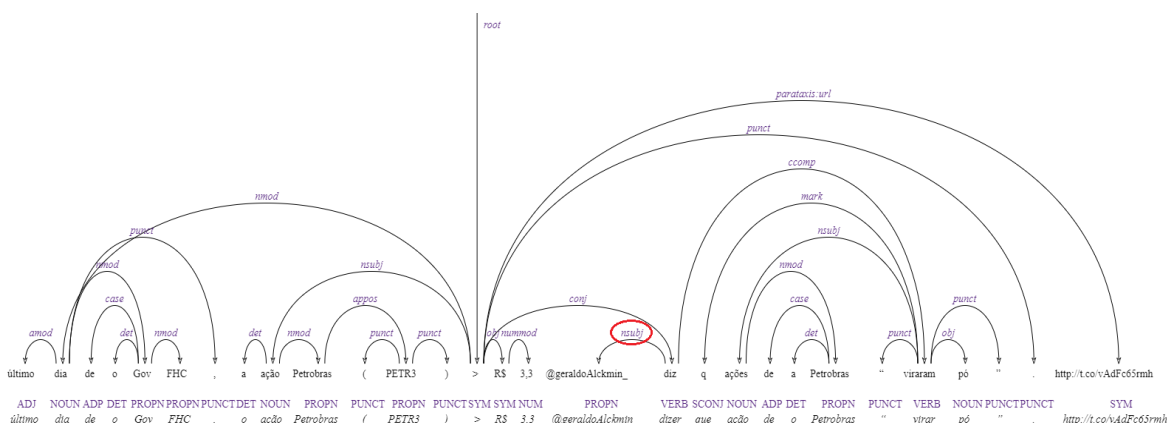


Figura 14 – Menção integrada à sintaxe anotada com **nsubj**.

Quando *standalone*, uma menção não tem relação sintática explicitada com seu *head* e, por isso, deve ser anotada como dependente por **parataxis** do *head* (que pode ser o *root*). Nesses casos, a menção pode ocorrer no começo, no meio ou no final do *tweet*. Quando no começo e no meio, tende a ser seguida por dois pontos, como nas Figuras 15 e 16. E, quando a ocorrência se dá ao final do *tweet*, pode aparecer entre parênteses (Figura 17).

A conexão de uma menção ao *root* ou a outro *head* por **parataxis** depende da interpretação do conteúdo do *tweet*. Nas Figuras 15 e 17, por exemplo, as menções “@JornalMercantil” e “@oswaldosena_” conectam-se ao *root* por **parataxis**, uma vez que dizem respeito à informação central da postagem. O mesmo não ocorre com “@Info_BMFBOVESPA” (meio) (Figura 16), que é dependente do *head* “fecha”.

Exemplo:

- (14) “ @JornalMercantil : China planeja reduzir importação de minério de ferro <http://t.co/Yqyjr6Z2fB> ” #vale5

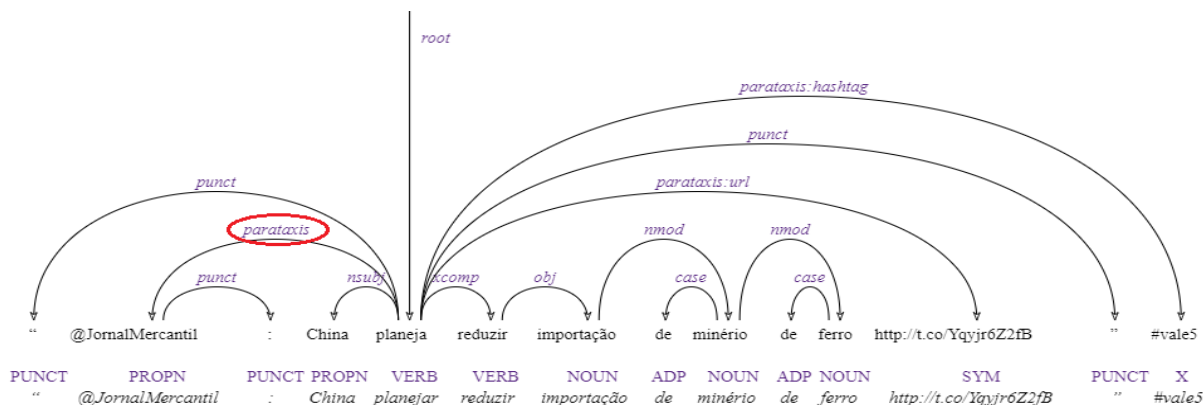


Figura 15 – Menção com **parataxis** no início do *tweet* e seguida de pontuação.

Exemplo:

- (15) \$PETR4 R\$ 13,11 - 1,58 % / \$VALE5 R\$ 25,9 - 2,12 % , ' @Info_BMFBOVESPA : #Ibovespa fecha em baixa de 0,91 % a os 45.443 pontos . '

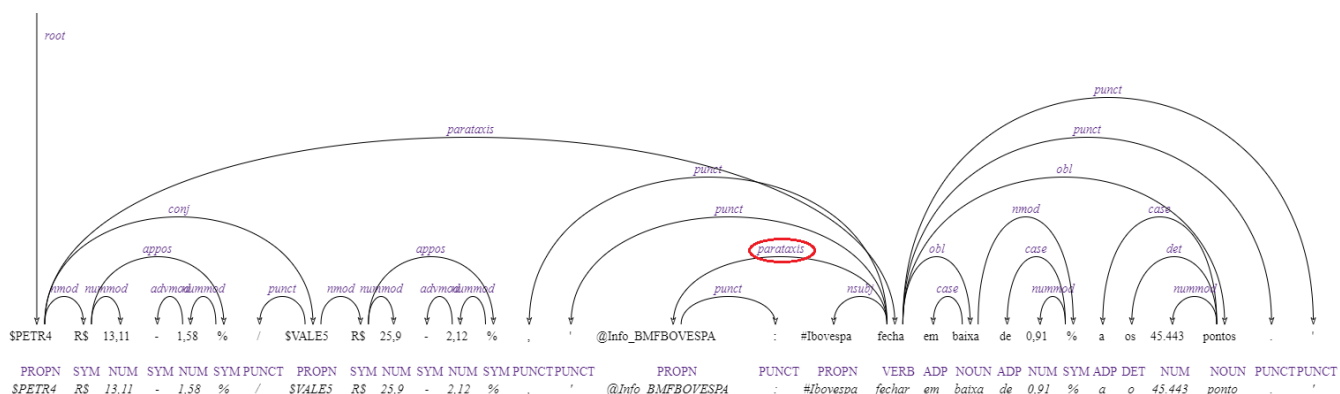


Figura 16 – Menção com **parataxis** no meio do *tweet* e seguida de pontuação.

Exemplo:

- (16) Petrobras (#PETR) , Rossi (#RSID3) e BB (#BBAS) batem mínimas de até 10 anos . <http://t.co/O0X4s9y6Ab> (@oswaldosena_1)

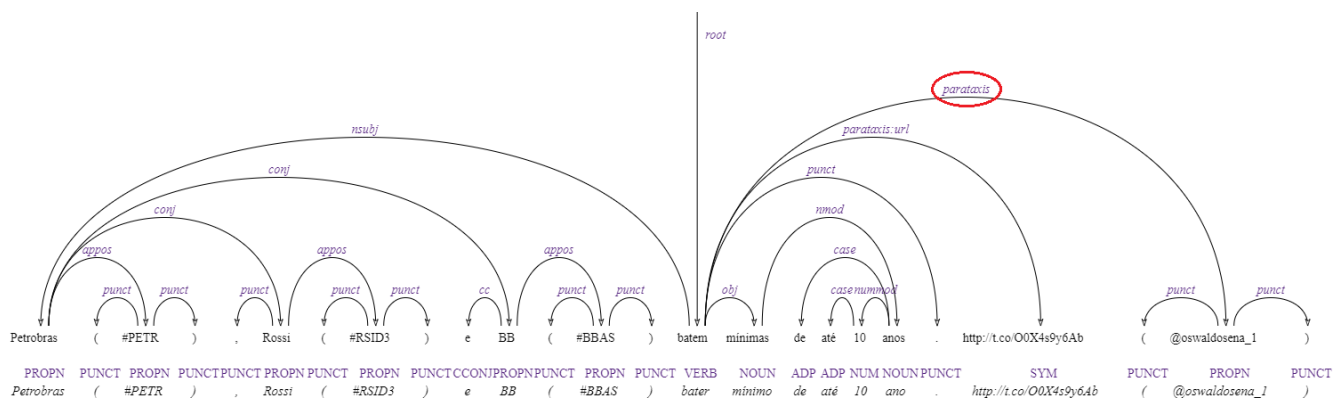


Figura 17 – Menção com **parataxis** no final do *tweet* e entre parênteses.

3. URL: endereço web

Uma URL tem sempre PoS SYM e pode ocorrer integrada à sintaxe ou *standalone*. Se integrada, pode ocorrer precedida de preposição (Figura 18) ou de dois-pontos (Figura 19). Nesses casos, a URL deve ser conectada a um *head* por **obl**.

Exemplo:

- (17) Publiquei estudo de a #HGTX3 em o gráfico diário . Rompendo tendência de baixa ? ? ? Veja em <http://t.co/oRA4bA8Qye>

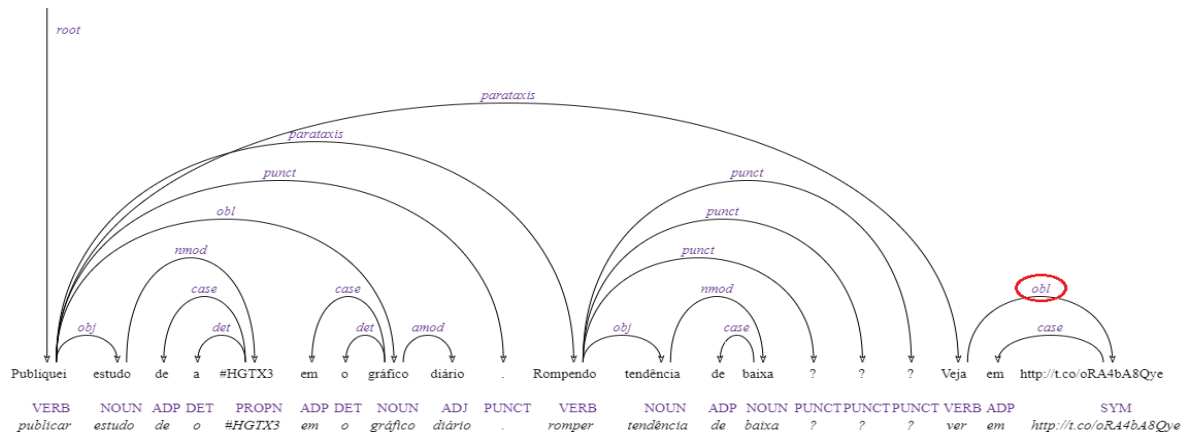


Figura 18 – URL precedida por preposição e conectada a um *head* por **obl**.

Exemplo:

- (18) Em os últimos 5 pregões #CSNA3 acumulou uma baixa de 17.1 % enquanto #USIM5 - 14.8 % e #VALE5 - 10 % . Veja o ranking : <http://t.co/XjRazUAN9b>

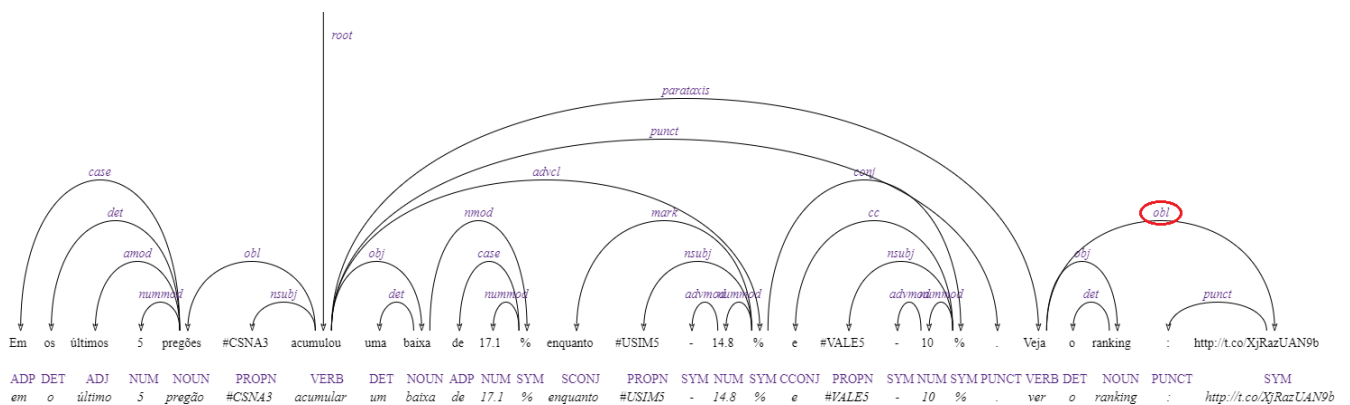


Figura 19 – URL precedida por dois pontos e conectada a um *head* por **obl**.

Ainda quando integrada, uma URL pode completar o sentido de um verbo, comumente no imperativo, ligando-se, portanto, ao seu *head* por meio da *deprel* apropriada.

Exemplo:

- (19) Hoje , a as 9h , o Bradesco realizará teleconferência sobre os resultados de o 1T14 . Acesse <http://t.co/FHE37z52Q2> e participe ! #BBDC4

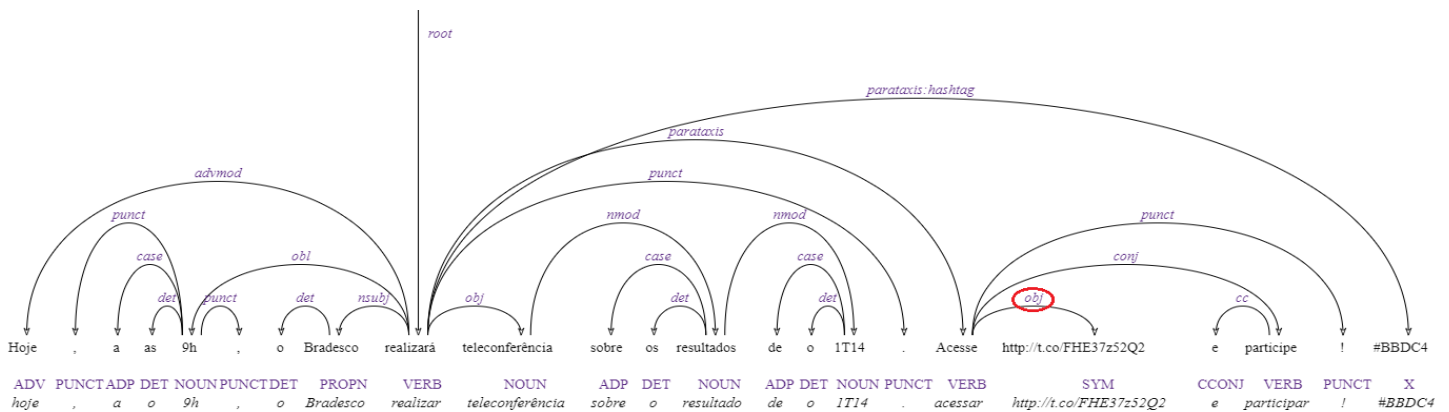


Figura 20 – URL com preposição implícita e conectada a um *head* por **obj**.

Se *standalone*, uma URL ocorre ao final dos *tweets*, podendo ser ou não precedida por um ponto final (Figura 21) ou reticências (Figura 22).

Nesses casos, deve-se conectá-la ao **root** por **parataxis**, acrescida da sub-relação **url**.

Exemplo:

- (20) @garimpodeacoes \$RSID3 4Q13 (N) Geração Op de Caixa forte , com desalavancagem financeira . Margens se recuperam YoY . <http://t.co/CC8MGagICJ>

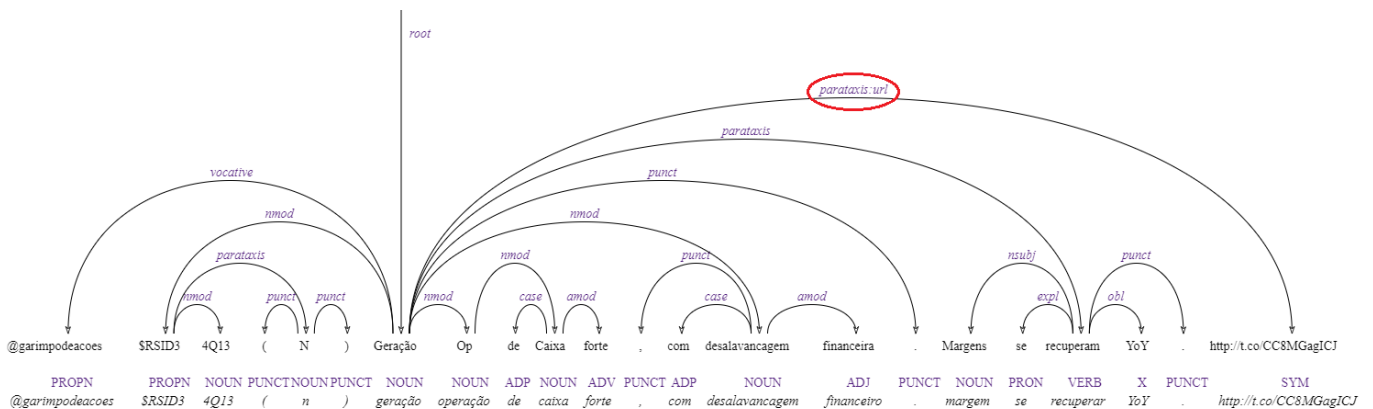


Figura 21 – URL *standalone* precedida de ponto final e conectado por **parataxis:url** (exemplo 20).

Exemplo:

(21) Petrobrás Pn (Petr4) , Gráfico Diário . Ação registr ... <http://t.co/Zsml5piTaT>

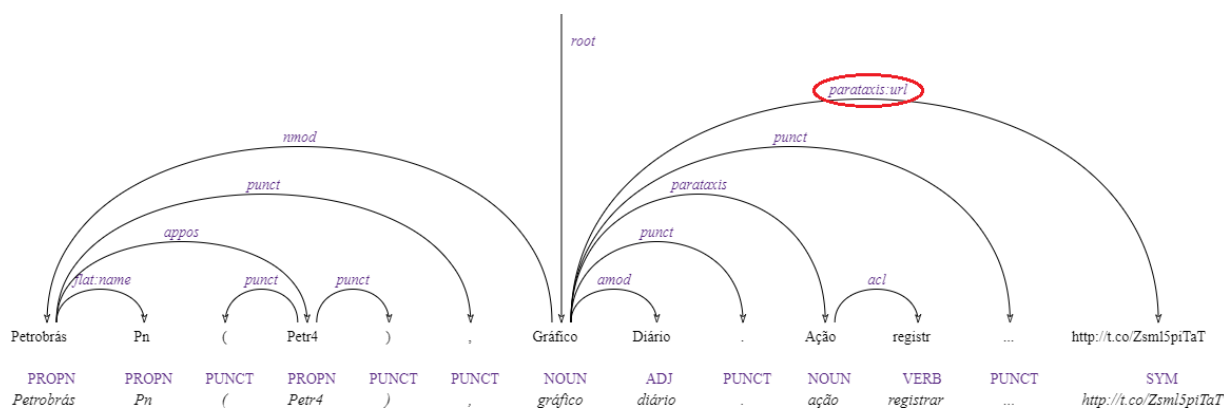


Figura 22 – URL *standalone* precedida de reticências e conectada por **parataxis:url**. (exemplo 21).

Exemplo:

(22) #BBAS3 #BBDC4 RT @passagensaereas : Banco de o Brasil e Bradesco fazem parceria para criar novo programa de fidelidade <http://t.co/zeQF7pMYMo>

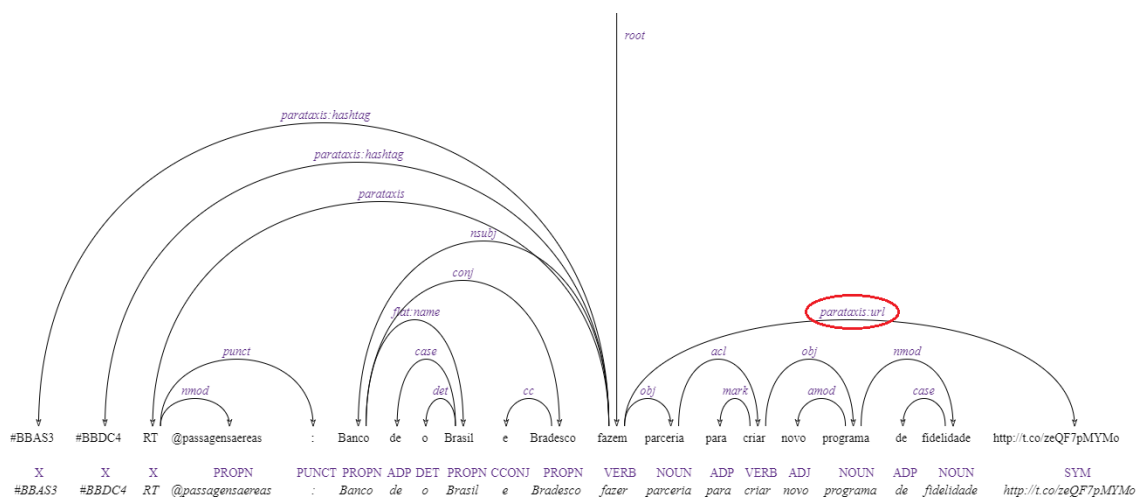


Figura 23 – URL *standalone* não precedida de pontuação e conectada por **parataxis:url** (exemplo 22).

4. Hashtag

As *hashtag* são *tokens* comumente compostos por uma palavra precedida pelo símbolo “#” (como #petr4). Elas são típicas da linguagem das redes sociais, podendo ocorrer integradas ou *standalone*.

Quando integradas (PoS PROPEN) devem ser anotadas com base na sua função/posição sintática. A seguir, estão as diretrizes para as diferentes funções sintáticas (*deprel*) que as *hashtags* podem ter quando integradas à sintaxe do *tweet*, a saber: **root**, **nmod**, **nsubj**, **obj**, **obl**, **appos** e **vocative**.

4.1. Root: raiz

Exemplo:

(23) #petr4 stop se fechar acima 17,775 e romper

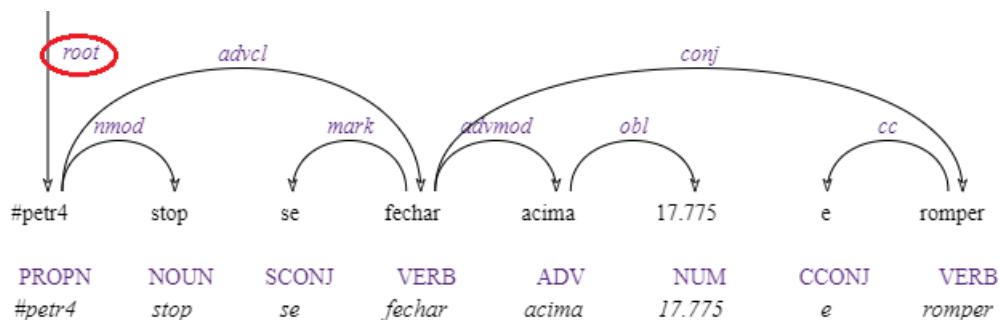


Figura 24 – Hashtag anotada com **root**.

4.2. Nmod: modificador nominal

Exemplo:

(24) Lembra de o sinalzinho de fundo de a #petr4 ?

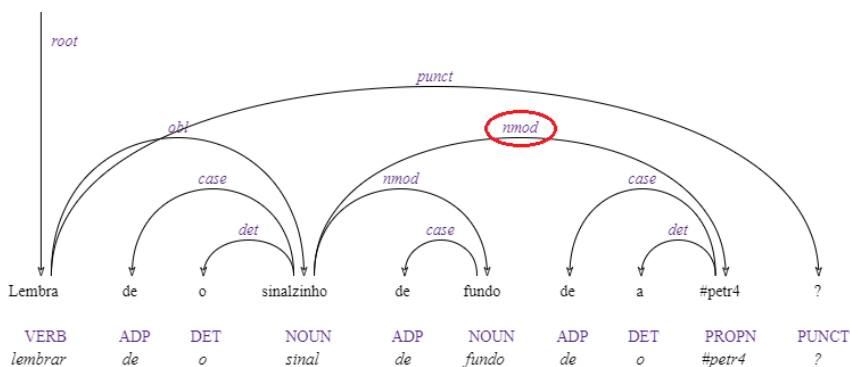


Figura 25 – Hashtag anotada com **nmod**.

4.3. Nsubj: sujeito

Exemplo:

(25) RT @Live_Trade : **#petr4** cumpriu certinho e ...

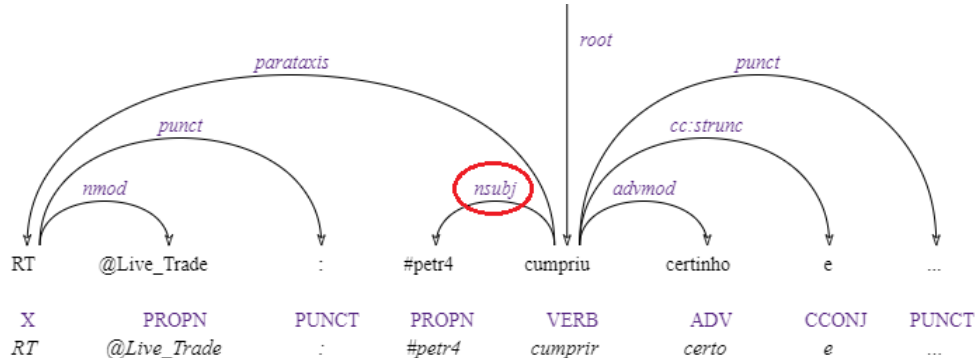


Figura 26 – Hashtag anotada com **nsbj**.

4.4. Obj: objeto direto

Exemplo:

(26) Agora entendem pq eu estava tranquilo comprando **#petr4** #BBAS3 e #vale5 Agora ficou desenhado para os q não entendiam .

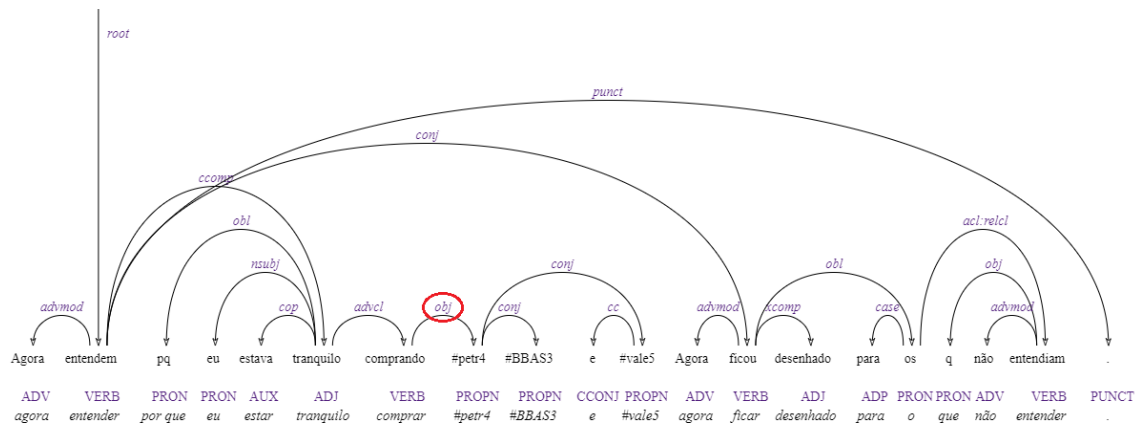


Figura 27 – Hashtag anotada com **obj**.

4.5. Obl: nominal oblíquo

Exemplo:

- (27) Entrei em **#petr4** , com projeção de crescimento . #hype3 ta maravilhosa , continua em a tendência de alta .

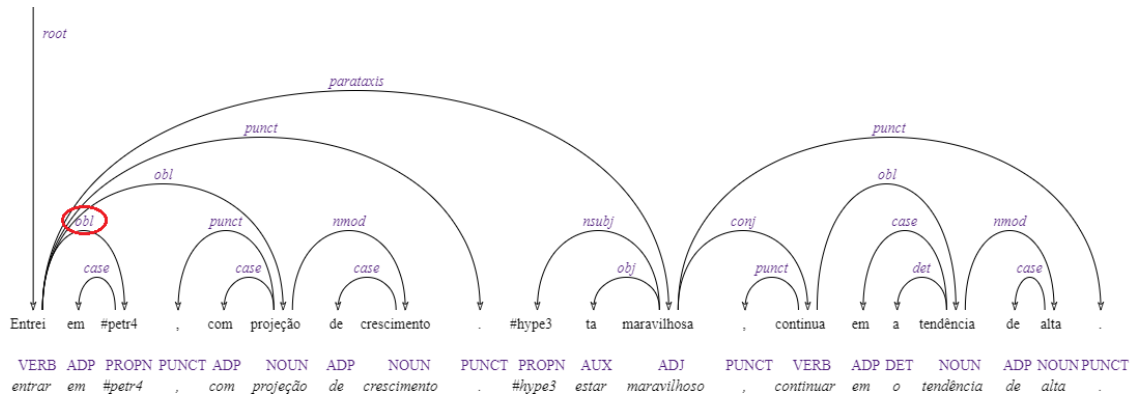


Figura 28 – Hashtag anotada com **obl**.

4.6. Appos: modificador apositivo

Exemplo:

- (28) @Live_Trade Bom dia meu amigo , e o KING KONG (**#petr4**) , sabe me dizer suporte e resistência ? abraço

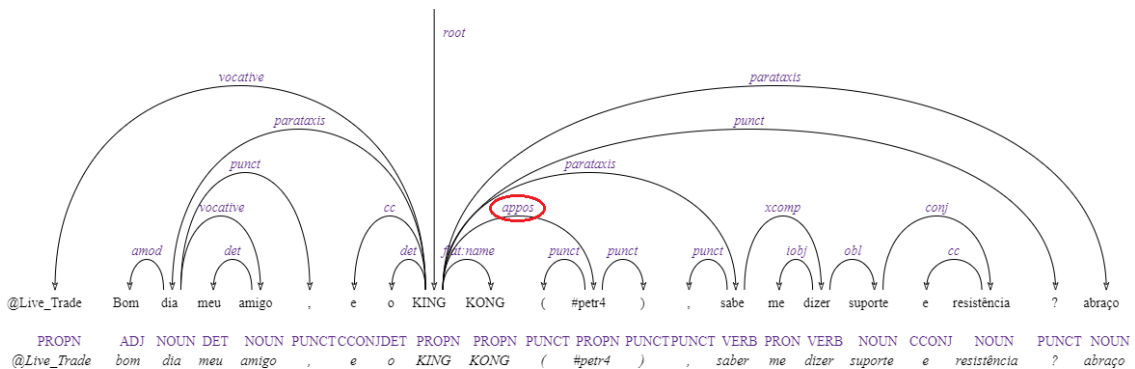


Figura 29 – Hashtag anotada com **appos** (exemplo 28).

Exemplo:

(29) @lambari_trader E não é que a Petrobras #petr4 visitou os R\$ 12 ... !!!



Figura 30 – Hashtag anotada com **appos** (exemplo 102).

4.7. Vocative: vocativo

Exemplo:

(30) #vale5 vc é uma putinha mesmo .. rsrs .. em o finalzinho que vc resolve abrir as pernas .. rsrs

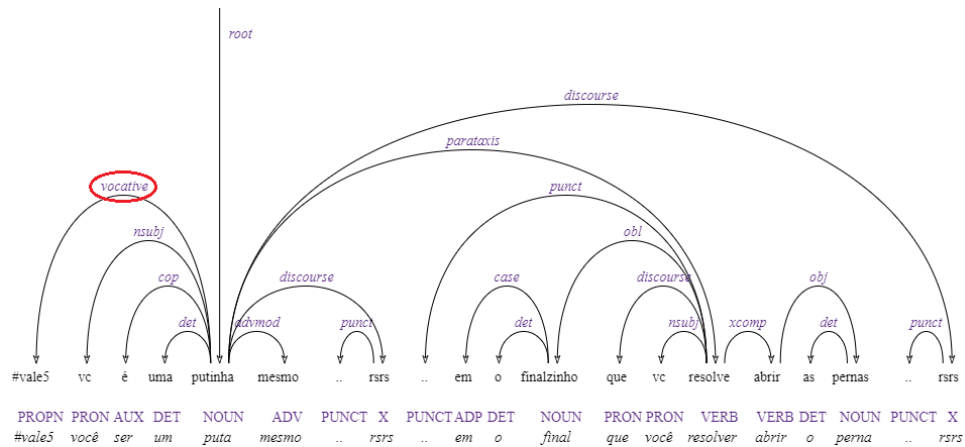


Figura 31 – Hashtag anotada com **vocative**.

Quando *standalone*, as *hashtags* podem ocorrer no início ou no final dos *tweets*. Nesses casos, a PoS é **X** e, no nível sintático, devem ser conectadas a um *head* por **parataxis**, acrescida pela sub-relação **:hashtag**. Caso haja mais de uma *hashtag*, cada uma deve ser conectada ao *head*.

Exemplo:

- (31) **#PETR4** RT @valor_economico : Câmara derrota governo e aprova comissão para investigar Petrobras <http://t.co/p37D1t2ARU>

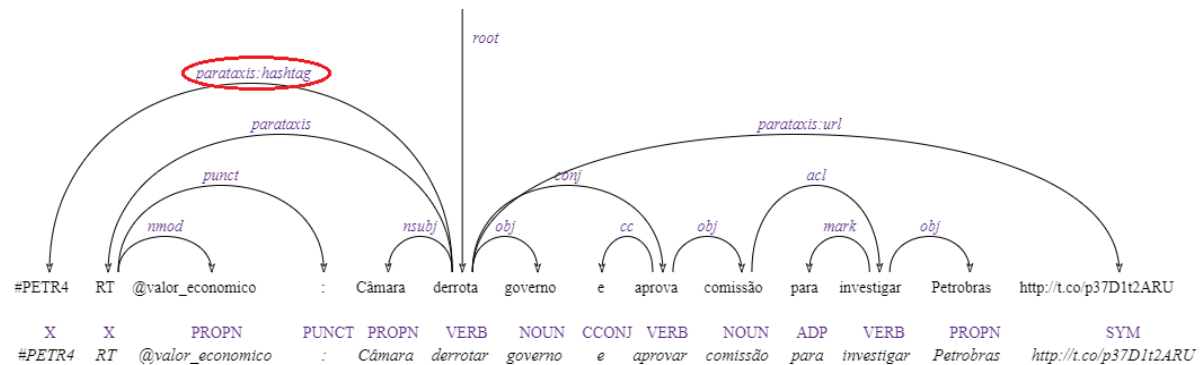


Figura 32 – *Hashtag standalone* no início do *tweet* e anotada com **parataxis:hashtag** (exemplo 31).

Exemplo:

- (32) **#IBOV** **#KLB4** Klabin fecha contrato com a Pöyry para Projeto Puma : <http://t.co/WPqvXaEQB2>

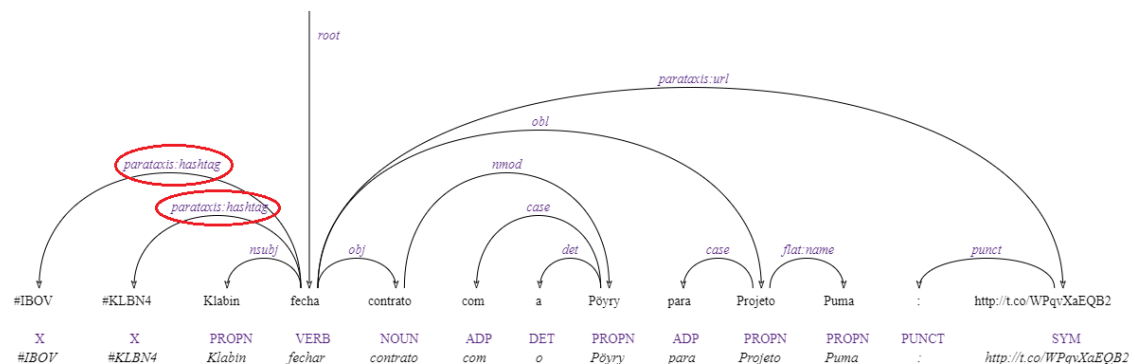


Figura 33 – Duas *hashtags standalone* no início e anotadas com **parataxis:hashtag** (exemplo 32).

Exemplo:

- (33) Agora é oficial ... Ministério de a Fazenda publicou portaria que aumenta a tributação de bebidas frias . #ABEV3

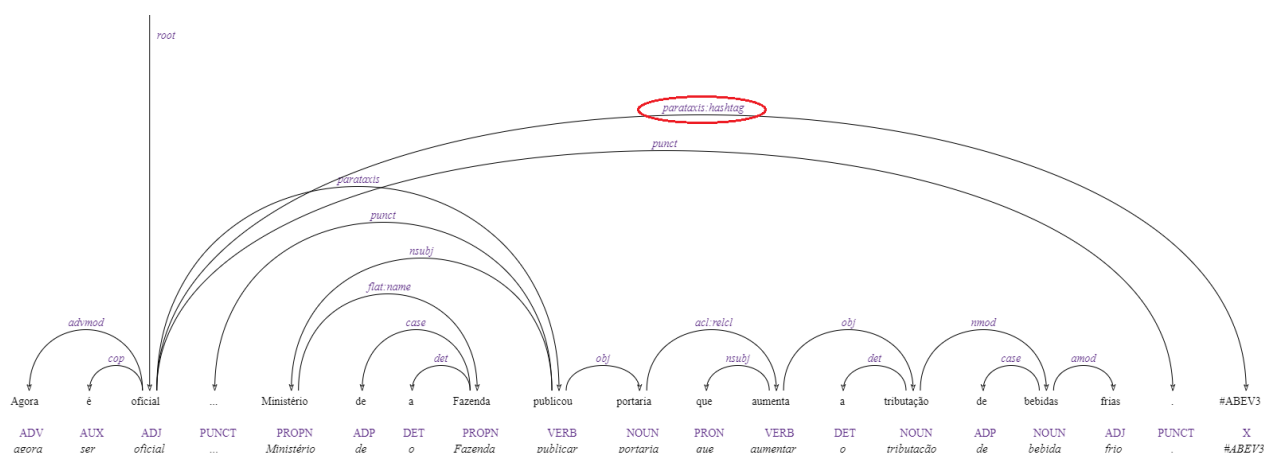


Figura 34 – Hashtag standalone no final do tweet anotada com **parataxis:hashtag** (exemplo 33).

5. Cashtag

As *cashtags* são *tokens* comumente compostos por uma palavra precedida pelo símbolo “\$” (como \$petr4). Elas são características da linguagem das redes sociais no domínio financeiro, podendo ocorrer integradas ou *standalone*.

Quando integradas (PoS PROPN) devem ser anotadas com base na sua função/posição sintática. A seguir, estão as diretrizes para as diferentes funções sintáticas (*deprel*) que as *cashtags* podem ter quando integradas à sintaxe do *tweet*, a saber: **root**, **nmod**, **nsubj**, **obj** e **appos**. Diferentemente das *hashtags*, as *cashtags* não ocorrem como **vocative** e **obl**.

5.1. Root: raiz

Exemplo:

- (34) \$PETR4 R\$ 13,11 - 1,58 % / \$VALE5 R\$ 25,9 - 2,12 % , ' @Info_BMFBOVESPA : #Ibovespa fecha em baixa de 0,91 % a os 45.443 pontos . '

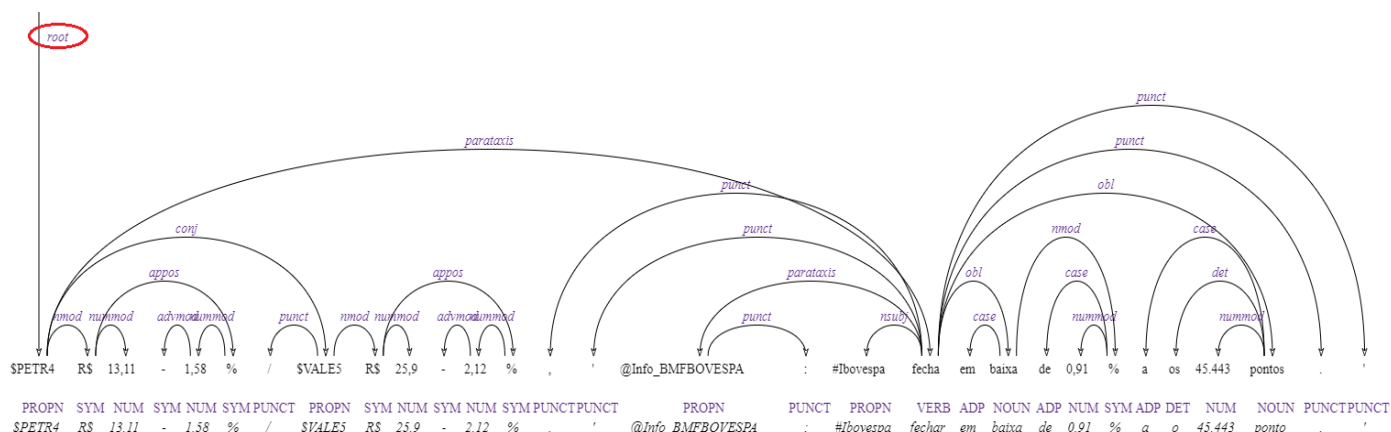
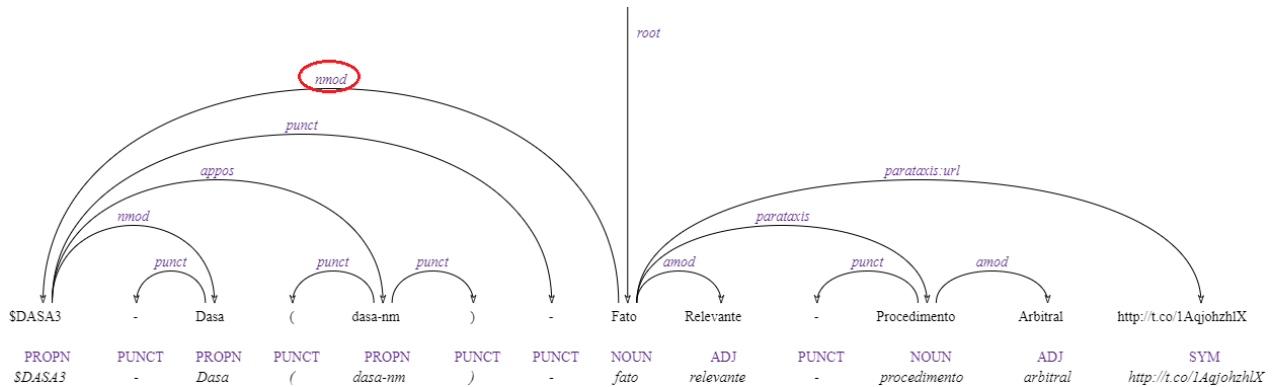


Figura 35 – Cashtag anotada com **root**.

5.2. Nmod: modificador nominal

Exemplo:

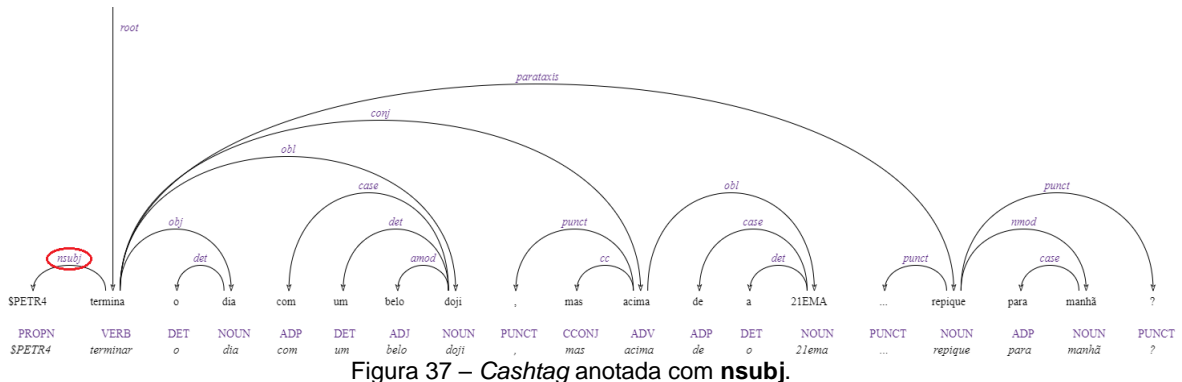
(35) **\$PETR3** - Atencao Para O Preco Petr73 (petr3) <http://t.co/frle2l3TBF>



5.3. Nsubj: sujeito

Exemplo:

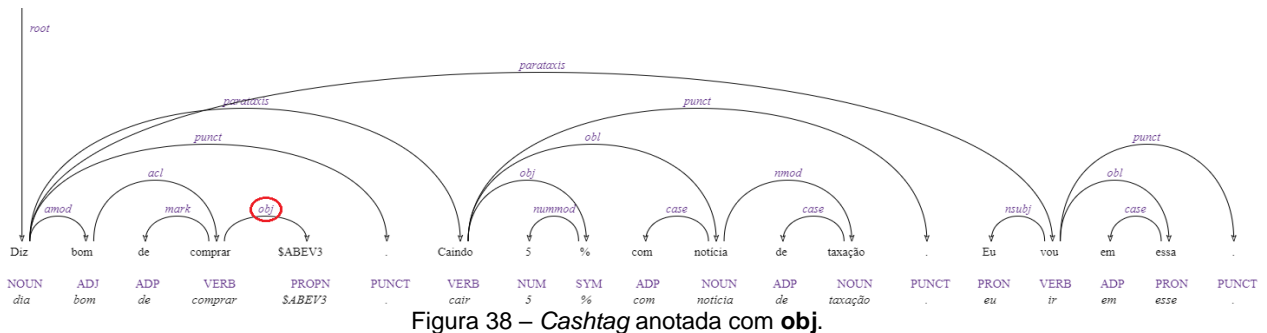
(36) **\$PETR4** termina o dia com um belo doji , mas acima de a 21EMA ... repique para manhã ?



5.4. Obj: objeto direto

Exemplo:

(37) Diz⁹ bom de comprar **\$ABEV3** . Caíndo 5 % com notícia de taxaço . Eu vou em essa .



⁹ Tokens com desvio ortográfico (ou erro de digitação) não foram alterados no corpus, mas seus traços morfosintáticos (incluindo PoS e lema) estão descritos em função de sua forma correta, a qual, aliás, também é indicada na coluna MIRC do CoNLL-U pelo traço adicional CorrectForm. Na anotação sintática, assume a função da forma correta no tweet.

Figura 40 – Relação que chega no predicado principal do trecho truncado com sub-relação **:strunc** e última palavra do trecho truncado com **:wtrunc**.

Exemplo:

- (40) Elétricas sofrem ' apagão ' e levam junto siderúrgicas , 11 ações caem mais de 4 % :
Apenas 9 **ação** ... <http://t.co/UfrAbJS2fS> #infomoney #vale5

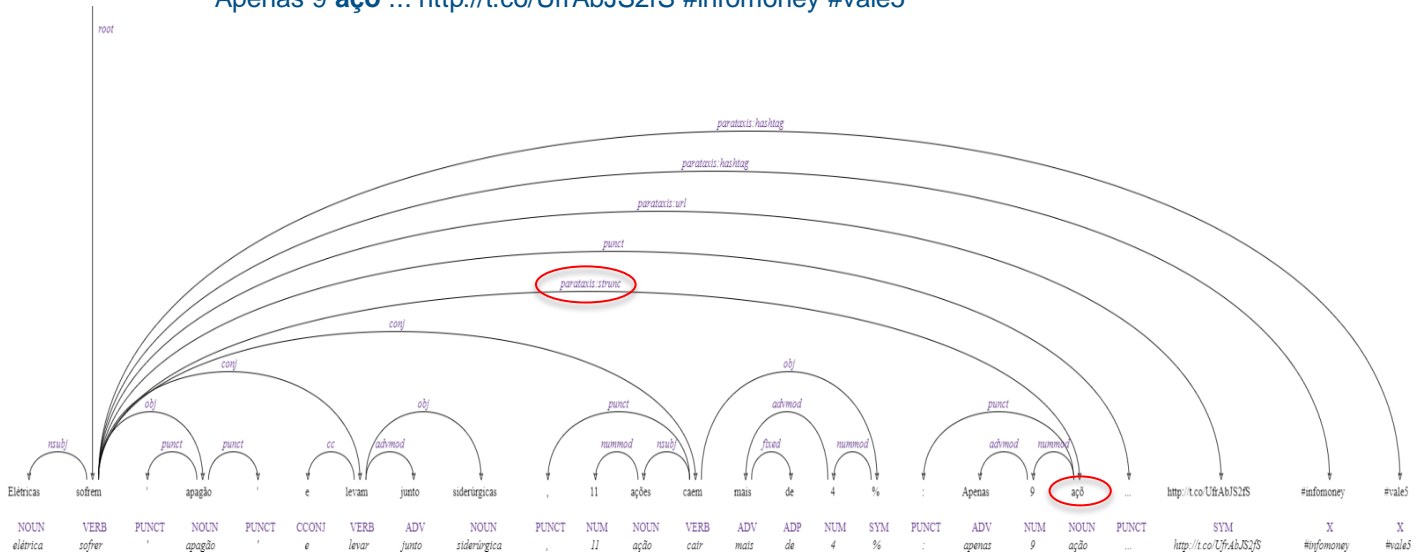


Figura 41 – *Head* truncado de trecho truncado: relação da qual o *head* é dependente tem sub-relação :strunc.

2. Se as reticências (“...”) indicarem um truncamento estrutural, mas o texto/trecho estiver sintaticamente completo, sugere-se, nesse caso, não usar a sub-relação :strunc.

Exemplo:

- (41) Notas gerais A VALE (VALE5) pretende pagar , como primeira parcela de remuneração mínima a os acionistas em 2014 , ... <http://t.co/RB6XViLTFD>

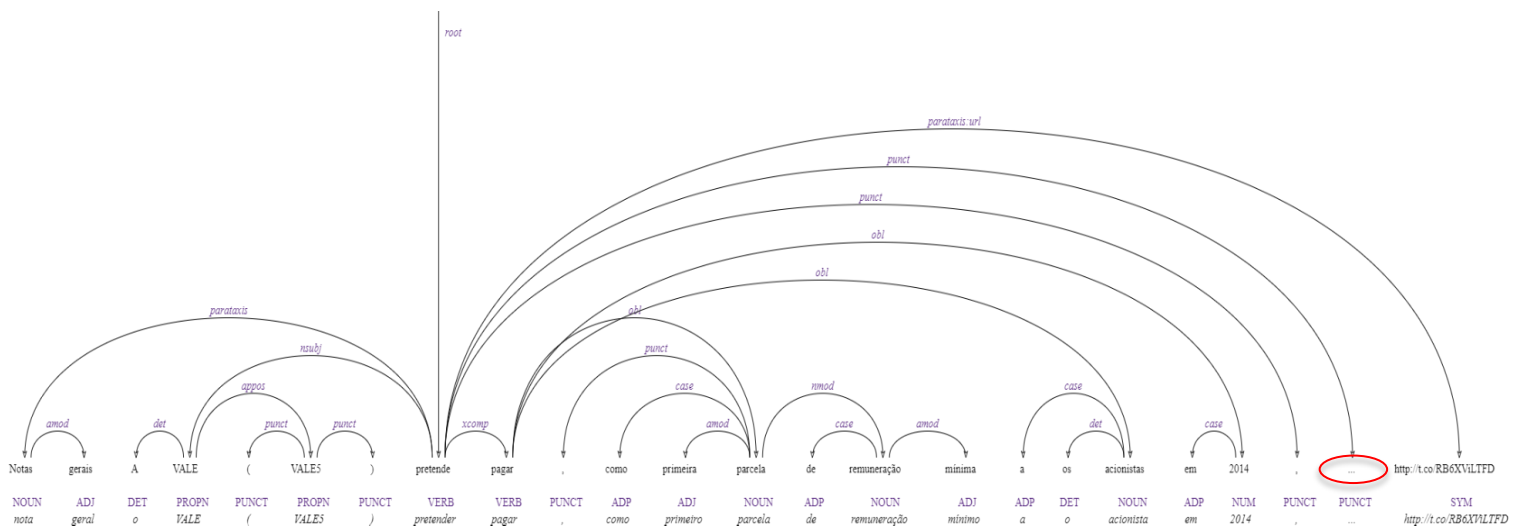


Figura 42 – Truncamento estrutural sem quebra de conteúdo não requer sub-relação :strunc (exemplo 41).

Exemplo:

(42) Notas gerais A PETROBRAS (PETR4) demitiu o diretor financeiro de a subsidiária Petrobras Distribuidora , Nestor ... <http://t.co/CMZD46pQOq>

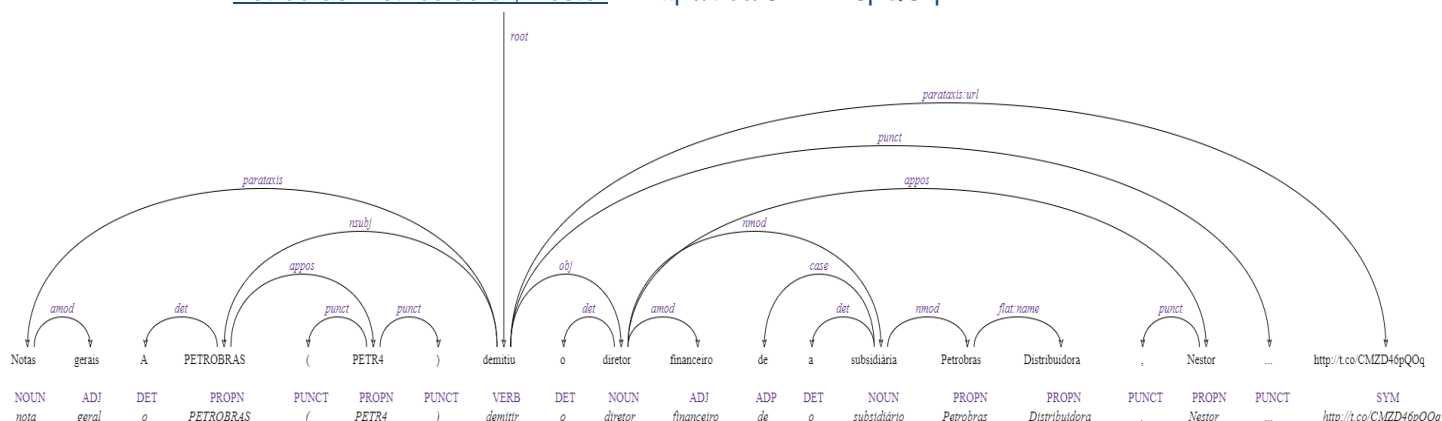


Figura 43 – Truncamento estrutural sem quebra de conteúdo não requer sub-relação :strunct (exemplo 116).

2.7. Índice de (des-)valorização das ações

Segundo Di Felippo *et al* (2021), os *tweets* sobre o mercado financeiro apresentam os chamados “índices de (des-)valorização das ações”, como - 1,3 %, que são expressões sempre compostas pela seguinte sequência de PoS: SYM, NUM, SYM.

A anotação de *deprel* entre esses 3 elementos segue as diretrizes (exemplos 43):

1. símbolo (PoS SYM) “%” é *head*
2. valor numérico (PoS NUM) “1,3” é dependente de “%” por **nummod**
3. símbolo “-” (PoS SYM) é dependente de % por **advmod**

Os índices ocorrem em dois contextos sintáticos distintos, sendo que em cada um deles o *head* (“%”) será dependente de outro *token* por meio de *deprels* diferentes.

Um índice pode aparece inserido em contexto nominal, sem que um verbo ocorra ou seja inferido. No *tweet* do exemplo (43) a expressão nominal “maiores altas (seguida de dois pontos)” possui o *root* (“altas”) e a sequência de índices ocorre depois dos dois pontos.

Nesses casos, o *head* do índice (“%”) é **nmod** do *ticker* que o precede e o primeiro *ticker* da sequência é dependente por **appos** do *root*.

Exemplo:

(43) Maiores Altas : LLXL3 + 7,79 % | OIBR4 + 6,14 % | JBSS3 + 6,11 % | RSID3 + 5,67 % | ELET3 + 4,83 % .

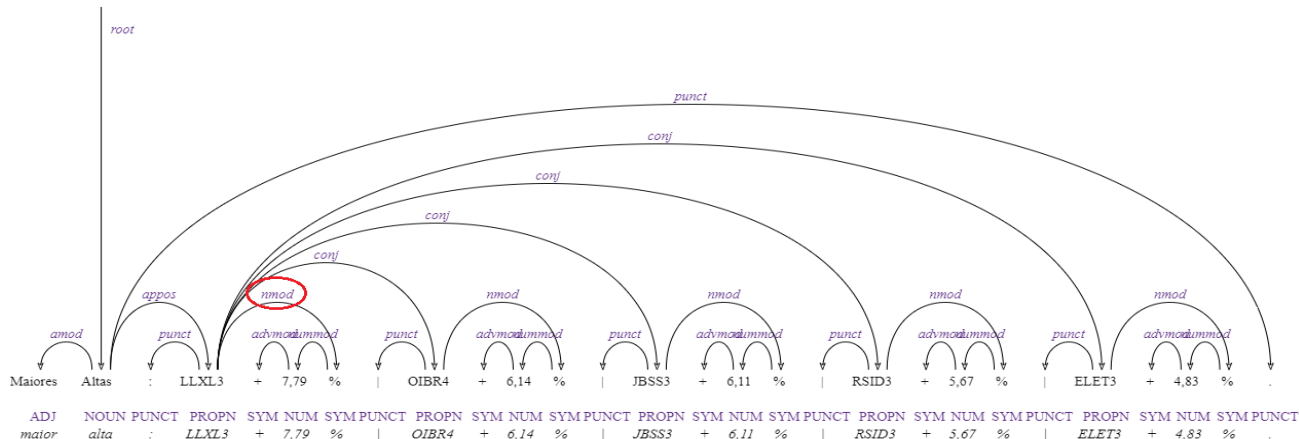


Figura 44 – Índice de (des-)valorização como dependente por **nmod** do *ticker* que o precede.

Em outro contexto, um índice pode aparecer em contexto verbal, com verbo explícito ou implícito. Em (44), por exemplo, pode-se inferir que a sentença inicial seja “Ações da VALE também estão em alta”, sendo que cada índice herda o predicado inferido, como “Vale3 está em alta de...”. Diante dessa interpretação, o **root** do *tweet* é o *token* “alta”, sendo o *head* do índice (“%”) dependente por **orphan**¹⁰ do *ticker* que o precede, pois o símbolo “%” e o *ticker* ficam “órfãos” de *head* em função do predicado elíptico. O primeiro *ticker* da sequência, nesse caso, é dependente por **parataxis** do **root**.

Exemplo:
(44) Ações de a VALE também em alta . Vale3 , + 2,1 % e Vale5 , + 1,6 ...

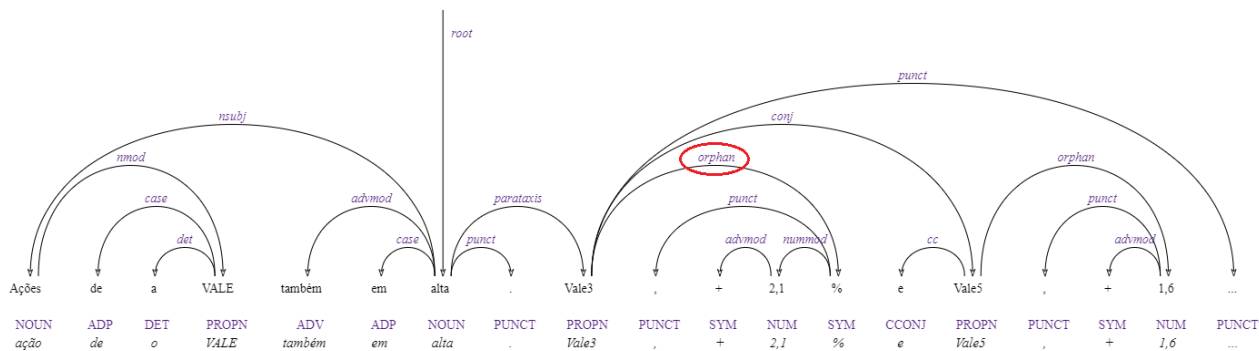


Figura 45 – Índice de (des-)valorização como dependente por **orphan** do *ticker* que o precede.

2.8. Emoticon e emoji

Os *emoticons* (como “:.”, “:|”, “o.O” e “:D”) e *emojis* (como 🍕) só ocorrem *standalone* no *corpus* que deu origem a este manual.

Nesses casos, *emoticons* e *emojis* devem ser conectados ao **root** por **discourse**, pois não têm uma relação clara com a estrutura sintática do *tweet*, exceto de maneira expressiva.

A identificação do *head* adequado, no entanto, é contextual. No caso da Figura 46, por exemplo, a ocorrência das aspas indica que a sequência de *emojis* se refere ao *head* (“enviada”) do trecho entre aspas e não ao **root** do *tweet*.

Para sequências de um mesmo *emoticon* ou *emoji* (exemplo 45), cada um deles deve ser conectado ao seu *head* por **discourse**.

Exemplo:
(45) Vendido ! :) “ @felipefdeaguiar : #ELPL4 ordem enviada R\$ 9,30 vendo ! 🍕 🍕 🍕 ”

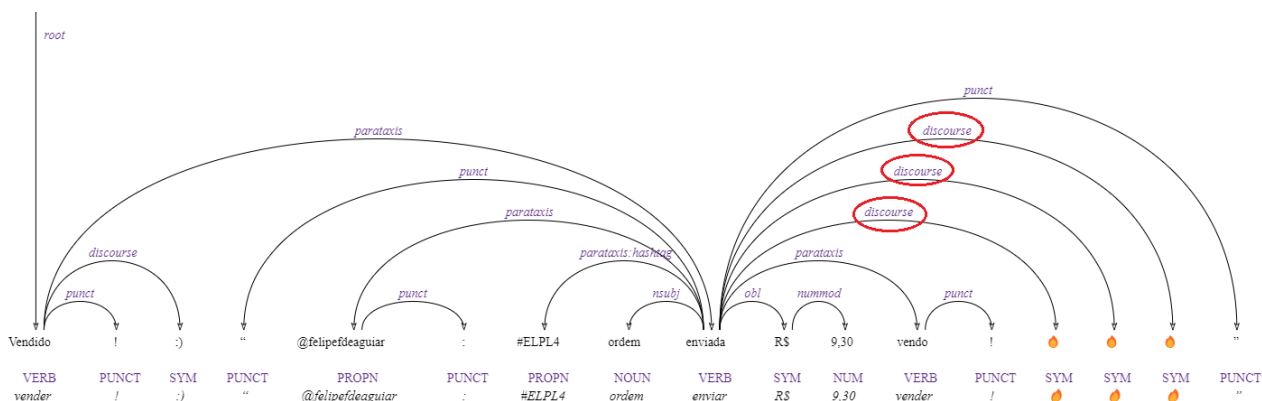


Figura 46 – Emoticon e emoji em contexto *standalone* conectados por **discourse** ao **root**.

¹⁰ A *deprel orphan* (do inglês, *orphaned dependente*) liga dois elementos que ficaram “órfãos” de *head*, em função da elipse do *head* que tinham em comum. Essa relação é usada tipicamente quando há elipse de um predicado e pelo menos duas palavras de conteúdo que se ligariam a esse predicado (DURAN, 2022).

2.9. Onomatopeias

As onomatopeias de risos (como hehehe, kkkkk, haha e outras) (PoS X) nos *tweets* do mercado financeiro ocorrem em contexto *standalone* e, assim como *emojis/emoticons*, relacionam-se à estrutura sintática do *tweet* apenas de maneira expressiva, sendo, portanto, conectadas ao **root** pela *deprel discourse*.

Exemplo:

(46) #Vale5 Opções é caça níquel , igual tem em os botecos **kkk**

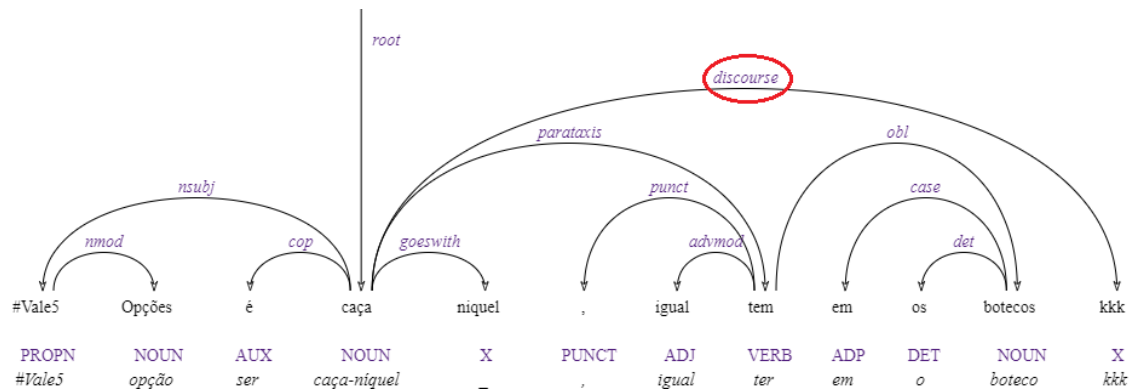


Figura 47 – Onomatopeia *standalone* conectada por **discourse** (exemplo 46).

Exemplo:

(47) #petr4 King Kong me acordem qdo bater em 12,50 que tenho interesse ... **rsrsr**

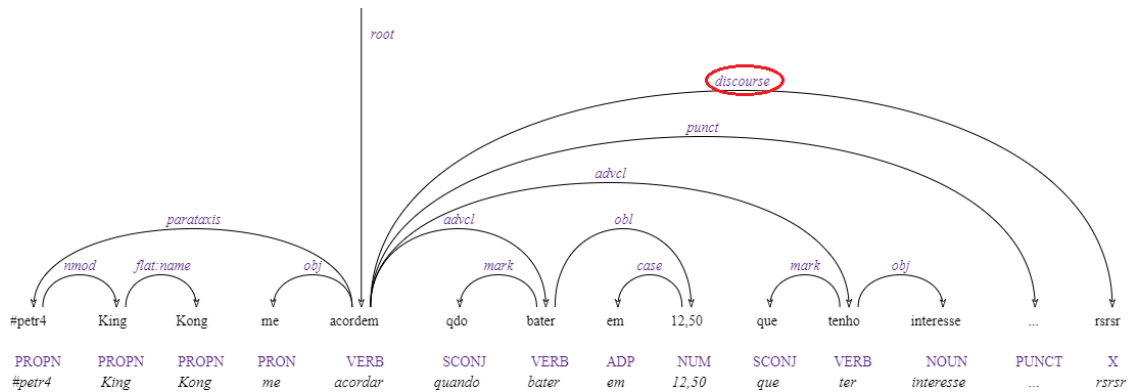


Figura 48 – Onomatopeia *standalone* conectada por **discourse** (exemplo 47).

2.10. Repetição de sinal de pontuação

Para sequências de um mesmo sinal de pontuação (como “? ? ?”), cada um deles deve ser conectado ao seu *head* por meio de **punct**.

Exemplo:

(48) @Live_Trade Marcos , projeta algum fundo para OIBR4 ? ? ?

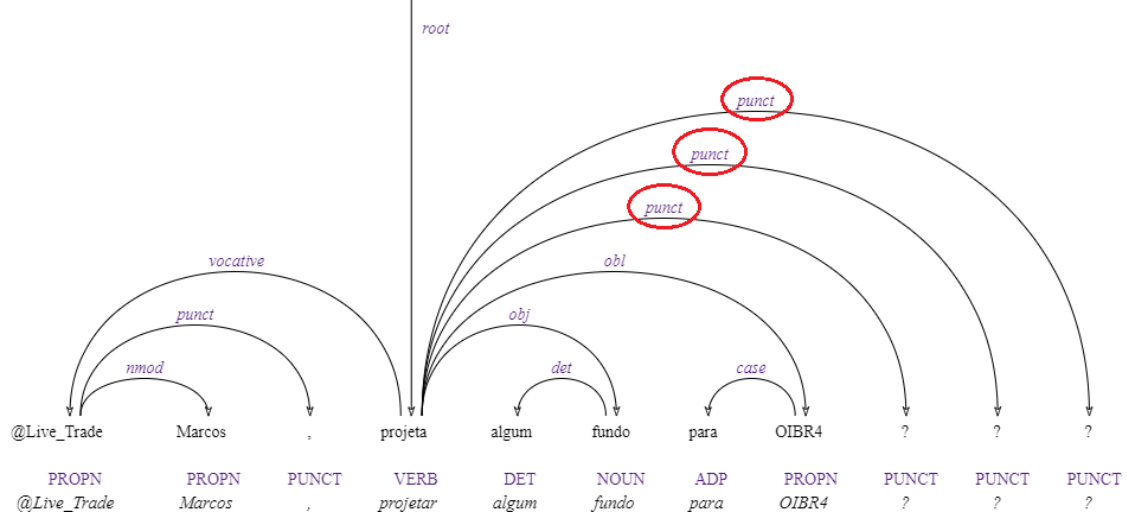


Figura 49 – Cada sinal de pontuação repetido em sequência conectado por **punct** ao *head* (exemplo 48).

2.11. Substituição lexical por símbolo (SYM)

Repetições em sequência do cifrão “\$” podem ocorrer integradas à sintaxe ou *standalone*. Quanto integradas, ocorrem em substituição à palavra “dinheiro”, com função de objeto direto. Nesses casos, a opção foi por conectar apenas o primeiro cifrão da sequência por **obj** ao *head* e os outros símbolos da sequência por **goeswih**¹¹ ao primeiro.

Exemplo:

(49) Quando Lula mandou trabalhador aplicar \$\$ de o FGTS em a #PETR4 , ações valiam R\$ 24 . Hoje valem R\$ 15,5 e contando para baixo : - 34 % !

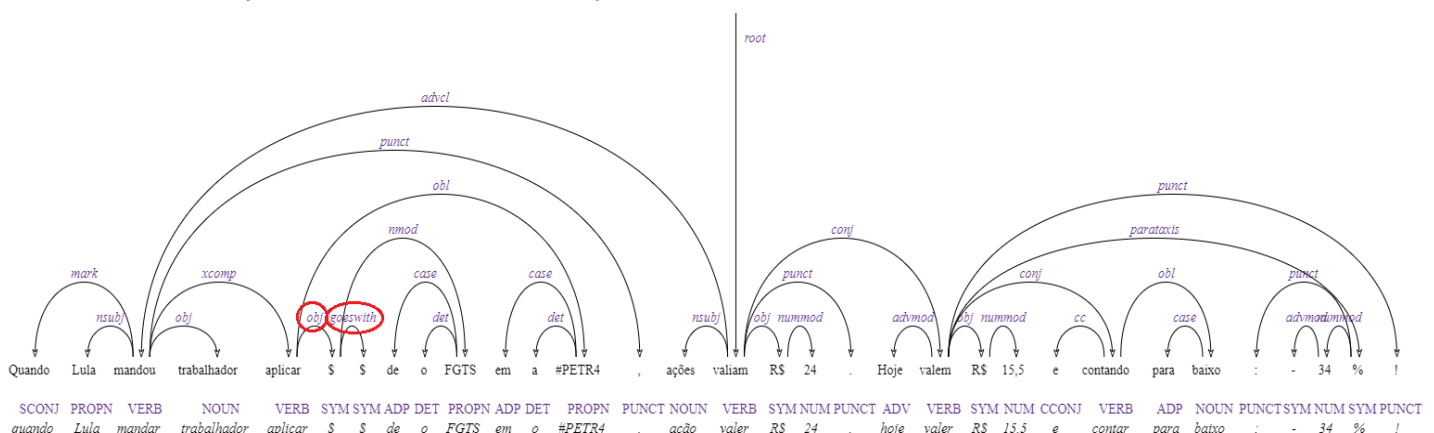


Figura 50 – Primeiro cifrão conectado por **obj** ao *head* e por **goeswith** com o outro cifrão (exemplo 49).

¹¹ Nos *tweets* originais, os cifrões repetidos em sequência formam um único *token*, sem espaço em branco entre eles (p.ex.: “&&”). Assim, optou-se por usar **goeswith**, pois, segundo Duran (2022), essa *deprel* pode ser empregada para conectar elementos que foram *tokenizadas* indevidamente, como parece ser o caso.

Quando *standalone*, as repetições sequenciais do cifrão devem ser conectadas a um *head* por **discourse**, pois não têm uma relação clara com a estrutura sintática do *tweet*, exceto de maneira expressiva. Nesses casos, cada um dos símbolos da sequência deve ser conectado ao *head* por **discourse**.

Exemplo:

- (50) #PETR4 indo p picas @SakaSakamori : Denúncia . Estrangulamento \$\$\$ de a Petrobras <http://t.co/RMBjR9d80F> Beira de o colapso ! #DilmaTemMedoDaCPI

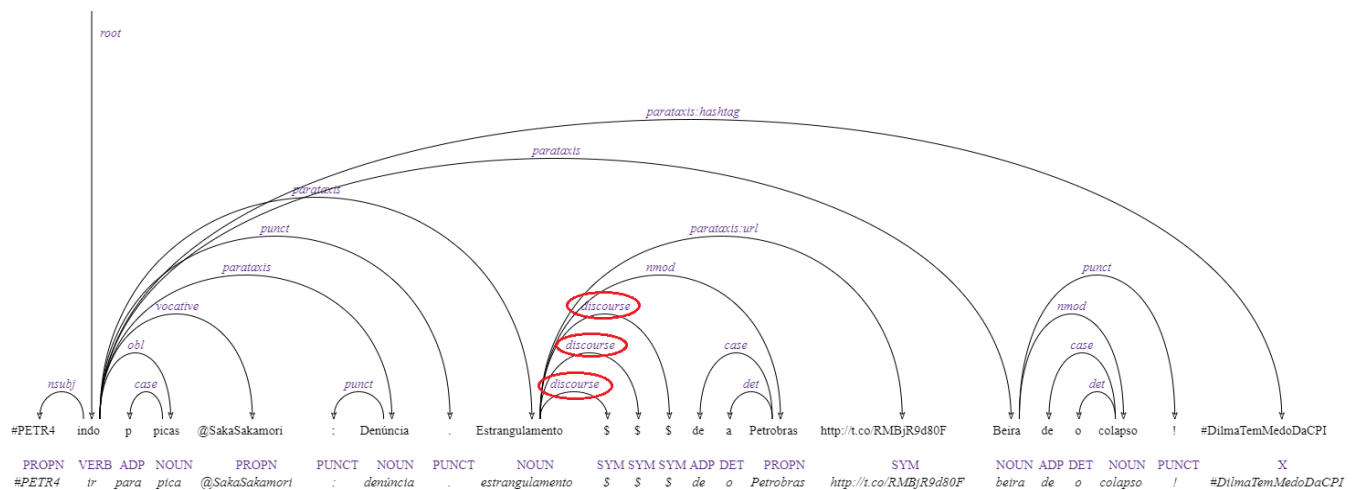


Figura 51 – Cada cifrão repetido em sequência conectado por **discourse** ao *head* (exemplo 50).

Sintaxe não-padrão

Sujeito (nsubj) separado do predicado (*root*) por dois pontos

Exemplo:

- (51) @jcvolemos VALE5 : Sentiu a resistência de 29,90 e lateralizou . Acima de 29,90 pode chegar a 31,33 e abaixo de 29,02 pode chegar a 27,59 .

(Interpretação: “[...] VALE5 sentiu a resistência de 29,90 e lateralizou . [...]”.)

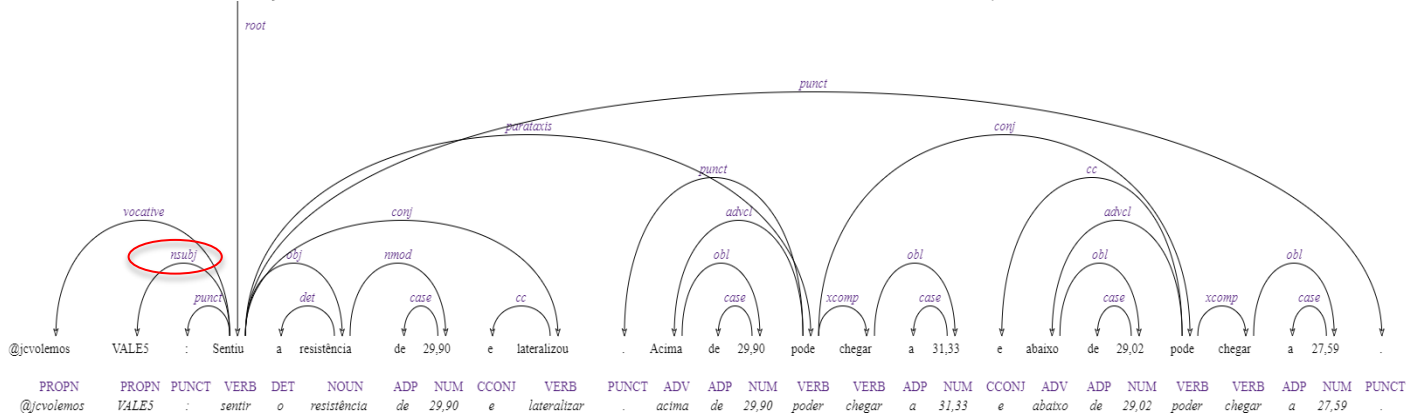


Figura 52 – Atribuição de **nsubj** com sujeito separado do *root* por dois pontos (exemplo 51).

Exemplo 2:

- (52) **#CRUZ3 : pode estar revertendo tendência de baixa após romper canal .**
 Melhor se superar os 23.52 . #Whoknows ? <http://t.co/tZVAMpMQVG>

(Interpretação: “CRUZ3 pode estar revertendo tendência de baixa após romper canal. [...]”.)

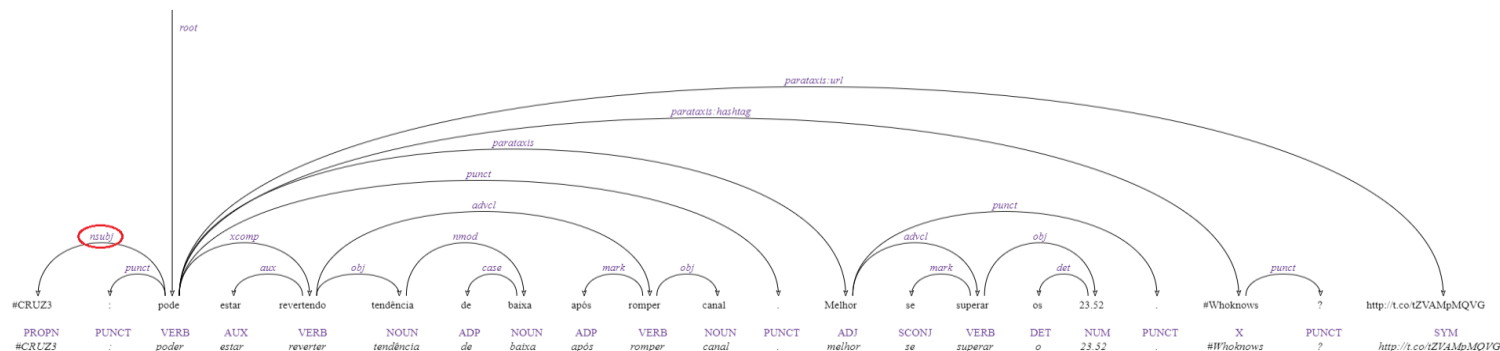


Figura 53 – Atribuição de **nsbj** com sujeito separado do **root** por dois pontos (exemplo 52).

Advcl sem predicado

Como mencionado, para que a anotação sintática (via *deprel*) de um *tweet*, é preciso, por vezes, inferir uma interpretação do *post*. Esse é o caso do trecho em negrito do *tweet* abaixo, do qual se infere a expressão “se estiver” em “(se estiver) acima de 29,90” e “(se estiver) abaixo de 29,02”, o que permite, de acordo com Duran (2022), anotar “acima_ADV (de 29,90)” como dependente por **advcl** de “pode_VERB”, pois **advcl** liga o predicado de uma oração matriz ao predicado de uma oração que a modifica.

Exemplo:

- (53) **@jcvolemos VALE5 : Sentiu a resistência de 29,90 e lateralizou .** **Acima de 29,90 pode chegar a 31,33 e abaixo de 29,02 pode chegar a 27,59 .**

(Interpretação: [**@jcvolemos VALE5** sentiu a resistência de 29,90 e lateralizou. Pode chegar a 31,33 se estiver acima de 29,90 e pode chegar a 27,59 se estiver abaixo de 29,02].)

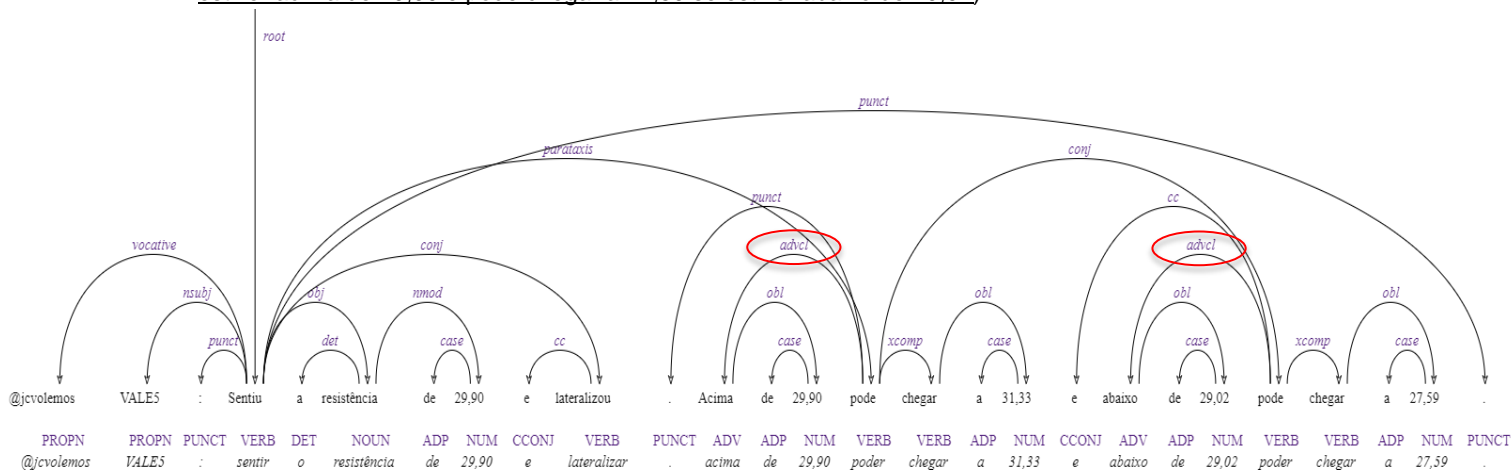
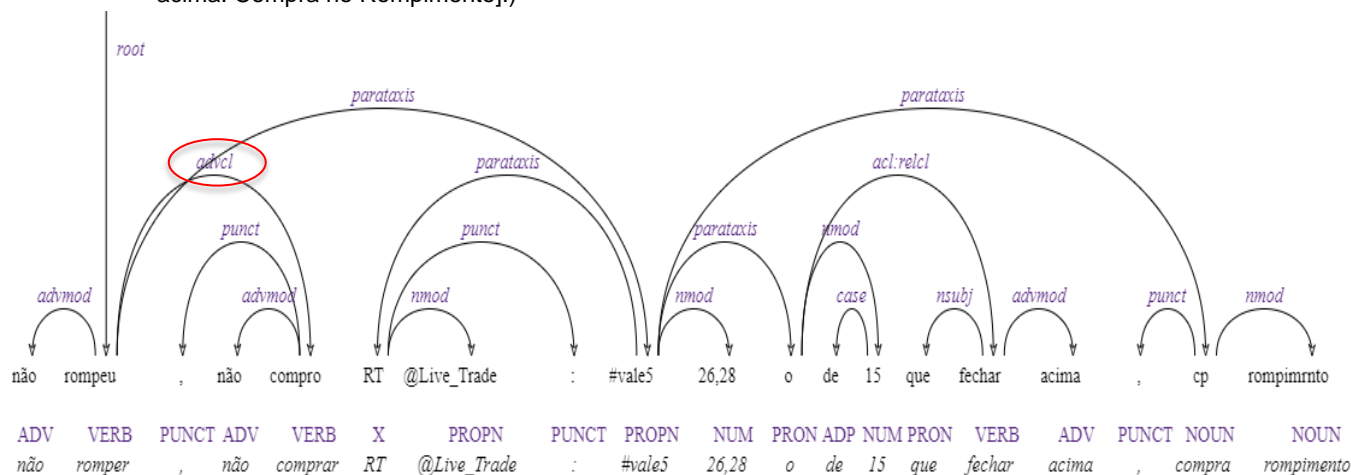


Figura 54 – Atribuição de **advcl** sem predicado (exemplo 53).

Exemplo:

- (54) não rompeu, não compro RT @Live_Trade : #vale5 26,28 o de 15 que fechar acima , cp rompimrnto.

(Interpretação: Não rompeu, portanto, não compro. [RT @Live_Trade: #Vale a 26,28. O de 15 que fechar acima. Compra no Rompimento].)



Coordenação (conj) introduzida por pontuação ou símbolo

Sinais de pontuação (PoS PUNCT), como barras inclinada (“/”) e barra vertical simples (“|”) ou dupla (“| |”) e o símbolo (PoS SYM) “X”, podem funcionar como conjunção quando a interpretação do *post* sugerir que eles estejam introduzindo uma coordenação.

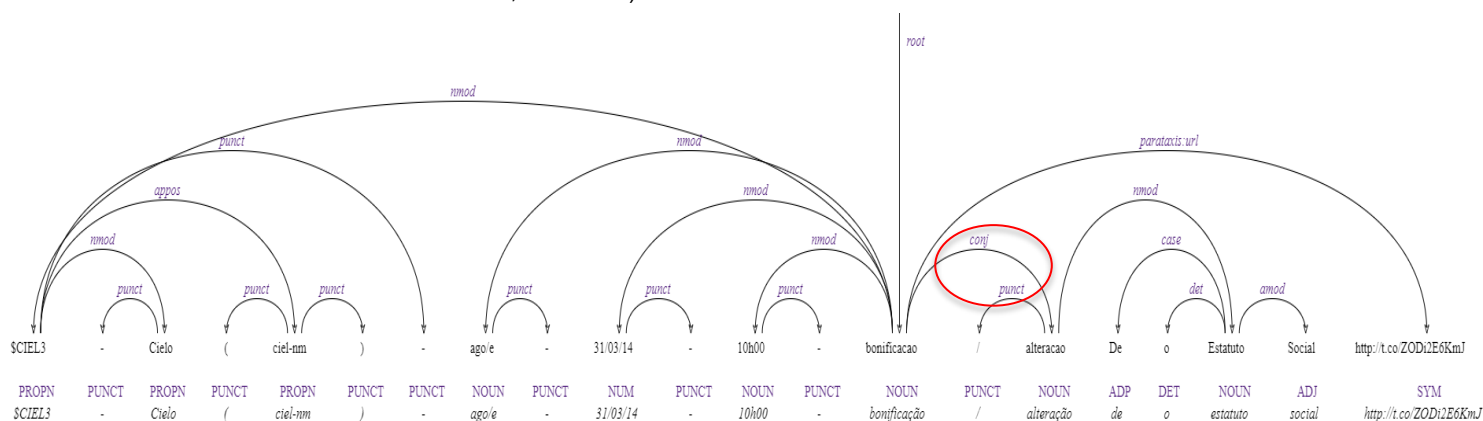
Nesses casos, a *deprel* entre o elemento coordenado e PUNCT é **punct**, e entre o elemento coordenado e SYM é **cc**.

Em ambos os casos, a *deprel* entre os elementos coordenados, seja por PUNCT ou SYM, é **conj**.

Exemplo (com barra inclinada):

- (55) \$CIEL3 – Cielo (ciel-nm) – ago/e – 31/03/14 – 10h00 – bonificacao / alteracao De o Estatuto Social <http://t.co/ZODi2E6KmJ>

(Interpretação: [...] Bonificação e alteração do estatuto social (decididas) na assembleia geral ordinária ou extraordinária do dia 31/03/14, às 10h00.)



Exemplo (com barra vertical simples e dupla):

(56) RT @dividendo_br : ELETROBRAS jscp | aprov 30/04/2014 | ex 02/05/2014 | | pg n/d |
ELET3 R\$ 0,399210837 | ELET5 R\$ 2,178256587 | ELET6 R\$ 1,63369244 htt ...

(Interpretação: a respeito dos juros sobre o capital próprio (jscp) da ELETROBRAS : aprovação em 30/04/2014; ex em 02/05/2014; pagamento indefinido, sendo: R\$ 0,399210837 para ELET3 e R\$ 1,63369244 para ELET6; fonte: <http://t.co/ljtHKLQlfr>)

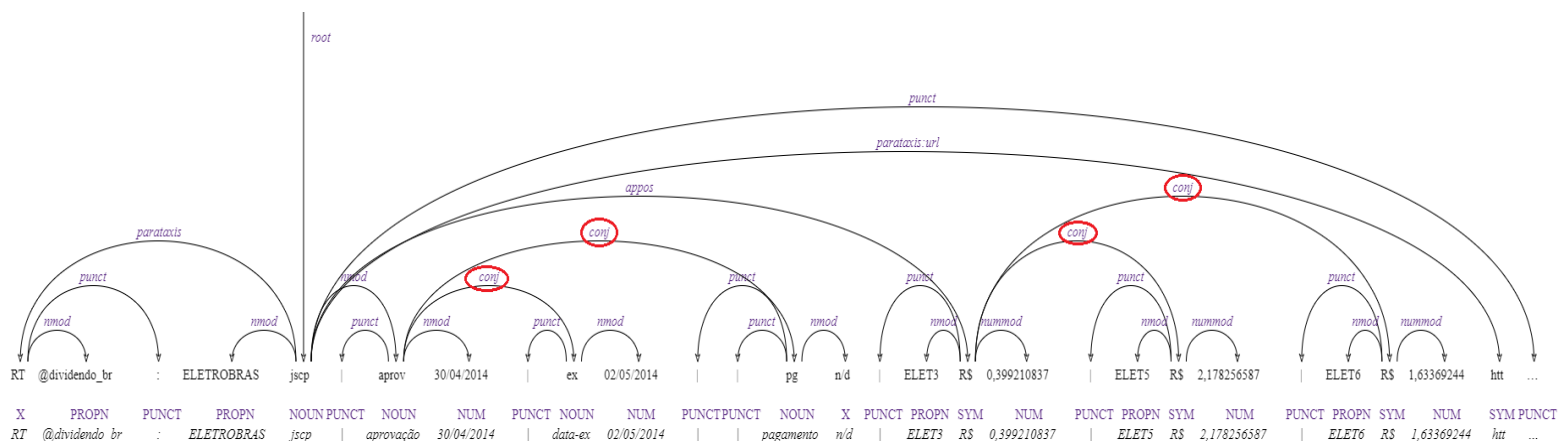


Figura 57 – Atribuição de **conj** introduzida por pontuação (barra vertical simples e dupla) (exemplo 56).

Exemplo (com símbolo):

(57) LONG&SHORT de PETR4 x PETR3 indo MTOOOO BEM !!!

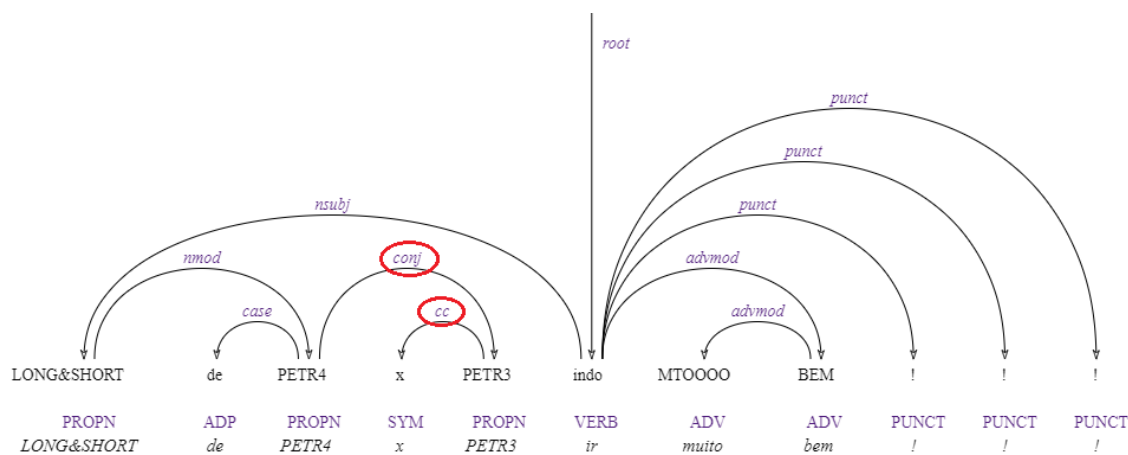


Figura 58 – Atribuição de **conj** introduzida por símbolo (SYM) com **cc** (exemplo 57).

Aposição (appos) sinalizada por PUNCT “/” (barra inclinada)

Padrão: **PUNCT “/” como elemento de aposição (appos)**, quando:

1. O dependente (elemento da direita) especifica ou descreve o **head** (elemento da esquerda)

Exemplo:

(58) \$LIGT3 - Light S/a (ligt-nm) - Aviso A os Acionistas / Distribuicao De Dividendo
<http://t.co/a8VHtsh8Xw>

(Interpretação: Aviso aos acionistas (: haverá) distribuição de dividendo [...])

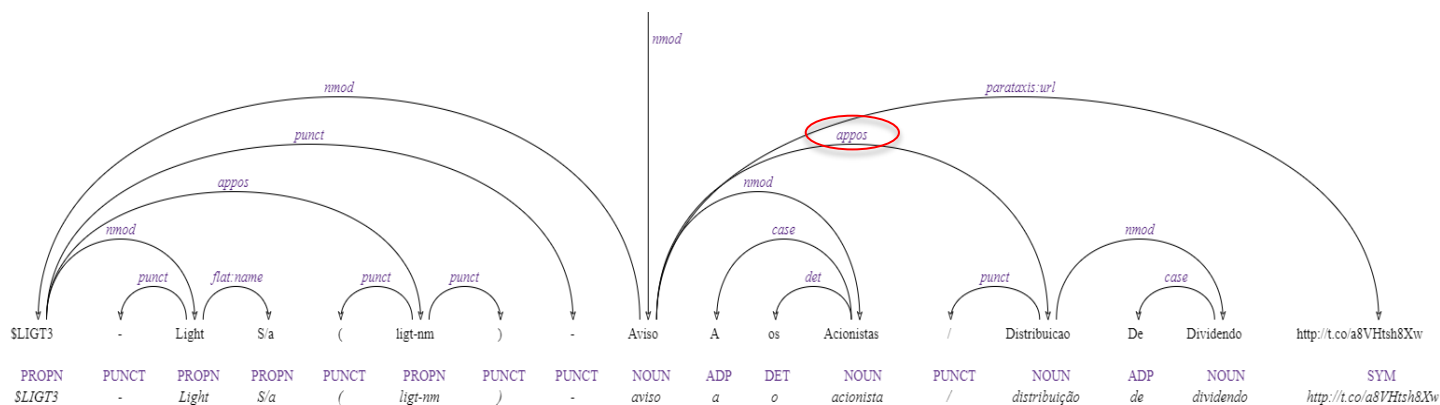


Figura 59 – Atribuição de **appos** introduzida por pontuação (barra inclinada).

Root expreso por símbolo (SYM)

Em certos contextos, o SYM “>” pode ser interpretado como verbo, funcionando como **root**. No exemplo da Figura 60, o referido símbolo foi “traduzido” para “foi (a)” (PoS VERB ADP), ao qual “R\$” foi conectado por **obj** (segundo argumento *core* do predicado).

Exemplo:

(59) último dia de o Gov FHC , a ação Petrobras (PETR3) > R\$ 3,3 @geraldoAlckmin_ diz q ações de a Petrobras “ viraram pó ” . <http://t.co/vAdFc65rmh>

(**Interpretação:** No último dia do governo FHC, a ação Petrobras (PETR3) **foi a** R\$3,3 e Geraldo Alckmin diz que as ações da Petrobras viraram pó.).

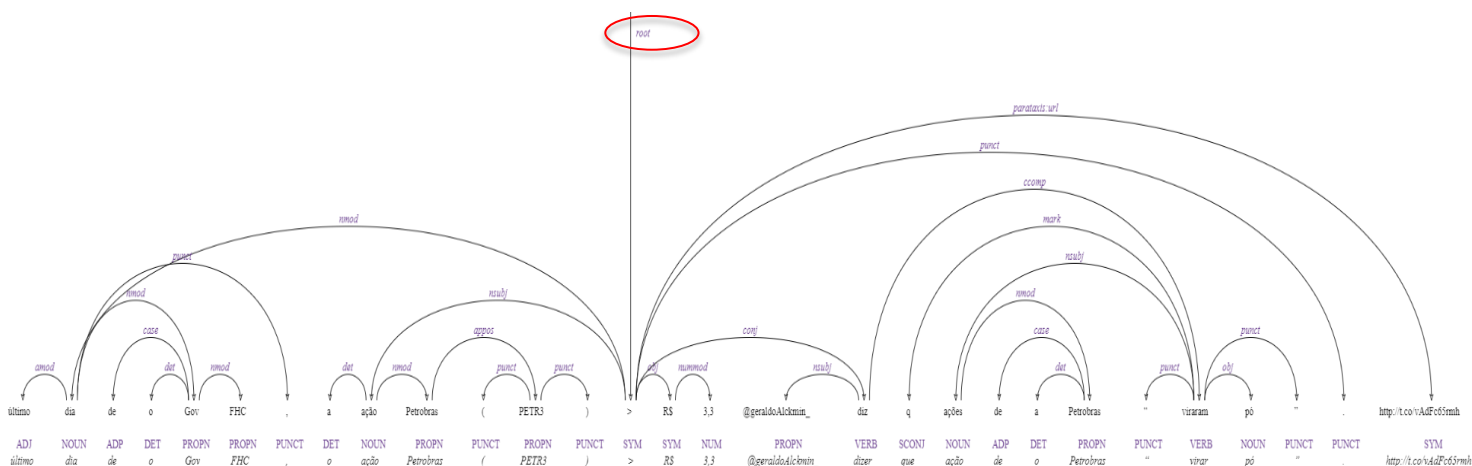


Figura 60 – Atribuição de **root** para símbolo (“>”).

Deprel por inferência

Inferência de conjunção (conj)

No trecho em negrito do exemplo a seguir, infere-se uma coordenação (**conj**) entre “Petrobrás” (representada pela *hashtag* #PETR4) e “Pasadena”, mesmo sem conjunção presente. A seta que a representa **conj** parte do primeiro elemento da série (*head*) em direção a cada dependente.

Exemplo:

- (60) **Ôrrrra , finarmenti , hein ! ! | FHC muda discurso e diz que apoia CPI para investigar #PETR4 #Pasadena** <http://t.co/33WJ6NBvWW> via @estadao

(Interpretação: “[...] FHC muda discurso e diz que apoia CPI para investigar **Petrobrás e Pasadena.**”)

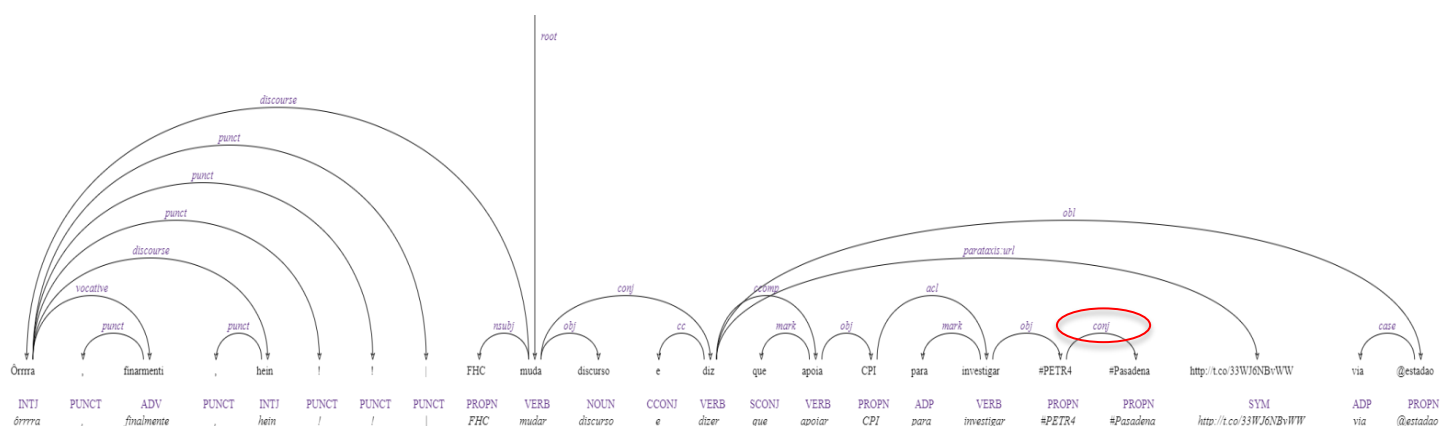


Figura 61 – Atribuição de **conj** sem ocorrência de conjunção (exemplo 60).

Exemplo:

- (61) @KatiaAbreu como assim CPI prejudicará #PETR4 ? imagino q Sra entenda a corrente gravidade e saiba o q seja CPI . @agenciapf @MP_PGR ctz fará

(Interpretação: “[...] . Agência PF (Polícia Federal) **e** o Ministério Público certamente farão.”)

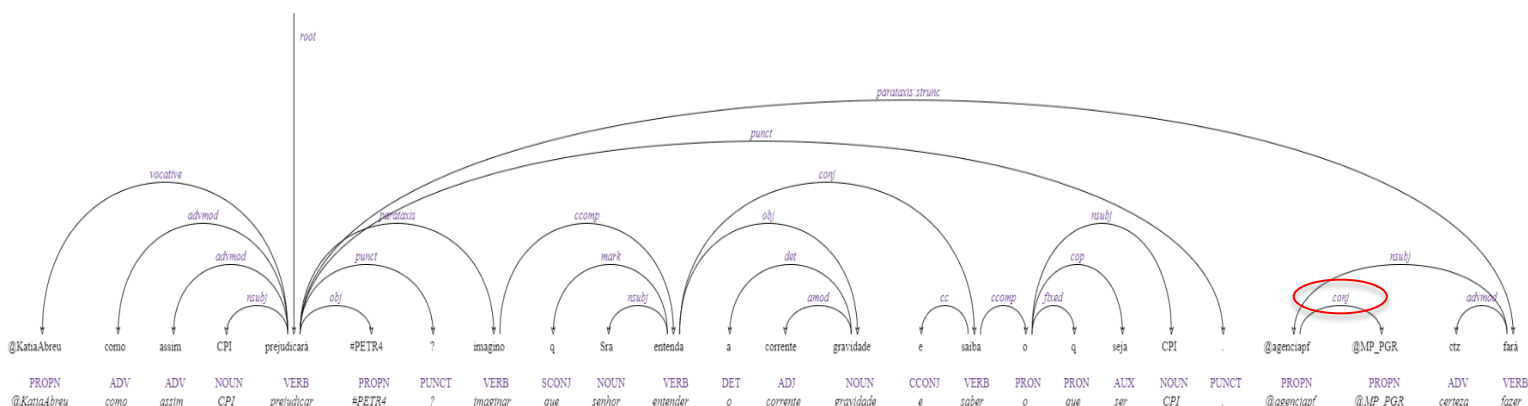


Figura 62 – Atribuição de **conj** sem ocorrência de conjunção (exemplo 61).

Inferência de conjunção e verbo

Exemplo:

(62) RT @reminiscences : Petr4 - CPI exclusiva de a Petrobras , Petr4 deve subir (pois afetará muito Dilma) CPI ampla , Petr4 deve cair , pois espalh ...

(**Interpretação:** **Se for** uma CPI exclusiva da Petrobras, Petr4 deve subir, pois afetará muito Dilma; **se for** uma CPI ampla, Petr4 deve cair, pois espalh....).

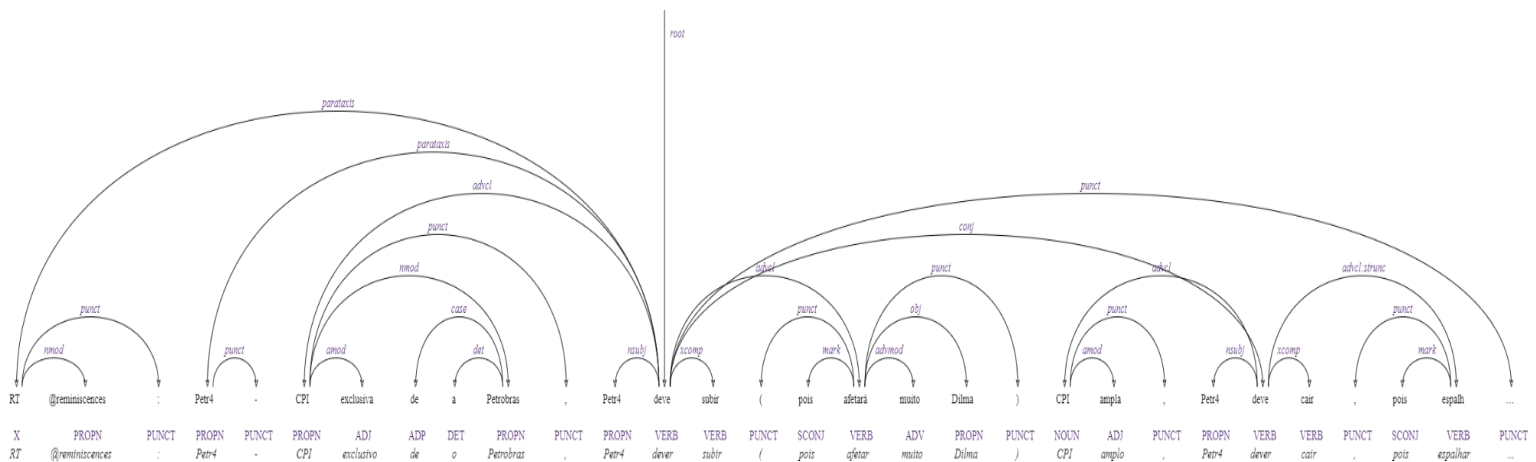


Figura 63 – Atribuição de **advcl** por inferência de conjunção e verbo (exemplo 62).

Exemplo:

(63) BVMF3 cancelado - bateu em o stop antes de acionar

(Interpretação: “BMF3 **foi** cancelado **pois** bateu no stop antes de acionar”) Essa interpretação explica a anotação de “cancelado” como **root**, o qual possui uma **advcl** causal).

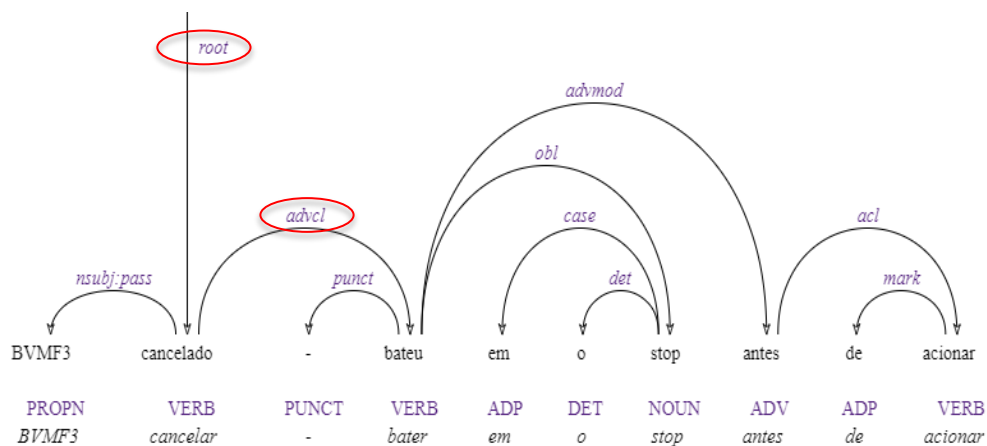


Figura 64 - Atribuição de **root** e **advcl** por inferência de conjunção e verbo (exemplo 63).

Inferência de verbo e preposição

Exemplo:

(64) #cyre3 postado hj antes de a abertura + 1,78

(Interpretação: “#cyre3 **foi** postado *hj* antes de a abertura **a** + 1,78”) Essa interpretação justifica a anotação de “postado” como **root** e de “+ 1,78” como seu dependente por **obl**.

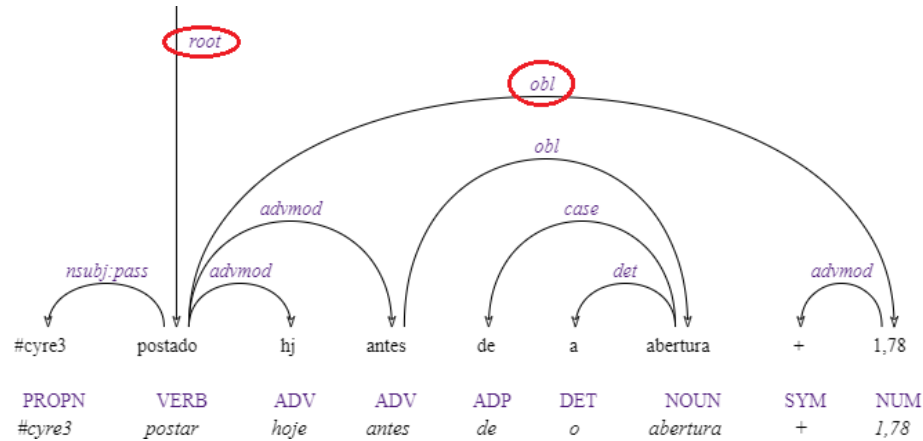


Figura 65 - Atribuição de **root** e **obl** por inferência de verbo e preposição (exemplo 64).

Exemplo:

(65) Day trade VALE5 Previsto e evitado zona de alto risco e falta de liquidez absurda 27-03-14 <http://t.co/FZV3HPrEG3>

(Interpretação: “Day trade VALE5 **foi** Previsto e **foram** evitado(as) zona de alto risco e falta de liquidez absurda **em** 27-03-14”).

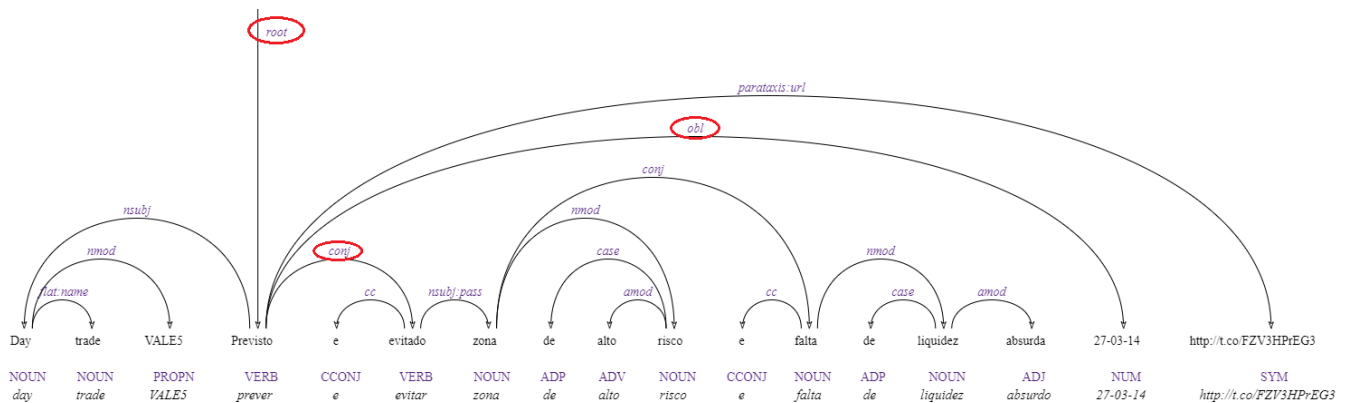


Figura 66 - Atribuição de **deprel** por inferência de verbos (exemplo 65).

(66) @edmilsonpapo10 achei , olha : PETR4 , 2008 , por volta de R\$ 45

[illegible]

Inferências de símbolo como fórmula ou palavra de conteúdo

Muitos *tweets*, por veicularem conteúdo originalmente expresso na forma de tabela ou lista, apresentam estrutura bastante fragmentada. Neles, o símbolo de igual (“=”) parece iniciar uma sequência (trecho sublinhado no exemplo da Figura 68) de valores, *tickers* e outros símbolos (como “\$”) na qual é extremamente complicado identificar uma hierarquia sintática entre os elementos. Assim, opta-se por anotar os elementos dessa sequência iniciada por “=” com **flat**.

(67) Ouro : AEDU3 Fusão só depende de o CADE em jun/14 1 AEDU = $\frac{0,4548 \text{ KROT } 43,7}{0,4548} = 19,87$ 19,87 / 13,13 = 51 % 51 % de espaço pra altas !!

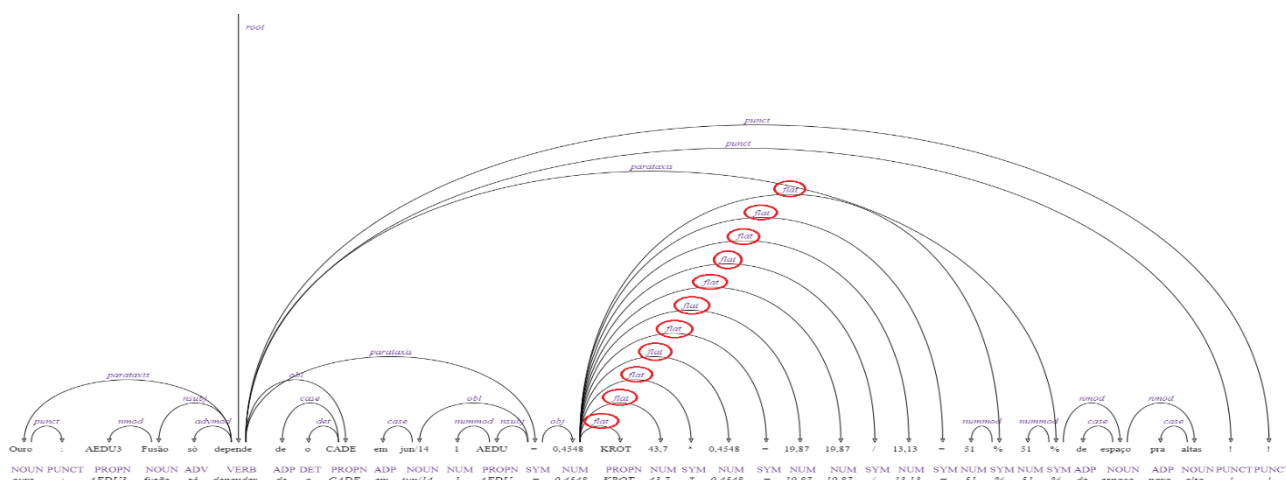
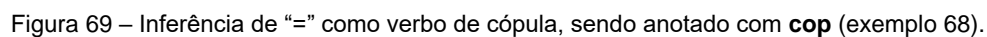


Figura 68 - Atribuição de **flat** a uma sequência de valores, *tickers* e outros símbolos iniciada por “=”.

Exemplo:

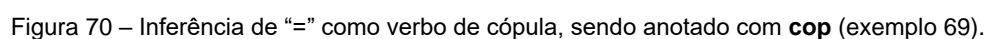
(68) @CaciqueInvest não , aviãozinho = #goll4

(Interpretação: “[...] aviãozinho é #goll4”.)



(69) Ladeira abaixo ! #petr4 = caminhão velho com problemas em os freios descendo a rua
CARREGADO de sardinhas atordoadas ! Vamos a os 10 pila msm ?

(Interpretação: “[...] #petr4 é caminhão velho com problemas nos freios descendo a rua [...]”)



Quanto aos operadores matemáticos “+/-” que compõem os índices de (des-)valorização das ações (por exemplo, +1,4%), os quais estão anotados com a PoS SYM, infere-se que eles substituem as palavras de conteúdo “mais” e “menos”, funcionando como advérbios. Dessa forma, opta-se por anotá-los com **advmod** (e conj se forem binários) (- + * /), e

Exemplo:

(70) Desempenho de as ações de a TIM em a semana passada : TIMP3 (Bovespa) : + 1,4 % , TSU (NYSE) : - 0.53 % , IBOV : - 1,8 % .

(Interpretação: “Desempenho das ações da TIM na semana passada: [...]”)

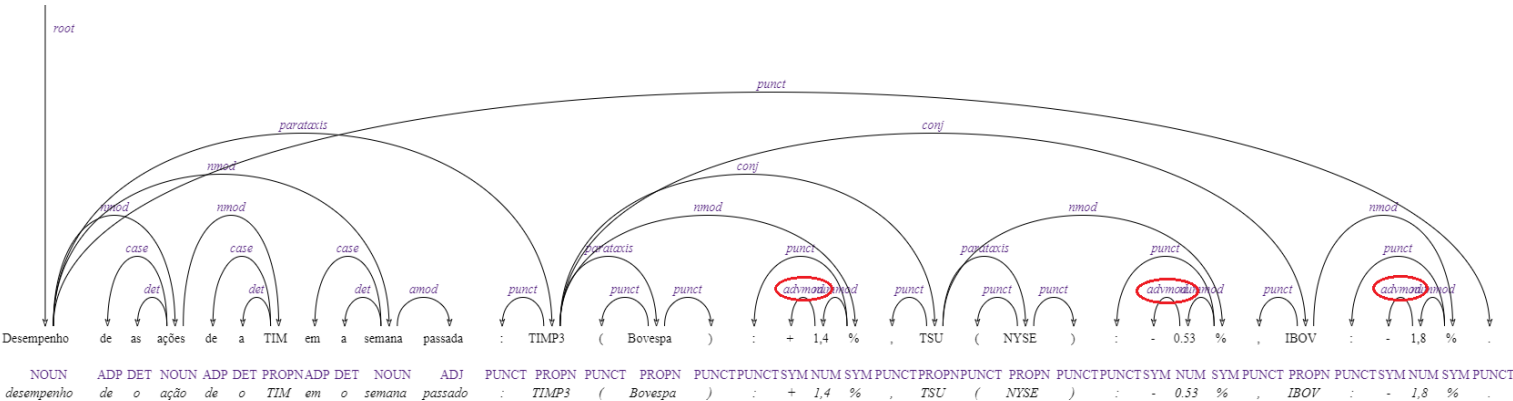


Figura 71 – Inferência de “+/-” como advérbios, sendo anotados com **advmod**.

SEGUNDA PARTE - PADRÕES ESTRUTURAIS

Como mencionado, as diretrizes descritas nesta segunda parte são específicas para certos padrões estruturais frequentes nos *tweets* do *corpus* DANTEStocks.

A maioria das diretrizes são descritas por um *template* composto por:

- padrão estrutural
- lista de elementos constitutivos do padrão com direcionamentos para a anotação de cada um
- subpadrão (se houver)
- ao menos 1 exemplo anotado extraído do *corpus* DANTEStocks

Alguns *templates*, cujos padrões são particularmente fragmentados e/ou que apresentam fenômenos de domínios, também possuem uma interpretação (ou glosa) de seu conteúdo, a qual busca explicitar a interpretação que levou à anotação sintática (e escolha das *deprel*) do exemplo.

Template 1

Padrão: **notas gerais <sentença_truncada> ... <url>**, em que:

- “notas gerais” é dependente por **parataxis** do **root**
- <sentença_truncada> contém o **root**
- ... (sinal de reticências que representa o truncamento) é dependente por **punct** do **root**
- <url> é dependente por **parataxis:url** do **root**
- Subpadrão
 - <sentença_truncada_com_mensagem_completa>**, em que:
 - Embora haja um truncamento, a sentença possui estrutura sintática completa, com análise e anotação UD válidas

Exemplo:

(71) Notas gerais A PETROBRAS (PETR4) concluiu a perfuração de o poço Pitu , localizado em águas profundas de a Bacia ... <http://t.co/DJCj8f8xTH>

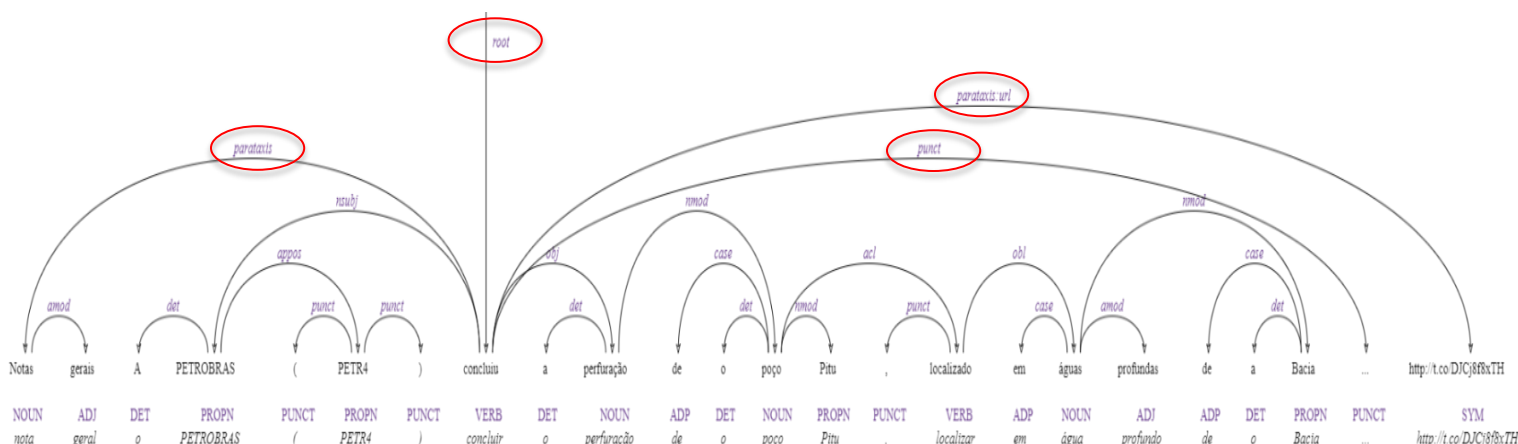


Figura 72 - Anotação do subpadrão 5.1 do *Template* 1.

5.2. <sentença_truncada_incompleta>, em que:

- O truncamento impede a sentença de ter estrutura sintática completa e, por isso, haverá uma *deprel*, dependente de cada caso (pode ser **obj**, **obl**, **nmod**, **ccomp** e **root**), rotulada com a sub-relação **:strunc** (usada para indicar um truncamento no nível sintático, em oposição **:wtrunc**, que indica truncamento de palavras).

Exemplo [**obj:strunc**]:

(72) Notas gerais A BR PROPERTIES (BRPR3) vendeu a a LPP Empreendimentos e Participações , sociedade de o grupo GLP , a ... <http://t.co/Ou2D3dYKDh>

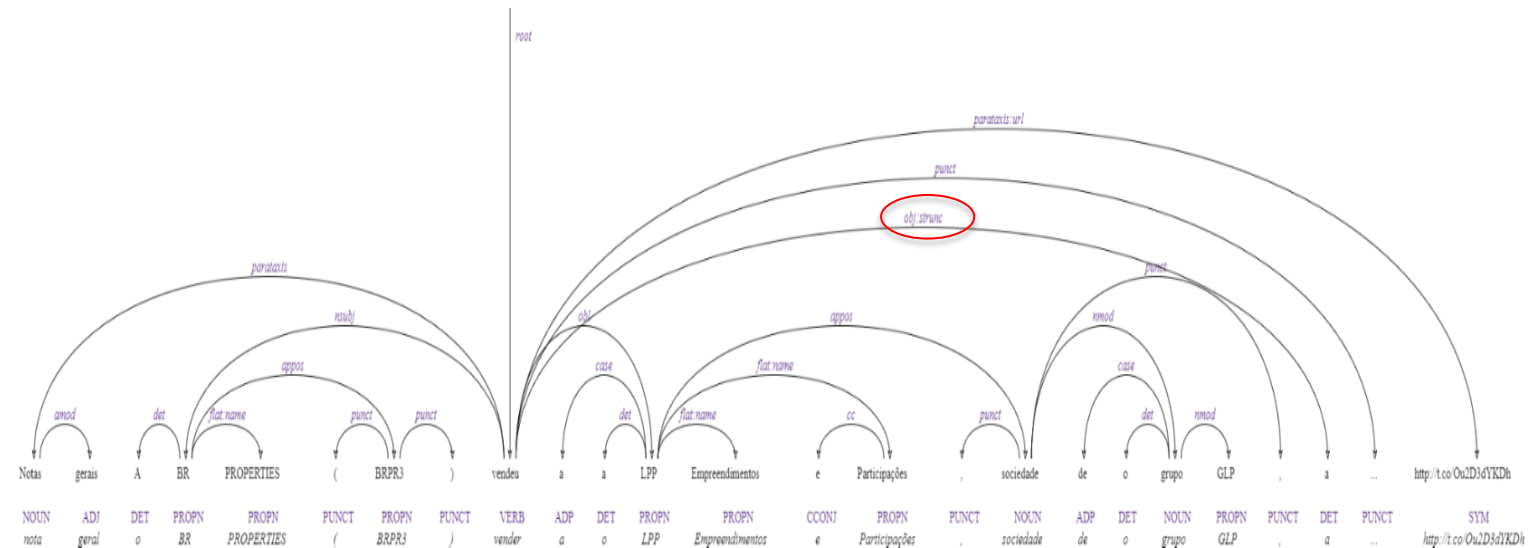


Figura 73 - Anotação do padrão 5.2 com **obj:strunc** do *Template 1*.

Exemplo [**obl:strunc**]:

(73) Notas gerais A LIGHT (LIGT3) pretender distribuir R\$ 32 milhões referentes a o dividendo mínimo obrigatório a os ... <http://t.co/p9jr05Re11>

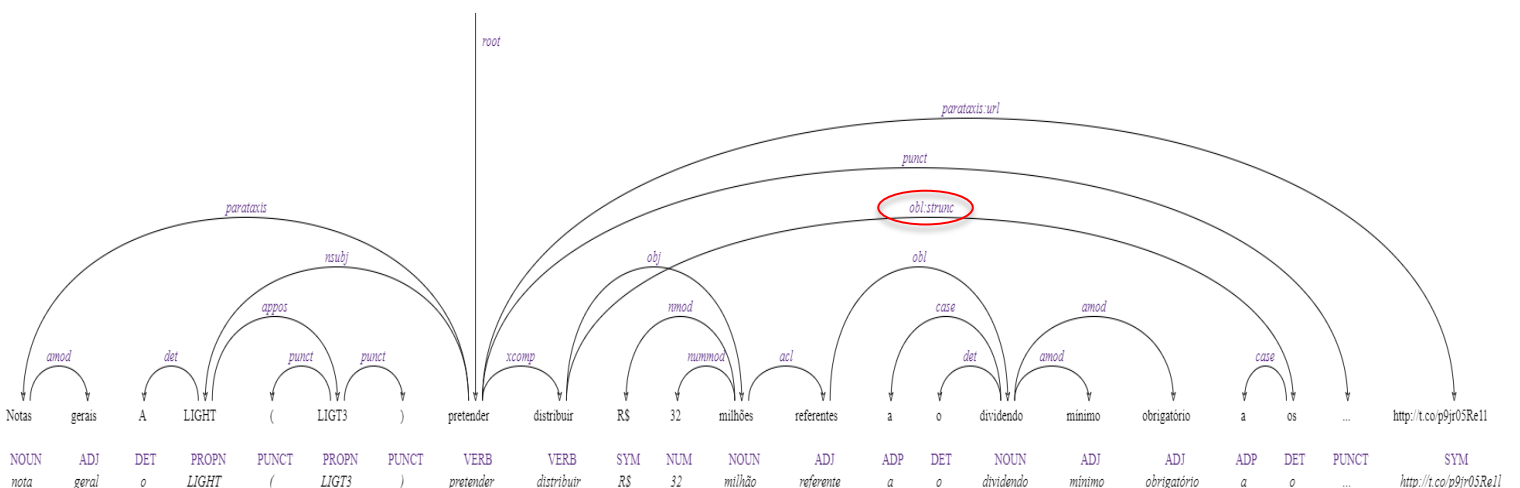


Figura 74 - Anotação do padrão 5.2. com **obl:strunc** do *Template 1*.¹²

¹² Em sentenças completas, não seria válido, pela UD, que um DET fosse dependente de uma relação **obl**, como na Figura 4. Isso só acontece porque DET é o último elemento antes do truncamento. Da mesma forma, um DET tampouco seria *head* de uma relação **case**, porém o truncamento faz com que essa seja uma alternativa para integrar a ADP "a" à árvore.

Exemplo [nmod:strunc]:

(74) Notas gerais A BROOKFIELD (BISA3) não conseguiu instalar a assembleia que definiria a saída de a companhia de a ... <http://t.co/P0Nc4AO2et>

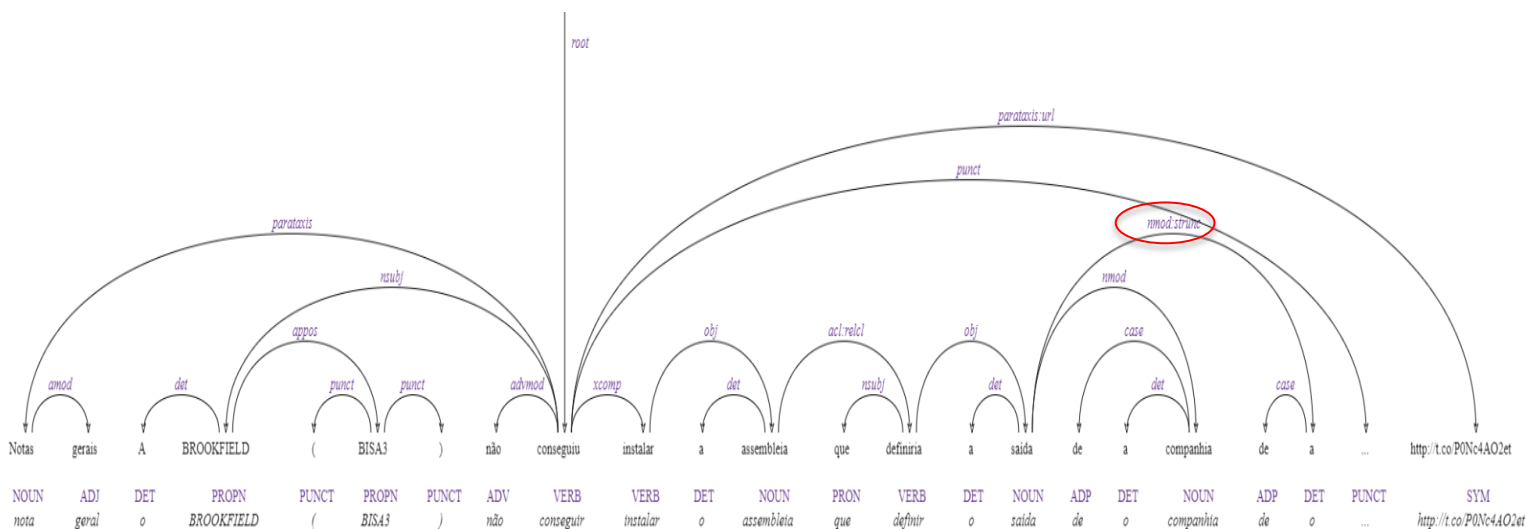


Figura 75 - Anotação do padrão 5.2. com **nmod:strunc** do *Template 1*.

Exemplo [ccomp:strunc]

(75) Notas gerais Por determinação de a justiça , a PETROBRAS (PETR4) precisará divulgar , em até cindo dias , qual é a ... <http://t.co/uGFrqxliWY>

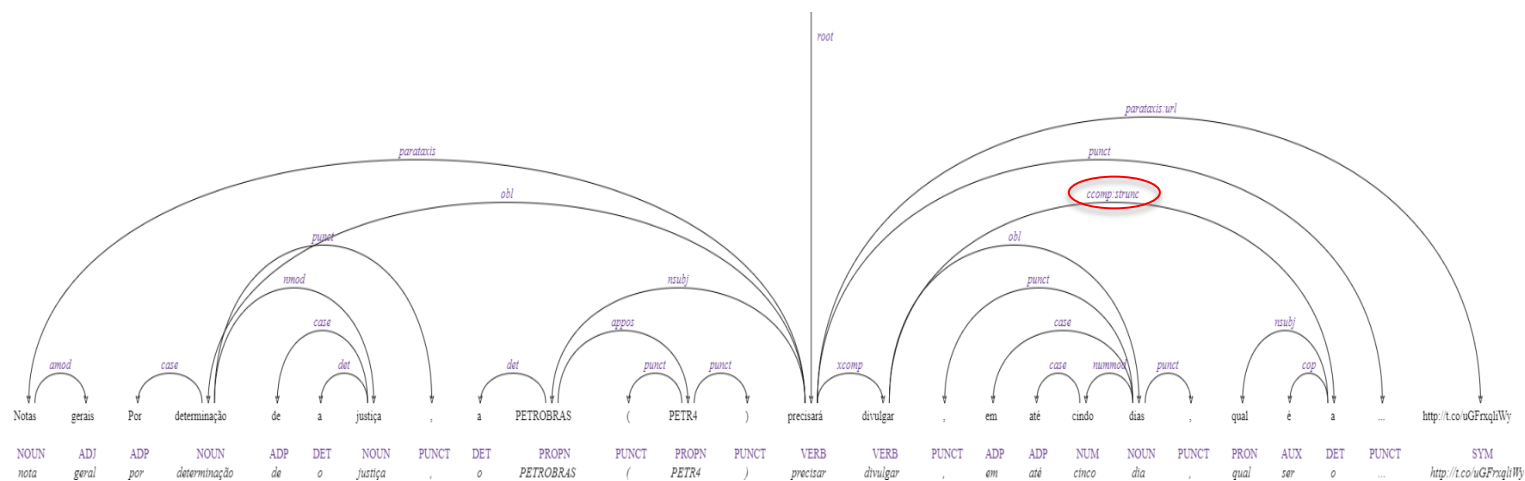


Figura 76 - Anotação do padrão 5.2 com **ccomp:struct** do *Template 1*.

Exemplo [**root:strunc**] (quando a última palavra antes do truncamento for o **root**):

(76) Notas gerais A produção total de petróleo e gás natural de a PETROBRAS (PETR4) em o Brasil em o mês de fevereiro foi ... <http://t.co/qEEHshr7en>

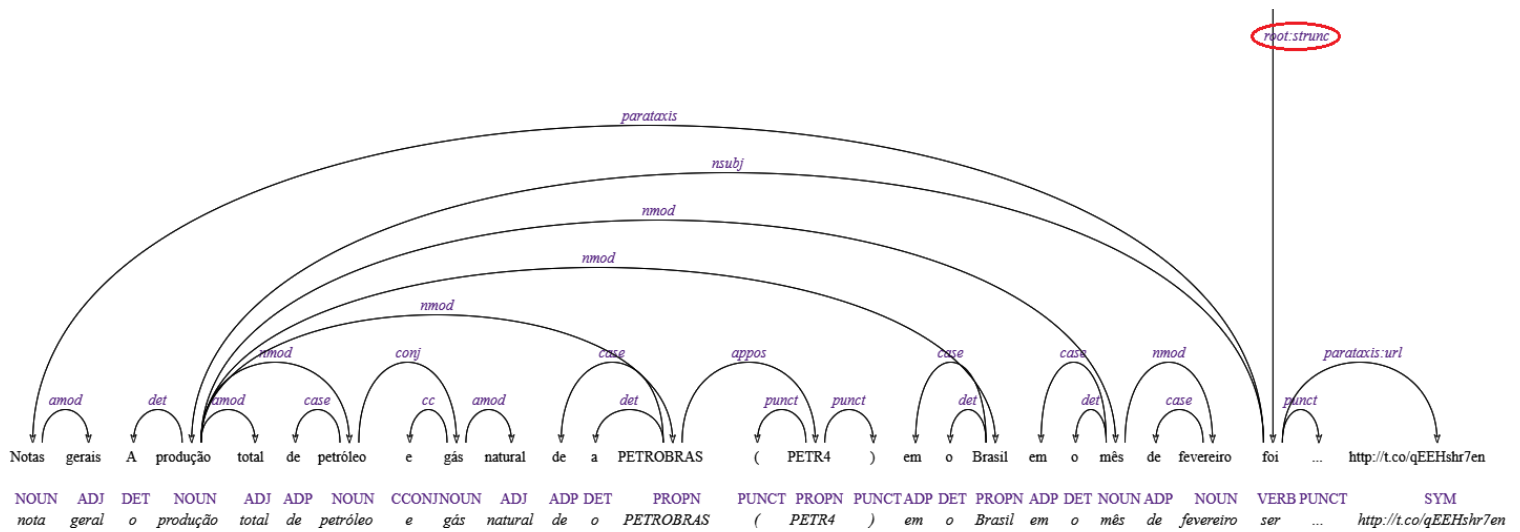


Figura 77 - Anotação do padrão 5.2. com **root:strunc** do *Template 1*.

Template 2

Padrão: **<hashtag-ticker> (mensagem:NNN) <url>**, em que:

1. **<hashtag-ticker>** será sempre **root**
2. (mensagem:NN) é dependente por **parataxis** de **root**
3. **<url>** é dependente por **parataxis:url** de **root**

Exemplo:

(77) **#vale5** (mensagem : 950904) <http://t.co/wfR8HEPu4k>

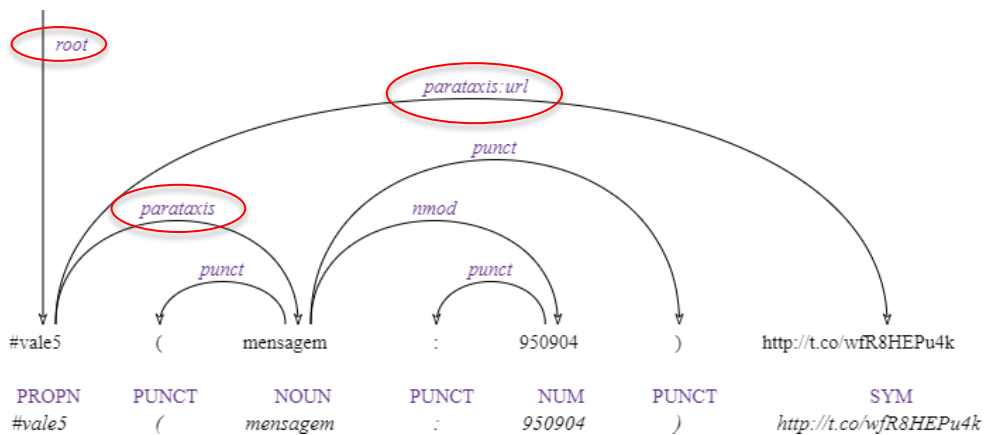


Figura 78 - Anotação do padrão do *Template 2*.

Template 3

Padrão: **<hashtag-ticker>** <complemento> (mensagem:NNN) <url>, em que:

1. *<hashtag-ticker>* é o **root**, a menos que ocupe um papel sintático em *<complemento>*
2. relação entre **root** e *<complemento>* depende da natureza deste último, podendo ser **nmod**, **appos**, **amod**, **advmod**, etc.
3. (mensagem:NN) é dependente por **parataxis** de **root**
4. *<url>* é dependente por **parataxis:url** de **root**

Exemplo: <complemento> do tipo **[nmod]**

(78) #PETR4 - 15 min (mensagem : 951348) <http://t.co/7A5UINu9Mu>

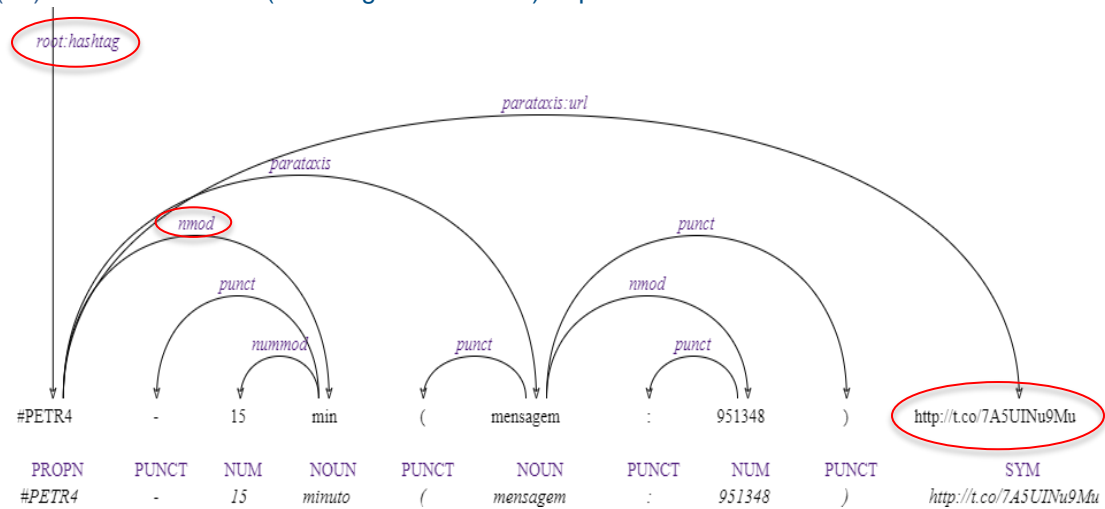


Figura 79 - Anotação do padrão do *Template 3* com <complemento> do tipo **nmod**.

Exemplo: <complemento> do tipo [**appos**]

(79) #BBAS3 Banco de a Brasil (mensagem : 956467) <http://t.co/75T8wtmEXw>

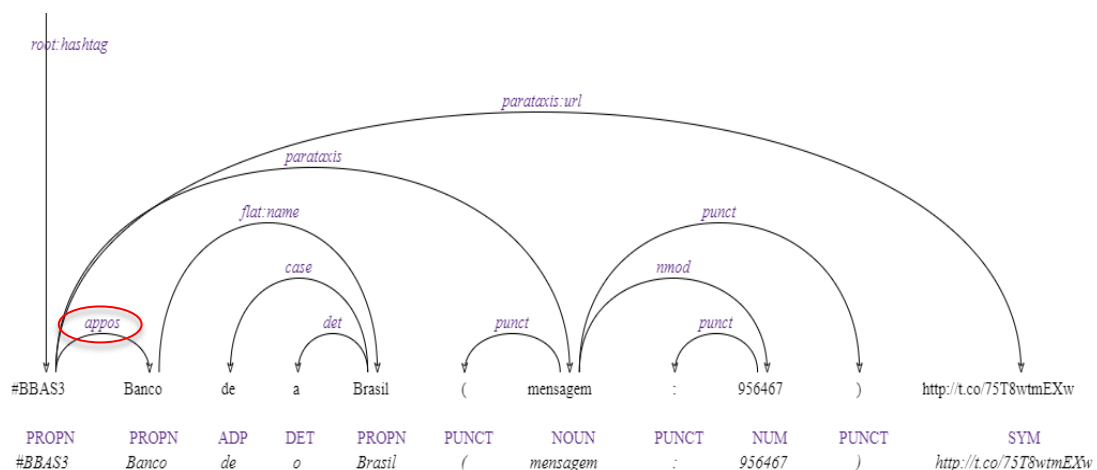


Figura 80 - Anotação do padrão do *Template 3* com <complemento> do tipo **appos.**

Exemplo: <complemento> do tipo [amod]

(80) #csna3 semanal (mensagem : 950998) http://t.co/suRkLOSBUz

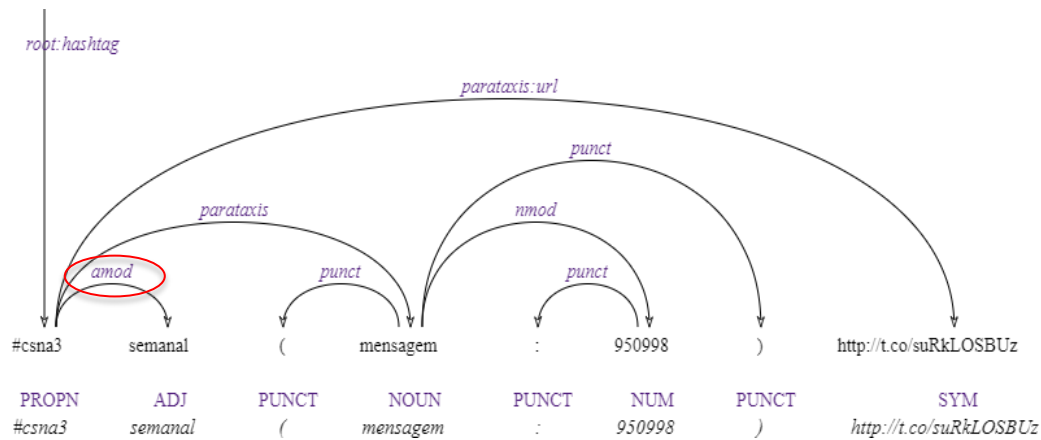


Figura 81 - Anotação do padrão do *Template 3* com <complemento> do tipo **amod**.

Exemplo: <complemento> do tipo [advmod]

(81) #LLXL3 - acima de 1 (um) (mensagem : 952921) http://t.co/11sdL24xTr

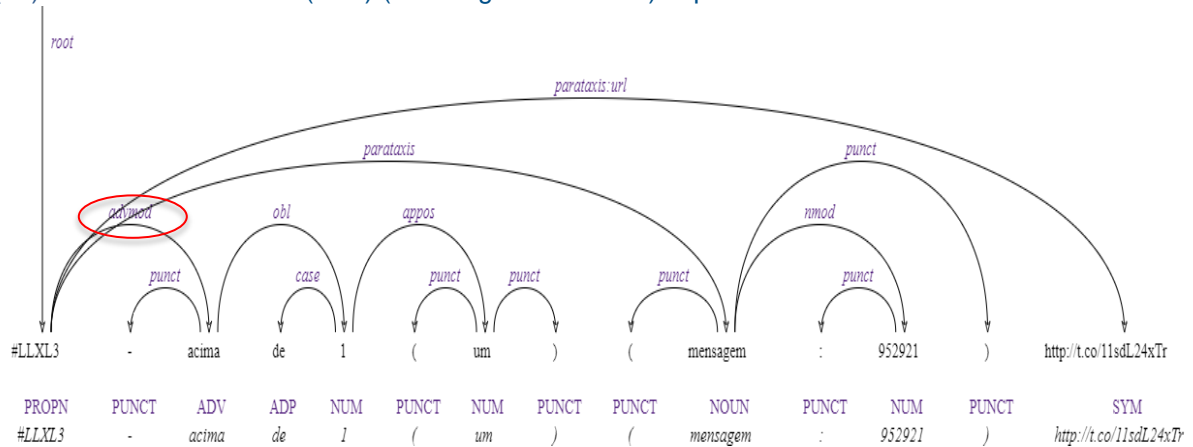


Figura 82 - Anotação do padrão do *Template 3* com <complemento> do tipo **advmod**.

Exemplo: <complemento> do tipo [nmod] e [parataxis]

(82) #PETR4 15 min - acho que nao ! (mensagem : 952919) http://t.co/32XqwNSA6Y

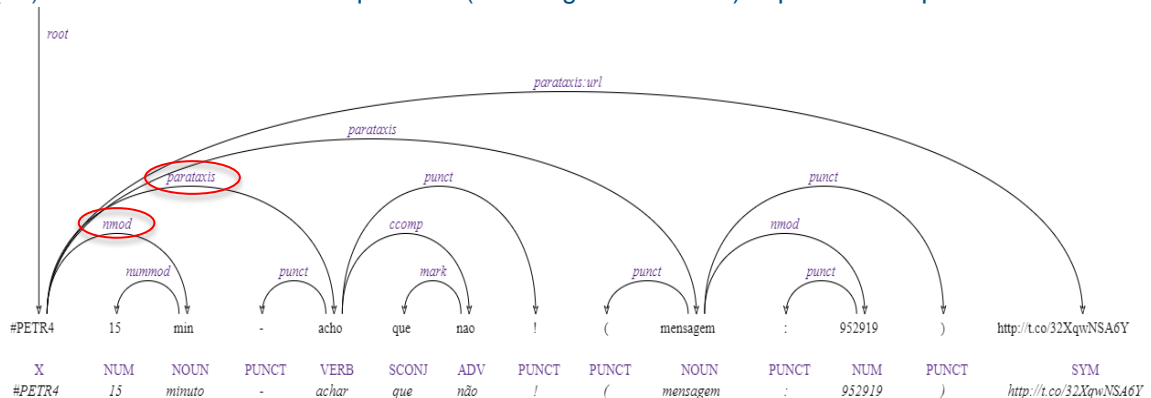


Figura 83 - Anotação do padrão do *Template 3* com <complemento> **nmod** e **parataxis**.

Template 4

Padrão: <sentença> (mensagem:NNN) <url>, em que:

1. <sentença> contém o **root**
2. (mensagem:NN) é dependente por **parataxis** de **root**
3. <url> é dependente por **parataxis:url** de **root**

Exemplo:

(83) DANDO ZOOM EM A #OIBR4 (mensagem : 954226) http://t.co/Pk10JNN9fv

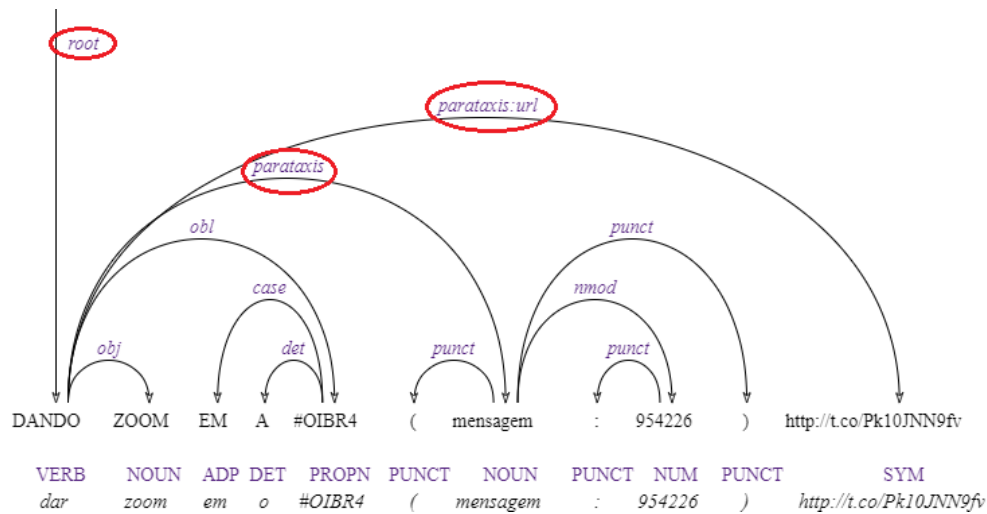


Figura 84 - Anotação do padrão do Template 4.

Template 5

Padrão: <ação&cia> dividendos <lista info&data> <lista ticker&valor> || <url>, em que:

1. **root** é o *token* “dividendos”
2. <ação&cia>: nome da ação e respectiva companhia (**appos**), ligados ao **root** por **nmod**
3. <lista info&data> são 3 elementos coordenados (**conj**) e ligados por **nmod** ao **root**:
|| x <data> || y <data> || z <data>, sendo
x= aprov ou apro; y = dataex ou ex; z = pagto ou pg; data é da forma dd/mm/aa ou n/d (data não definida)
4. <lista ticker&valor> é uma lista de um ou mais itens coordenados (**conj**) e ligados por **appos** ao **root**:
|| <ticker> <valor-reais> || <ticker> <valor-reais> || <ticker> <valor-reais>....
5. <ticker> é dependente por **nmod** de <valor-reais>
6. <url> é dependente por **parataxis:url** de **root**

Exemplo:

(84) ELPL ELETROPAULO dividendos || aprov 25/4/2014 || dataex 28/4/2014 || pagto n/d || ELPL3 R\$ 0,388977082 || ELPL4 R\$ 0,427874790 http://t.co/FhJ4SZQxMA

(Interpretação: a respeito dos dividendos da ELETROPAULO : aprovação em 25/05/2014; ex em 28/04/2014; pagamento n/d, sendo: R\$ 0,388977082 para ELP3 e R\$ 0,427874790 para ELP4; fonte: http://t.co/ljtHKIQlfr)

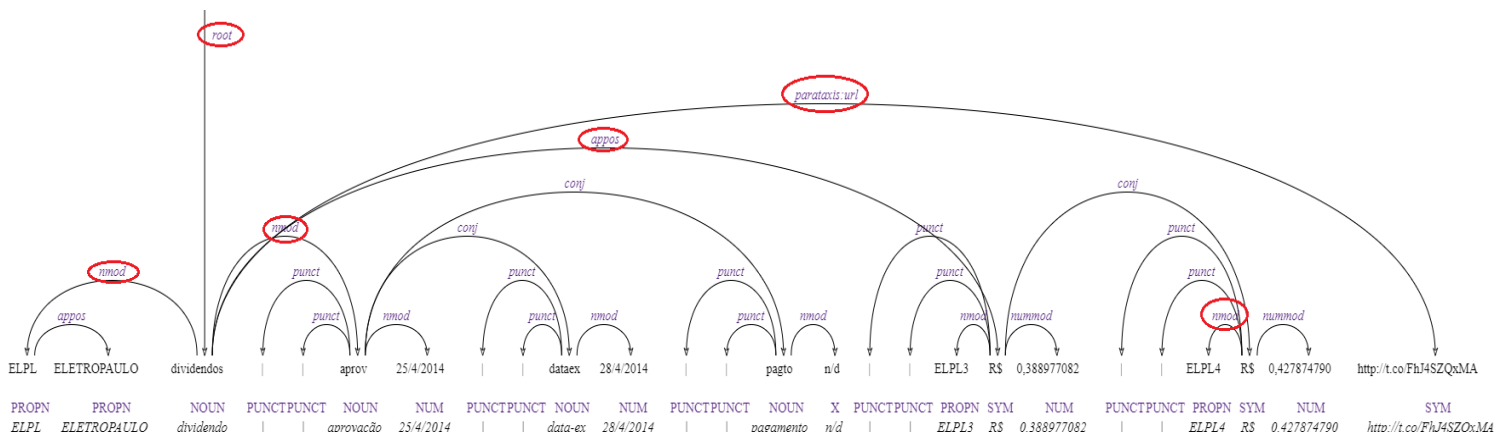


Figura 85 - Anotação do padrão do Template 5.

Template 6

Padrão: <RT @xxxx> : <ação e/ou cia> <dividendos/juros> <lista info&datas> <lista ticker&valor> || <url>, em que:

1. <RT @xxxx> é ligado por **parataxis** ao **root**
2. **root** é o *token* <dividendos/juros>
3. <ação cia>: nome da ação e/ou companhia ligado ao **root** por **nmod**
4. <lista info&data> são 3 elementos coordenados (**conj**) e ligados por **nmod** ao **root**:
|| x <data> || y <data> || z <data>, sendo
x= aprov ou apro; y = dataex ou ex; z = pagto ou pg; data é da forma dd/mm/aa ou n/d (não definida)
5. <lista ticker&valor> é uma lista de um ou mais itens coordenados (**conj**) e ligados por **appos** ao **root**:
|| <ticker> <valor-reais> || <ticker> <valor-reais> || <ticker> <valor-reais>....
6. <ticker> é dependente por **nmod** de <valor-reais>
7. <url> é dependente por **parataxis:url** de **root**

Exemplo (com url trancada, por isso, **parataxis:wtrunc**):

(85) RT @dividendo_br : ELETROBRAS jscp | aprov 30/04/2014 | ex 02/05/2014 | | pg n/d | ELET3 R\$ 0,399210837 | ELET5 R\$ 2,178256587 | ELET6 R\$ 1,63369244 htt ...

(Interpretação: **a respeito dos** juros sobre o capital próprio (jscp) da ELETROBRAS : aprovação em 30/04/2014; ex em 02/05/2014; pagamento indefinido, **sendo**: R\$ 0,399210837 para ELET3 e R\$ 1,63369244 para ELET6; **fonte**: <http://t.co/ljtHKlQlfr>)

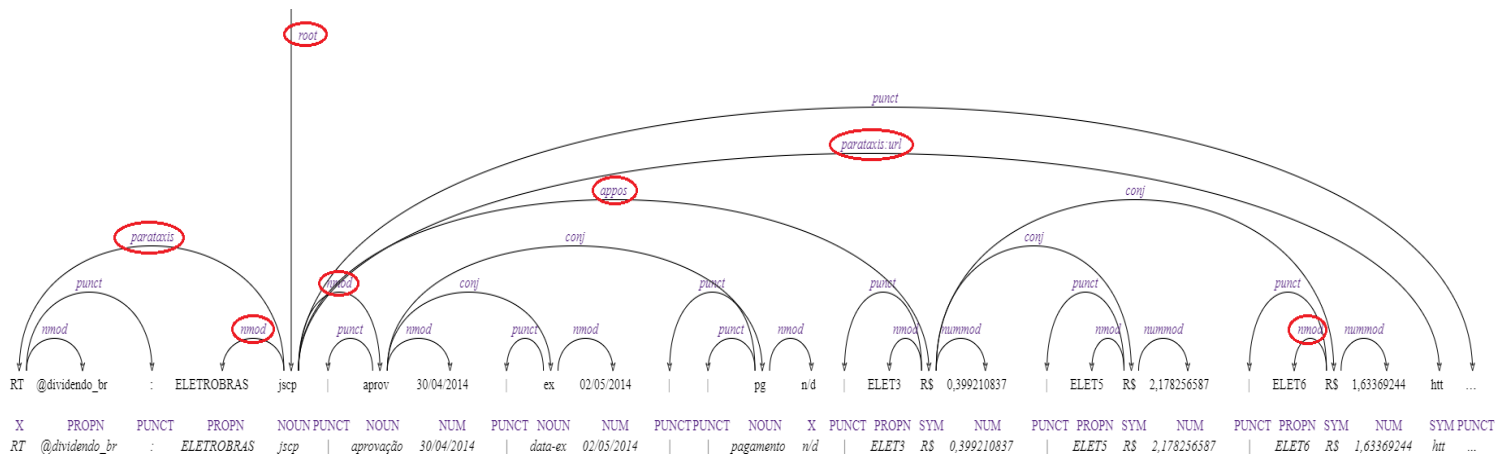


Figura 86 - Anotação do padrão do Template 6.

Template 7

Padrão: <sentença> : <porção_truncada> ... <url> [hashtag opcional], em que:

1. <sentença> tem a mensagem principal e contém o **root**
2. <porção_truncada> é ligada ao **root** por **parataxis:strunc**
3. o **head** da <porção_truncada> dependerá de cada caso; <porção_truncada> pode terminar com uma palavra truncada (**wtrunc**)
4. <url> é dependente por **parataxis:url** de **root**
5. [hashtag opcional] lista de uma ou mais *hashtags* ao final do *tweet*; cada uma delas conectada por **parataxis:hashtag** ao **root**

Exemplo:

(86) 'Salvação' de a OGX e OSX pode estar a a caminho , mais 5 empresas estão em o radar : Nível de o reser ... <http://t.co/WYotABIDa5> #infomoney #vale5



Figura 87 - Anotação do padrão do *Template 7* (exemplo 86).

Exemplo:

- (87) O que há de melhor em a Bovespa : as ações mais indicadas por os analistas : A Vale ficou em prim ... <http://t.co/tkdUiSqQUs> #infomoney #vale5

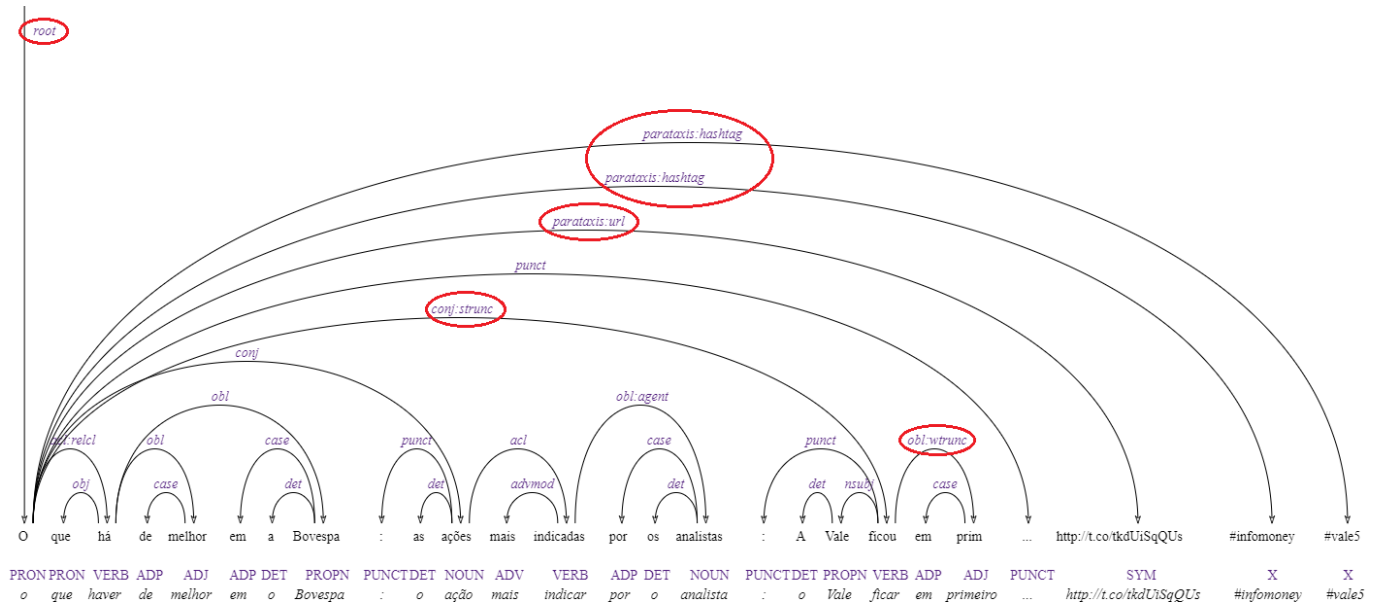


Figura 88 - Anotação do padrão do Template 7 (exemplo 87).

Exemplo 3:

- (88) EM TEMPO REAL : Bancos caem , Vale mostra recuperação , Petrobras sobe 2 % : Mais informações em breve <http://t.co/nQ3FofV6Jb> #infomoney #vale5

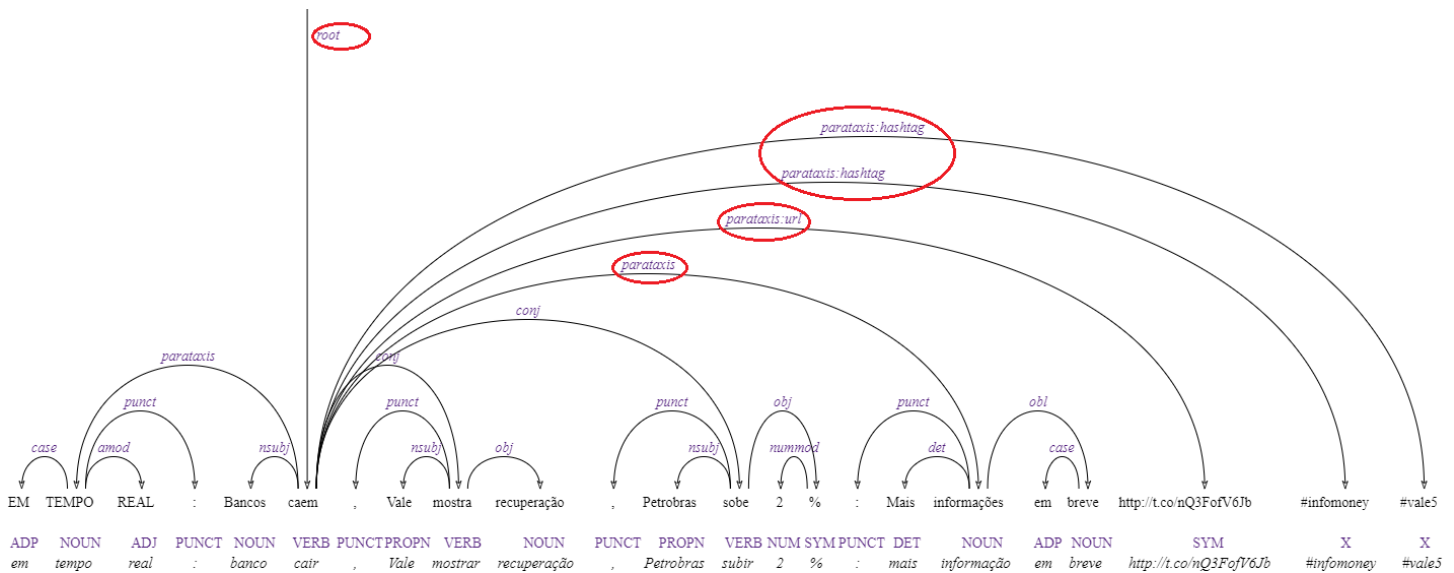


Figura 89 - Anotação do padrão do Template 7 (exemplo 88)

(89) Usiminas e CSN caem 5 % , Petrobras e Vale recuam quase 3 % , uma small cap ganha destaque : Aind ... <http://t.co/7vc5lBJ30> #infomoney #vale5



Padrão : **Rastreamento ações – Gráfico diário – 14h. Analise se romper : <lista *ticker&valor*>**, em que:

- Exemplo:

Figura 91 - Anotação do padrão do *Template 8*.

Template 9

Padrão: **Ativo c/ vol Financeiro Superior a a sua MM21 - <hora> : <lista tickers>**, em que:

1. token “Ativo” é **root**
 2. <hora> é dependente por **parataxis** de **root**
 3. As ações (tickers), conectadas por **conj**, estão como **appos** de **root**
- Exemplo:

(91) Ativo c/ vol Financeiro Superior a a sua MM21 - 14h : BBSE3 PINE4

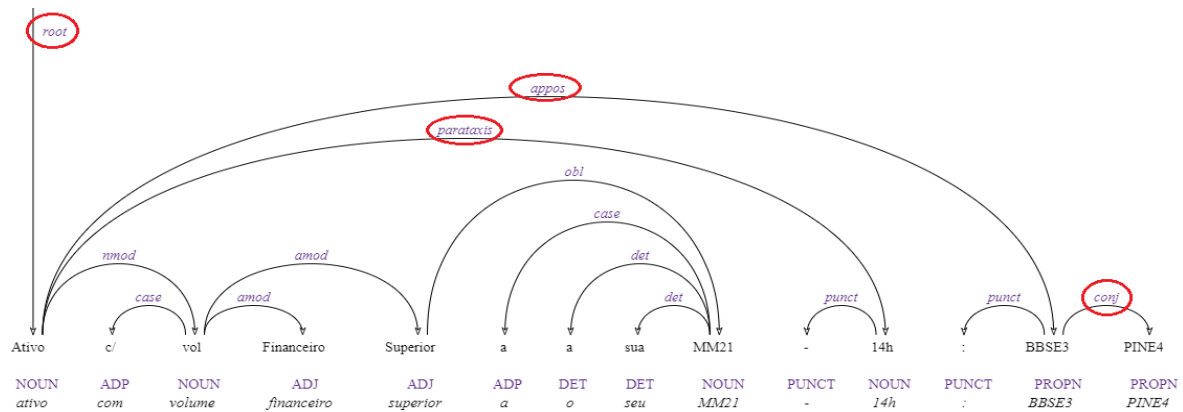


Figura 92 - Anotação do padrão do Template 9.

Template 10

Padrão: **<sentença>. Confira a nova indicação agora em url**, em que:

1. <sentença> é sintaticamente completa e contém o **root**
2. “Confira a nova indicação agora em url” é dependente por **parataxis** de <sentença>, e “Confira” é seu **head**

Exemplo:

(92) A última indicação de a #MRVE3 resultou em - 1.25 % . Confira a nova indicação agora em <http://t.co/kgt1YiTBf7>

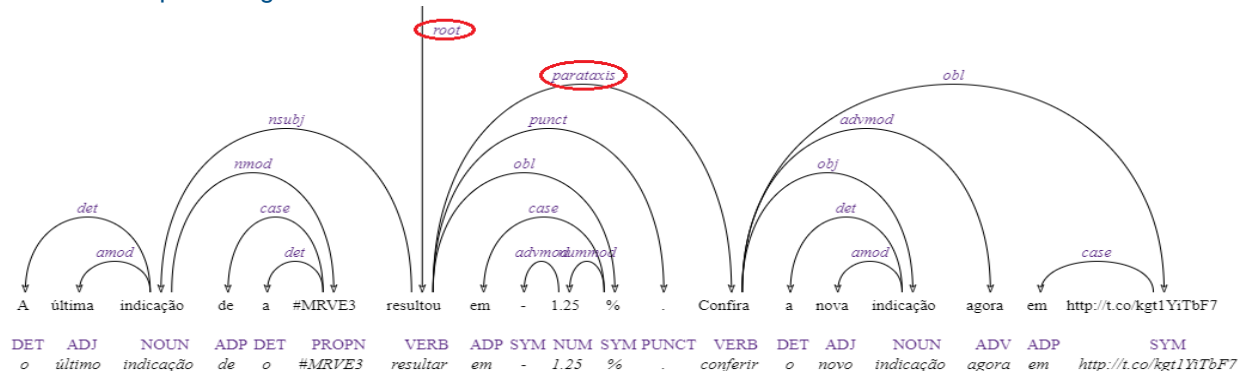


Figura 93 - Anotação do padrão do Template 10.

Template 11

Padrão: <ticker> <tema> <url>, em que:

1. <ticker> é dependente por **nmod** de **root**
2. <tema> contém “suportes e resistências”, sendo “suportes” o **root**
3. <url> é dependente por **parataxis:url** de **root**

Exemplo:

(93) #VALE5 suportes e resistências <http://t.co/c8OrWXrECN>

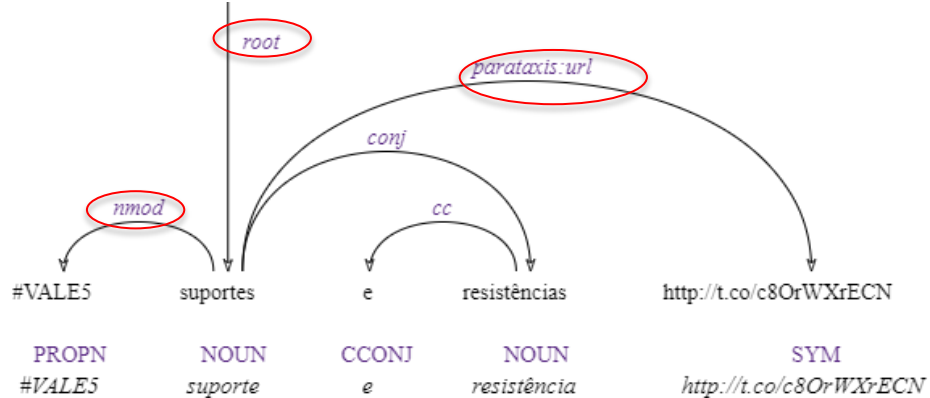


Figura 94 - Anotação do padrão do *Template 11* (exemplo 93).

Exemplo:

(94) #VALE5 suportes e resistências, veja ainda notícia em o comentário.
<http://t.co/sJrLzoBIUT>

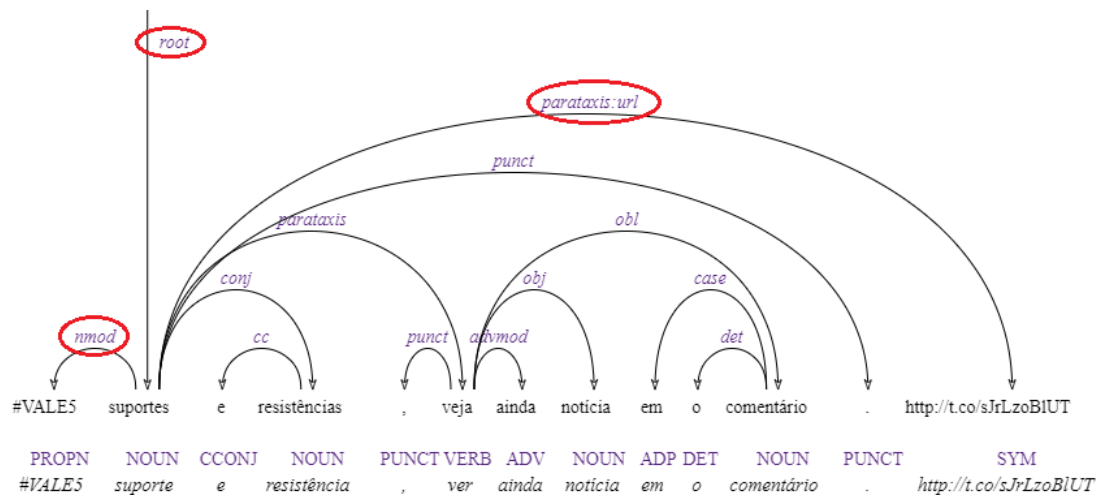


Figura 95 - Anotação do padrão do *Template 11* (exemplo 94).

Template 12

Padrão: **<prefixo> Prepare-se para o próximo pregão! <suífixo> <lista tickers>. Assista!**
url, em que:

1. <prefixo> e <sufixo> são opcionais; quando presentes, podem ser predicados ou sintagmas nominais, e se relacionam por **parataxis** ao **root**
2. **root** é o verbo “Prepare”
3. as ações em <lista *tickers*> estão em coordenação (**conj**) e podem estar em **parataxis** com **root**, se não houver <sufixo>, ou estão relacionadas apropriadamente ao <sufixo>

Exemplo:

(95) Prepare - se para o próximo pregão ! Análise : IBOV , PETR4 , VALE5 , TIMP3 , LPSB3 , MULT3 , GRND3 , LIGT3 e ABCB4 . Assista ! <http://t.co/OVfdQj4UPe>

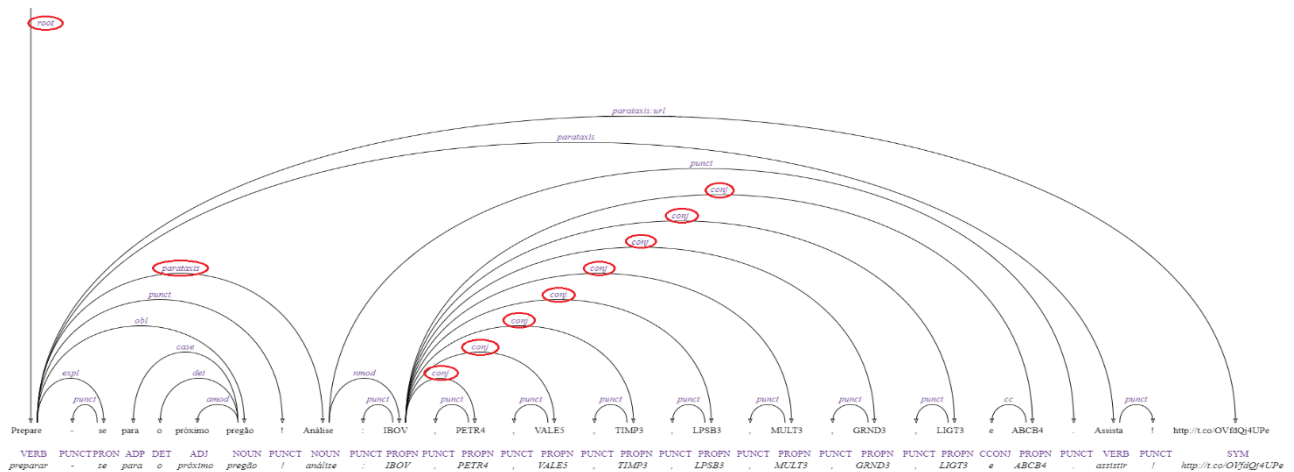


Figura 96 - Anotação do padrão do *Template* 12 (exemplo 95).

Exemplo:

(96) Bom dia ! Prepare - se para o pregão de esta 4ª . Assista : IBOV , PETR4 , VALE5 , EMBR3 , SMLE3 , HYPE3 , CRUZ3 , PMAM3 e NATU3 . <http://t.co/SbHM6qlUrO>

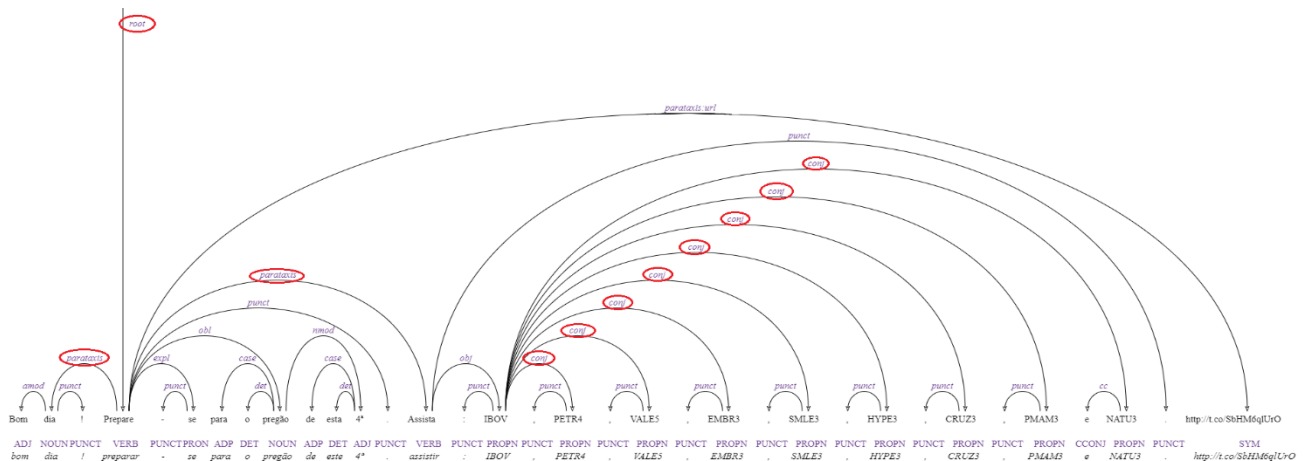
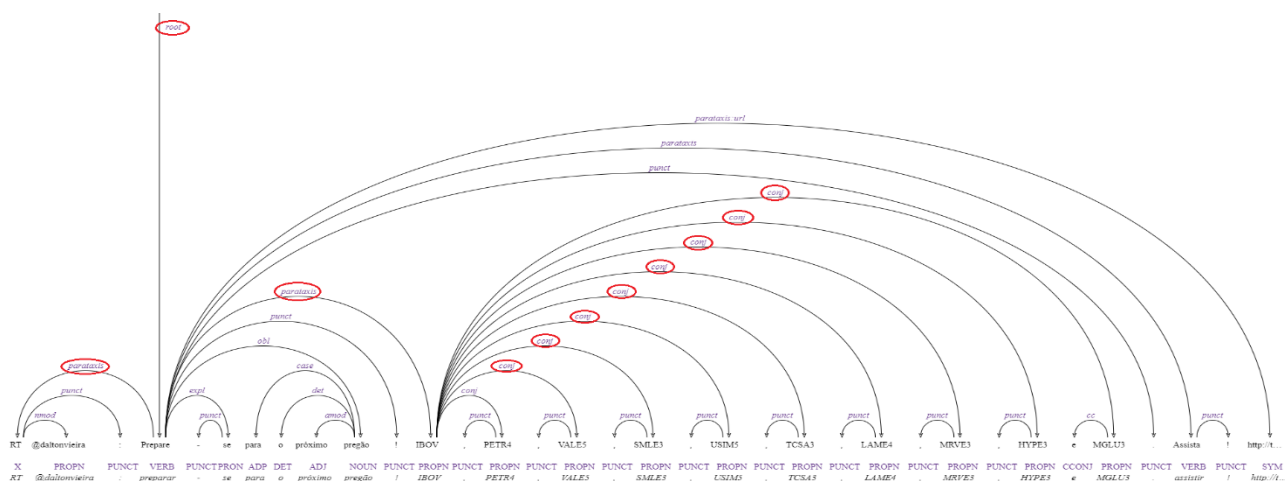


Figura 97 - Anotação do padrão do *Template* 12 (exemplo 96).

(97) RT @daltonvieira : Prepare - se para o próximo pregão ! IBOV , PETR4 , VALE5 , SMLE3 , USIM5 , TCSA3 , LAME4 , MRVE3 , HYPE3 e MGLU3 . Assista ! <http://t...>



Template 13

1. <prefixo> menciona empresa e ou ação
2. <sentença> é sintaticamente bem-formada e, por isso, contém o **root**
3. <prefixo> é dependente de **parataxis** de *root*, e seu **head** depende de sua forma
4. url é dependente por **parataxis:url** do *root*

(98) Sabesp (SBSP3) : Responsável por o abastecimento de 47 % de a região metropolitana de São Paulo , o Sistema Cantareira ... <http://t.co/xs8ji7To5z>

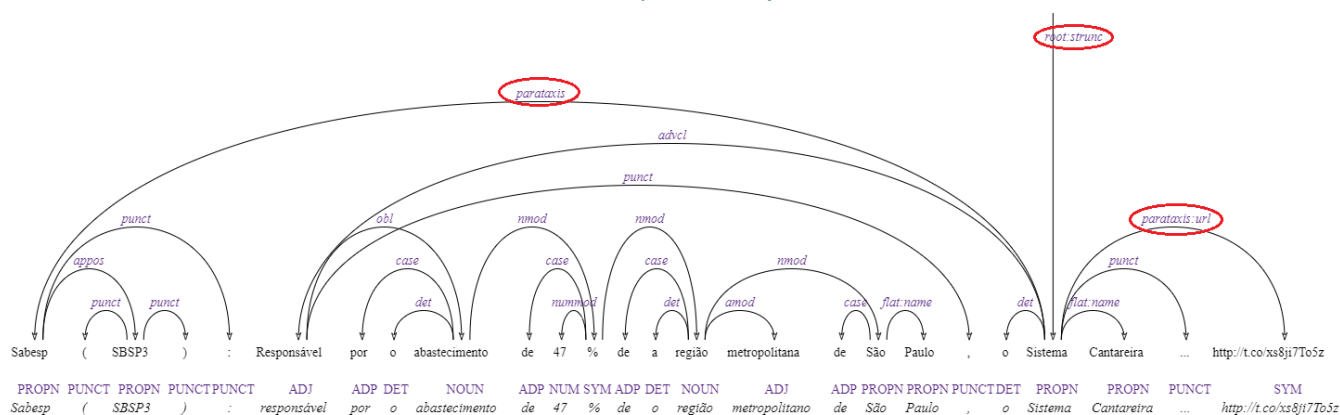


Figura 99 - Anotação do padrão do *Template* 13 (exemplo 98).

Exemplo:

(99) Notas gerais SABESP (SBSP3) : nível de o sistema cantareira caiu para 15 % , novo recorde de baixa segundo a ... <http://t.co/cQZeQ7P09u>

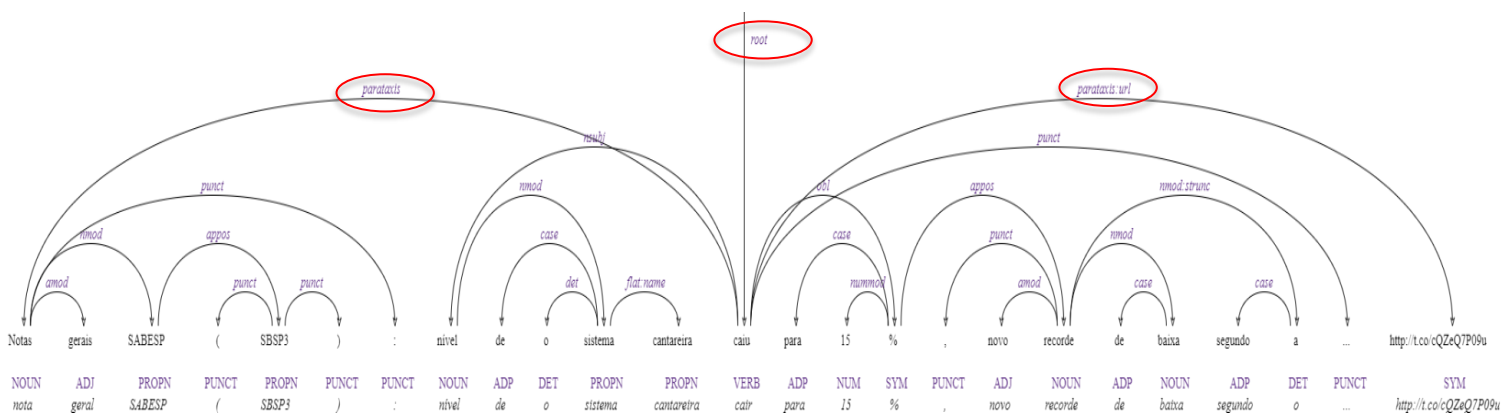


Figura 100 - Anotação do padrão do *Template 13* (exemplo 99).

Exemplo:

(100) #SBSP3 E cadê o racionamento !!! <http://t.co/ApULfZt5Sr>

(Interpretação: “E o racionamento está onde?”: onde é o predicativo, racionamento é o sujeito).

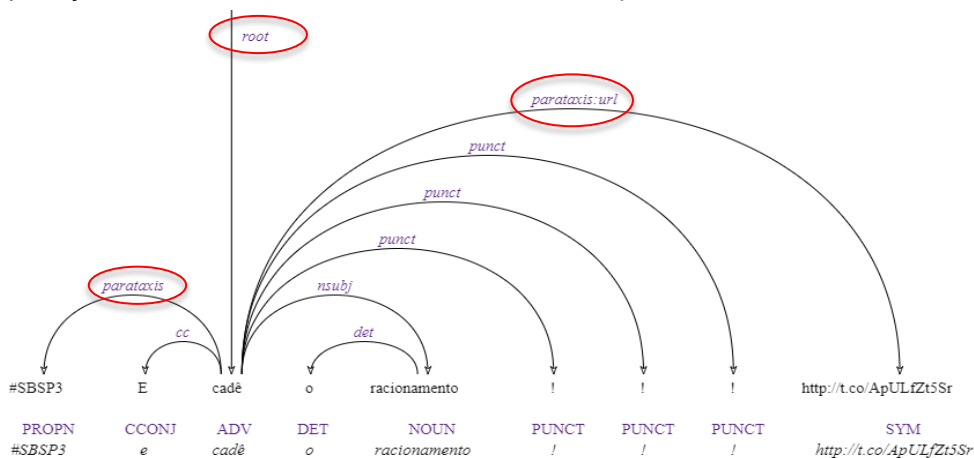


Figura 101 - Anotação do padrão do *Template 13* (exemplo 100).

Template 14

Padrão: <sentença> : <sentença truncada> url <lista *hashtags*>, em que:

1. <sentença> é sintaticamente completa e contém o **root**
2. <sentença truncada> é uma sentença truncada e seu **head** é dependente de <sentença> por **parataxis:strunc**
3. url é dependente por **parataxis:url** do **root**
4. <lista *hashtags*> são dependentes por **parataxis:hashtag** do **root**

Exemplo:

(101) Confira o potencial de valorização de 66 ações , upside supera 120 % : Foram utilizadas as prev ... <http://t.co/HbPlnX7Nxm> #infomoney #vale5

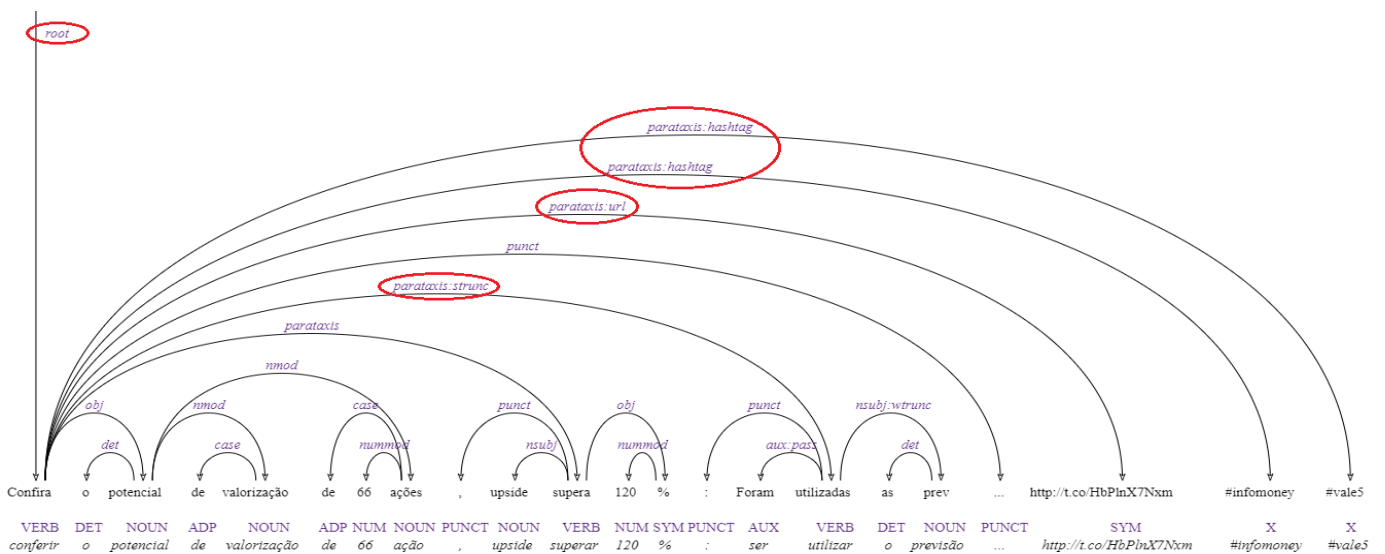


Figura 102 - Anotação do padrão do *Template 14* (exemplo 101).

Exemplo:

(102) ' Quem montar uma carteira de ações agora vai morrer de rir em 10 anos ', diz especialista : Se ... <http://t.co/CrU7flmRmc> #infomoney #vale5

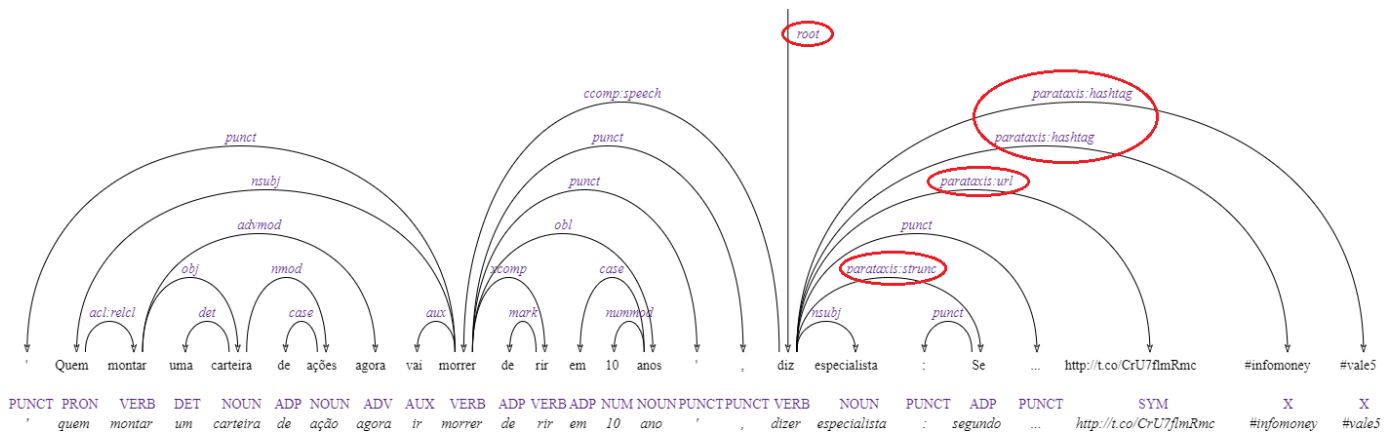


Figura 103 - Anotação do padrão do *Template 14* (exemplo 102).

Exemplo:

(103) Wintrade recomenda uma nova ação para março , saiba qual : A nova escolhida por os analistas é ... <http://t.co/fipOgc8hF4> #infomoney #vale5

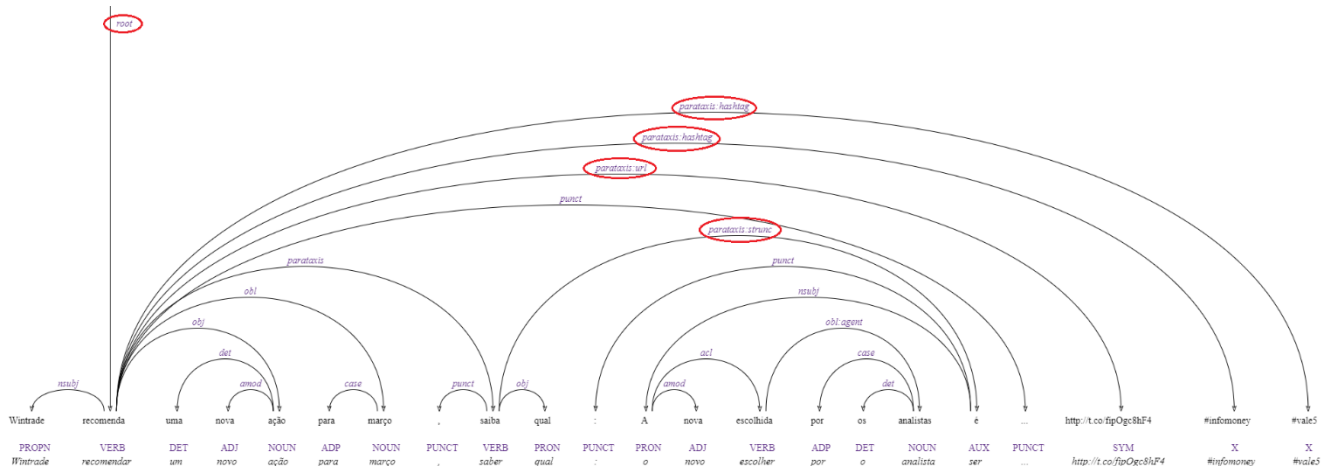


Figura 104 - Anotação do padrão do *Template 14* (exemplo 103).

Template 15

Padrão: **<cashtag> <empresa> (tipo de ação) - Fato Relevante - <fato> url**

1. “Fato Relevante”, se presente, é o tema, e <fato> é **root**; se ausente, **root** está em <fato>
2. <fato> é dependente de **root** pela relação **parataxis**
3. <cashtag> é dependente por **nmod** de **root**
4. <empresa> é dependente por **nmod** de <cashtag>
5. (tipo de ação) é dependente por **appos** de <cashtag>
6. url está ligada por **parataxis:url** ao root

Exemplo:

(104) \$MRFG3 - Marfrig (mrfg-nm) - Fato Relevante - Primeiro Aditivo A o Acordo De Acionistas <http://t.co/jBJzeY5U5V>

(Interpretação: Fato Relevante sobre ação \$MRFG3 da Marfrig no novo mercado (nm): Primeiro aditivo ao acordo de acionistas)

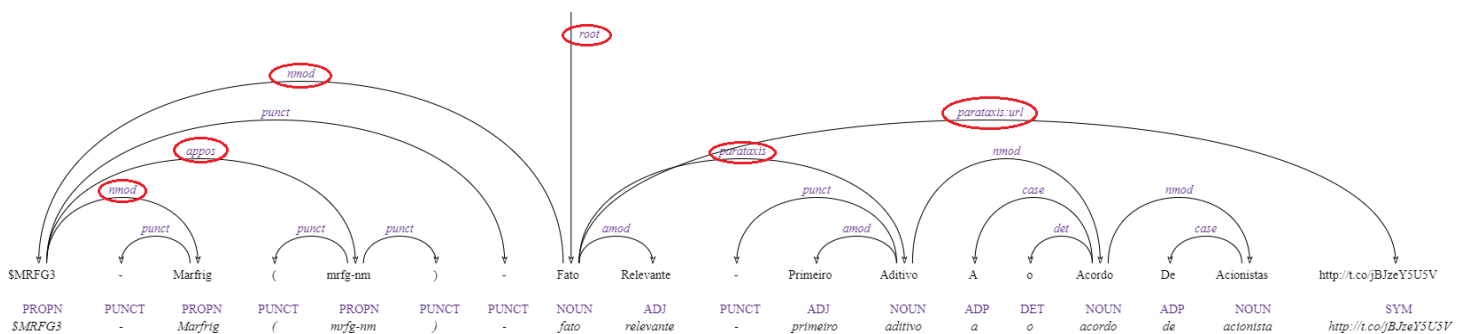


Figura 105 – Anotação do padrão do *Template 15* (exemplo 104).

Exemplo:

(105) \$RSID3 - Rossi Resid (rsid-nm) - Alteracao De a Politica De Divulgacao De Fato Relevante <http://t.co/gJVTbOY0uA>

(Interpretação: Alteração da política de divulgação de fato relevante sobre \$RSID3, da Rossi Resid, no novo mercado.)

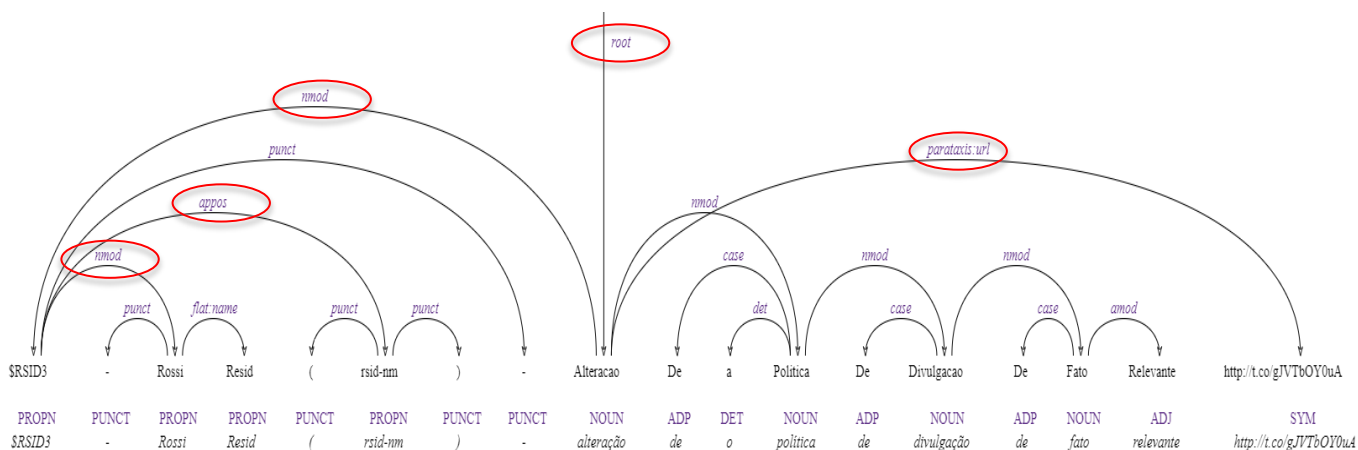


Figura 106 – Anotação do padrão do Template 15 (exemplo 105).

Exemplo:

(106) \$BBAS3 - Banco De o Brasil (bbas-nm) - Esclarecimentos A Consulta De a Bm&fbovespa <http://t.co/UaFXNfKifj>

(Interpretação: Esclarecimentos (sobre) a consulta da Bm&fbovespa (sobre) \$BBAS3 (do) Banco do Brasil (bbas-nm)

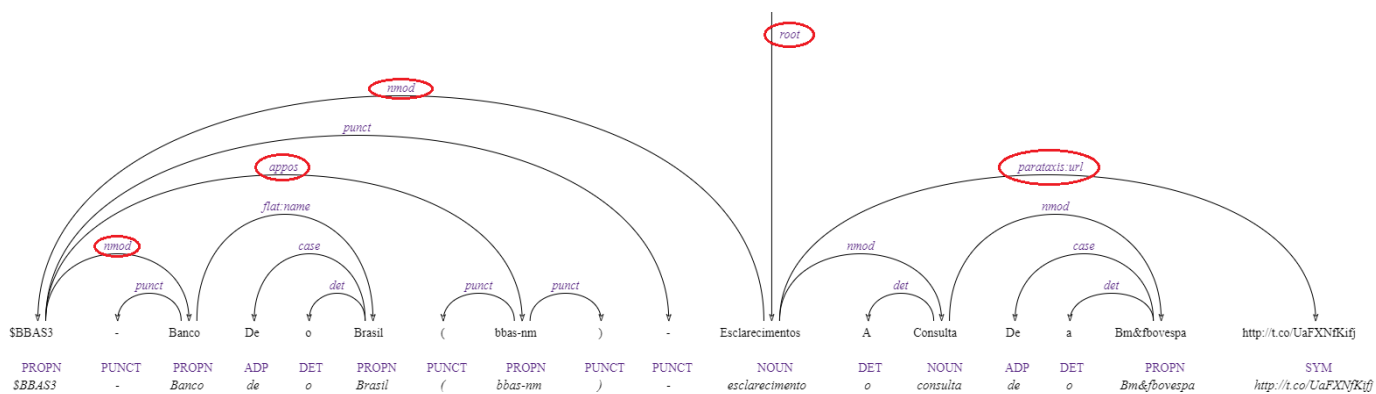


Figura 107 – Anotação do padrão do Template 15 (exemplo 106).

Template 16

Padrão: **[prefixo opcional] <sentença> url**, em que:

1. [prefixo] pode ter <lista *hashtags*>
2. <sentença> contém o *root*
3. cada elemento do [prefixo] é dependente por **parataxis** de *root*
4. url é dependente por **parataxis:url** de *root*

Exemplo :

(107) #BOVESPA #USIM5 S&P eleva perspectiva de nota de a Usiminas para estável :
http://t.co/JdfGy10phl #BR

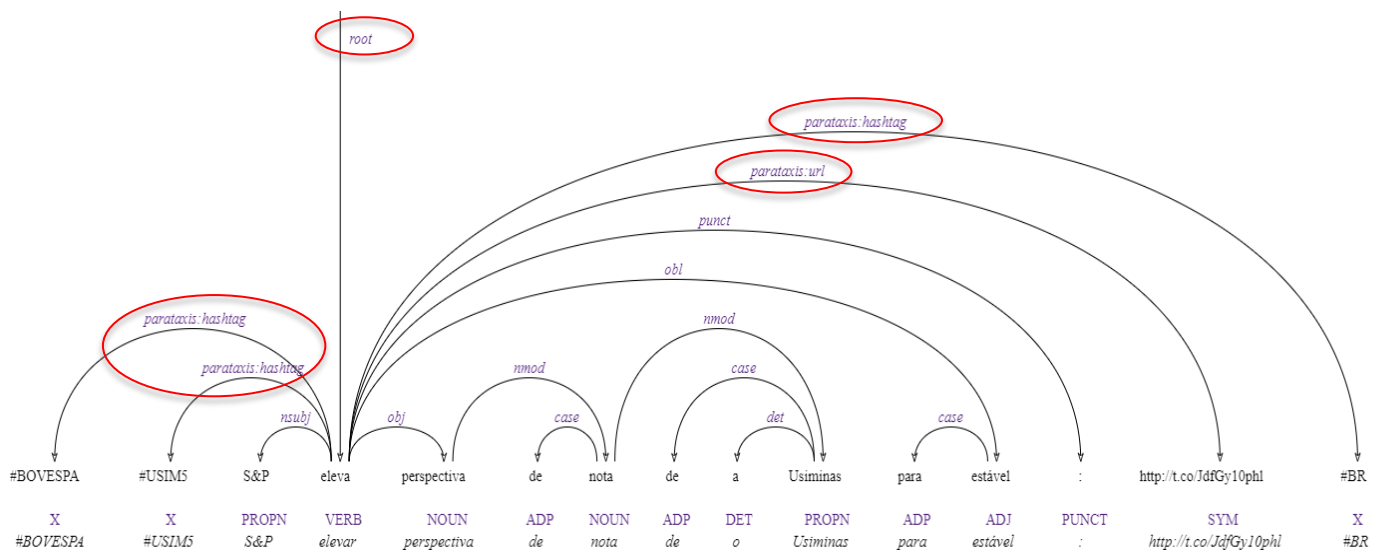


Figura 108 – Anotação do padrão com [prefixo] do *Template 16* (exemplo 107).

Template 17

Padrão: <lista *ticker*&[valor opcional]>, em que:

1. <lista *ticker*&[valor opcional]> é uma lista de um ou mais itens coordenados (**conj**), sendo:
|| <*ticker*> [valor-reais] || <*ticker*> [valor-reais] || <*ticker*> [valor-reais]....
2. [valor opcional]: cada *ticker* pode ou não ocorrer seguido do valor da ação que representa

Exemplo:

(108) Prepare - se para o próximo pregão ! Análise : **IBOV , PETR4 , VALE5 , TIMP3 , LPSB3 , MULT3 , GRND3 , LIGT3 e ABCB4** . Assista ! <http://t.co/OVfdQj4UPe>

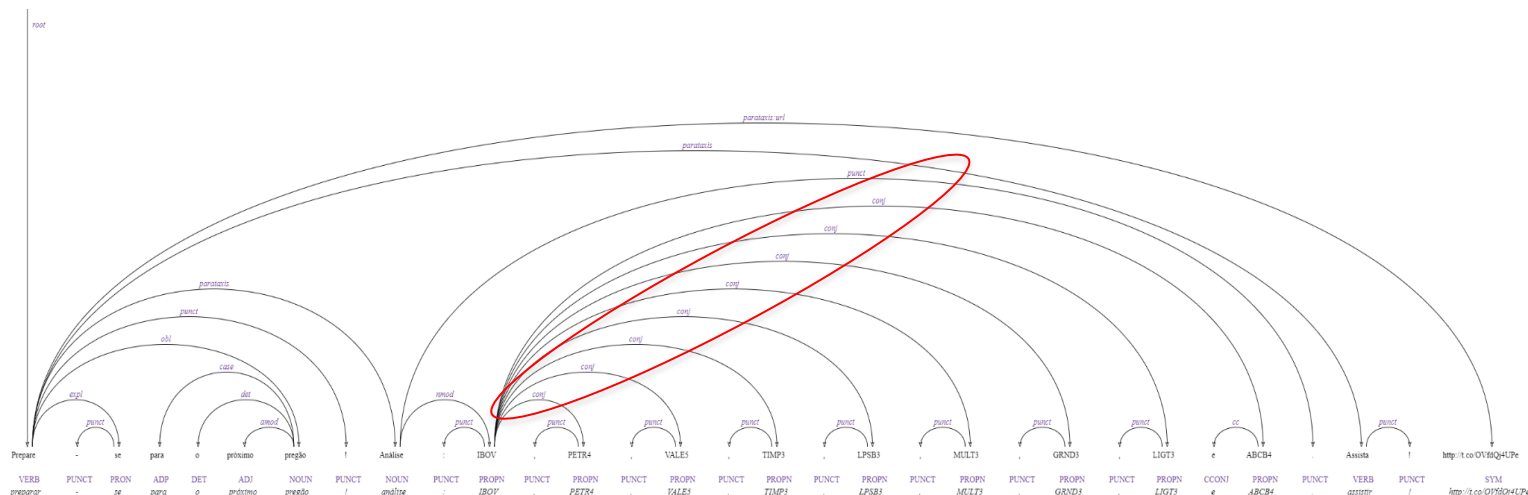


Figura 109 – Anotação do padrão do *Template 17* (sem os valores expressos).

Exemplo:

(109) 14/03/2014 - 17:19 : Maiores Baixas : **MRVE3 - 12,5 % R\$ 7,35 , DASA3 - 9,67 % R\$ 15,13 , CMIG4 - 5,69 % R\$ 12,94 , GFSA3 - 4,76 % R\$ 3 , ELPL4 - 4,03 % R\$ 7,62** .

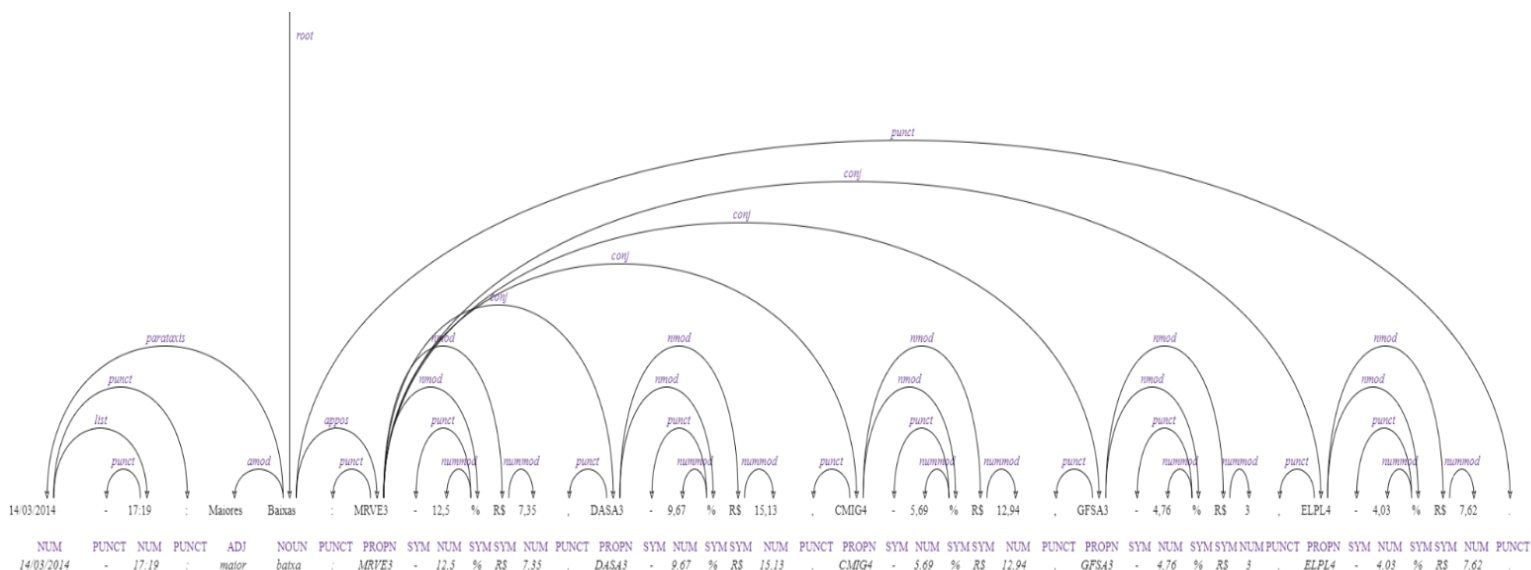


Figura 110 – Anotação do padrão do *Template 17* (com os valores expressos).

Template 18

Padrão: **Gatilho position , semana que vem : Ativo cmg4 - Venda - Validade 14/mar - Start 12,89 - Stop 13,75 - RP 12,03 <http://t.co/MPfzB6Vj0o>**

(110) Gatilho position , semana que vem : Ativo cmg4 - Venda - Validade 14/mar - Start 12,89 - Stop 13,75 - RP 12,03 <http://t.co/MPfzB6Vj0o>.



Figura 111 – Anotação do padrão do *Template 18*.

Template 19

Padrão: **INTRADAY PETR4 : Suportes 13,9 e 14,74 e resistências 15,99 e 16,4 INTRADAY VALE5 : Suportes 27,04 e 27,45 e resistências 28,12 e 28,38**

(111) INTRADAY PETR4 : Suportes 13,9 e 14,74 e resistências 15,99 e 16,4 INTRADAY VALE5 : Suportes 27,04 e 27,45 e resistências 28,12 e 28,38

(Interpretação: lista de informações de suportes e resistências “intraday” de ações)

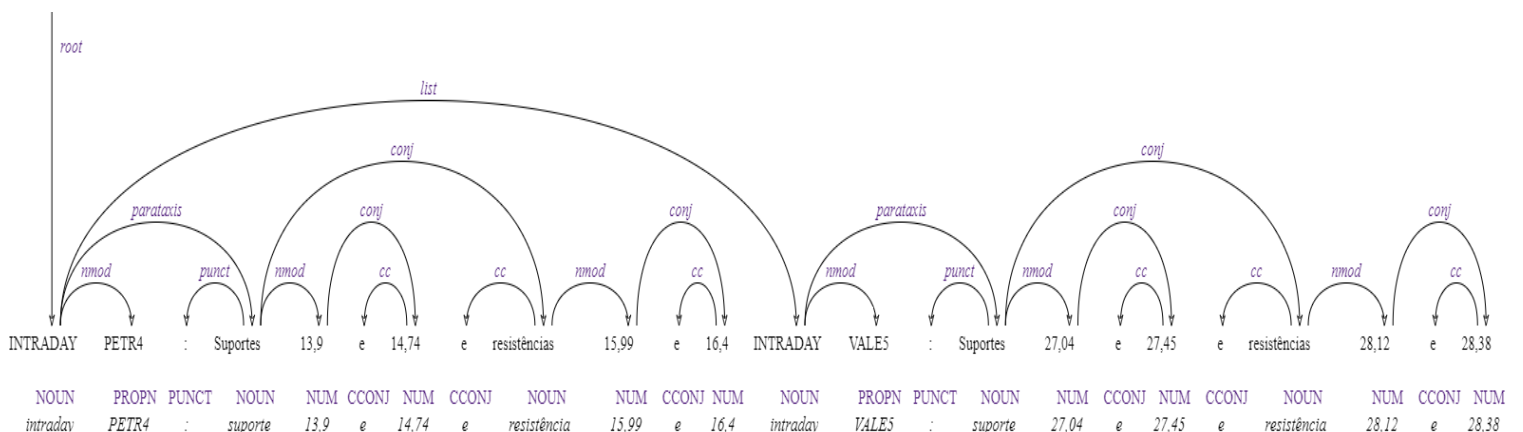


Figura 112 – Anotação do padrão do *Template 19*.

Template 20

Padrão: **Ações ex-dividendos hoje : GGBR3 , GGBR4 , GOAU3 , GOAU4 e MPLU3 . As cotações históricas foram ajustadas . Saiba mais ! <http://t.co/C7k4DuDID2>**

Exemplo:

(112) Ações ex-dividendos hoje : GGBR3 , GGBR4 , GOAU3 , GOAU4 e MPLU3 . As cotações históricas foram ajustadas . Saiba mais ! <http://t.co/C7k4DuDID2>

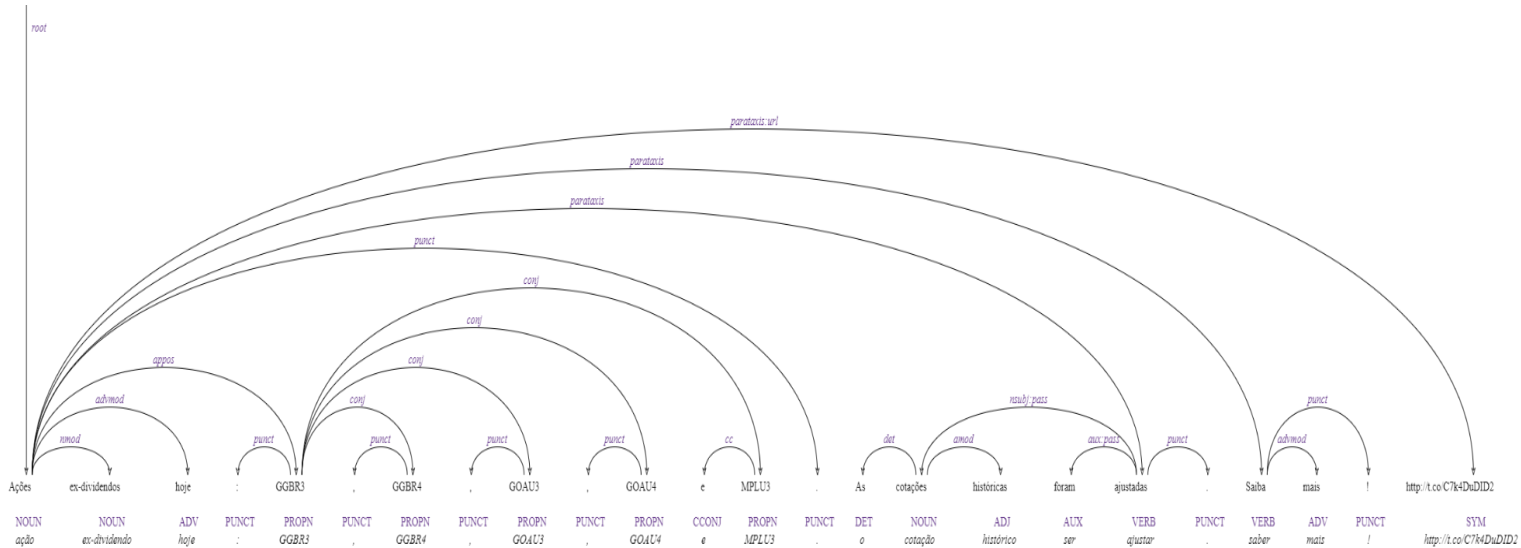


Figura 113 – Anotação do padrão do *Template 20*.

Template 21

Padrão: **Desempenho de as ações <empresa> em a semana passada : [bolsa] – <lista ticker&índice> , [bolsa – <lista ticker&índice>], em que:**

1. token “desempenho” é **root**
2. <empresa> indica o nome da empresa, que é **nmod** de “ações”
3. [bolsa] indica bolsa de valores e é opcional; se presente, liga-se por **parataxis** a **root**
4. <lista ticker&índice> são 2 ou mais elementos coordenados (**conj**) e ligados por **parataxis** a [bolsa] (quando presente) (Figura 45) ou ao **root** (Figura 46), sendo:
|| <ticker> <índice> || <ticker> <índice> ...
5. [bolsa - <lista ticker&índice>] é opcional e, quando presente, liga-se ao primeiro [bolsa] por **conj**

Exemplo:

(113) Desempenho de as ações de o Itaú Unibanco em a semana passada : em a Bovespa - ITUB4 : - 0,9 % , ITUB3 : 0,58 % , em a NYSE – ITUB : - 1,13 % , Índice Bovespa : - 1,8 %.

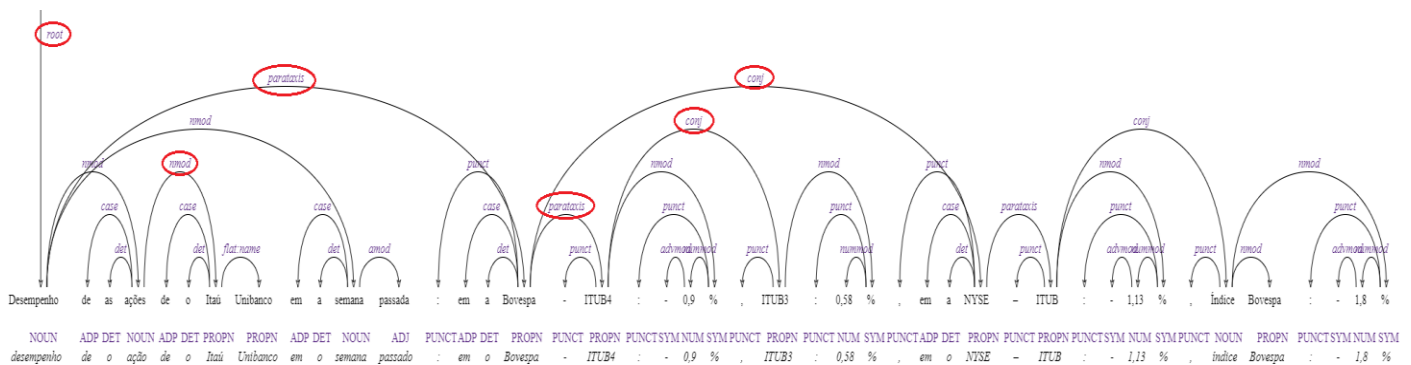


Figura 114 – Anotação do padrão (com os opcionais) do *Template 21* (exemplo 114).

Exemplo:

(114) Desempenho de as ações de a TIM em a semana passada : TIMP3 (Bovespa) : - 0.94 % , TSU (NYSE) : + 0.58 % , IBOV : + 2.64 % .

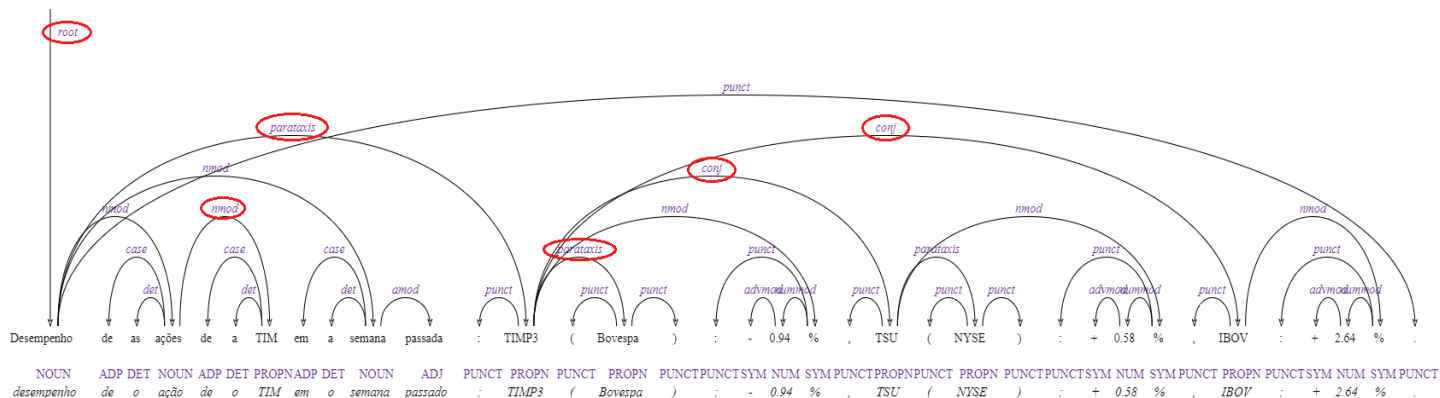


Figura 115 – Anotação do padrão (sem os opcionais) do *Template 21* (exemplo 115).

Template 22

Padrão: **MRVE3 vender a R\$ 7.99 indicado em 05/03/2014 13:29 e finalizou a compra com resultado de R\$ - 0.11 ou - 1.36 % <http://t.co/kgt1YiBf7>**, em que

1. MRVE3 é dependente por **obj** do **root**, já que é o objeto que complementa o verbo “vender” no infinitivo.

Exemplo:

(115) MRVE3 vender a R\$ 7.99 indicado em 05/03/2014 13:29 e finalizou a compra com resultado de R\$ - 0.11 ou - 1.36 % <http://t.co/kgt1YiTBf7>,

(Interpretação: "Vender (ação) MRVE3 a R\$7.99 (valor) indicado em 05/03/2014 (às) 13:29 e finalizou a compra com resultado de R\$-0.11 ou -1.36%")

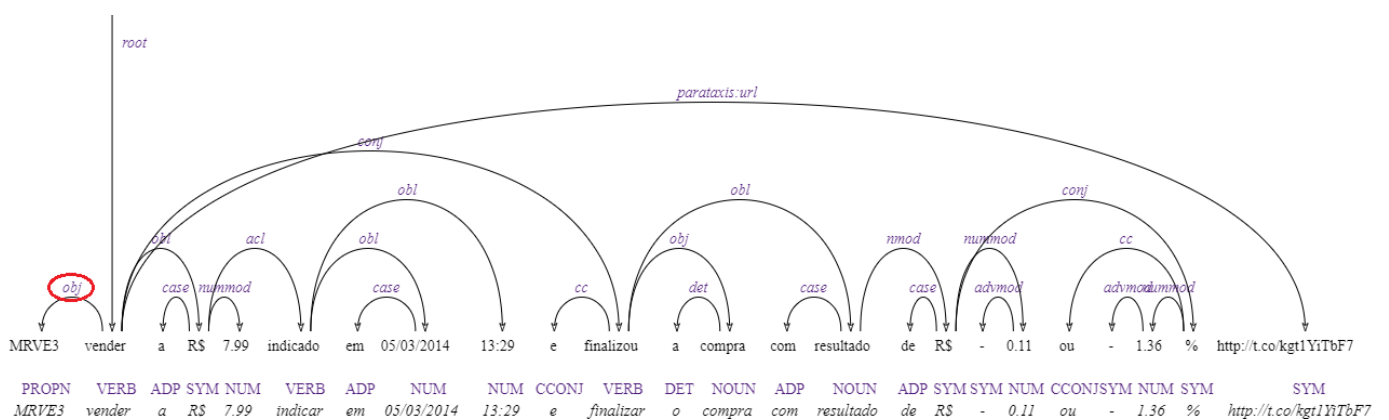


Figura 116 – Anotação do padrão do *Template 22*.

Bibliografia

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

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

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

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

Di Felippo, A.; Postali, C.; Cereghatto, G.; Gazana, L.S.; Silva, E.H.; Roman, N.T.; Pardo, T.A.S. (2021). Descrição Preliminar do Corpus DANTEStocks: Diretrizes de Segmentação para Anotação segundo Universal Dependencies. In the *Proceedings of the VII Workshop on Portuguese Description (JDP)*, pp. 335-343. December, 1.

Duran, M.S. (2022). Manual de Anotação de Relações de Dependência - Versão Revisada e Estendida: Orientações para anotação de relações de dependência sintática em Língua Portuguesa, seguindo as diretrizes da abordagem Universal Dependencies (UD). Relatório Técnico do ICMC 440. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo. São Carlos-SP, Outubro, 166p.

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

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

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

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

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

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

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