

POeTiSA

Corpus Folha-Kagle
Decisões no nível de POS tags

Magali Sanches Duran

10/06/2021

Decisões de Anotação - NUMERAIS

A POS tag NUM passa a ser entendida como uma super classe que abriga todos os numerais cardinais, que podem funcionar como quantificador, como pronome (quando a coisa quantificada está elíptica) e até como nome próprio. Isso está perfeitamente de acordo com as Guidelines da UD.

Tenho

| | |
|-------|------------------------------------------|
| (NUM) | cinco sobrinhos: |
| (NUM) | três deles fazem engenharia e |
| (NUM) | dois não estudam. Os |
| (NUM) | cinco têm menos de |
| (NUM) | vinte anos, mas todos têm mais de |
| (NUM) | quinze . |

Revisão de Decisões de Anotação - NUMERAIS

½ - meio

1 – um

1000 – mil

MAS

Um **terço** (NUM, **NOUN**)

Dois **quartos** (NUM, **NOUN**)

Três **oitavos** (NUM, **NOUN**)

Quatro doze **avos** (NUM, NUM, **NOUN**)

Revisão de Decisões de Anotação - NUMERAIS

Alguns fatos sobre o comportamento dos números em português:

- dois carrinhos (NUM, NOUN)
- duas bolinhas (NUM, NOUN)

Apenas **10 números flexionam em gênero**, concordando com o gênero da coisa quantificada:

- um, dois
- duzentos, trezentos, quatrocentos, quinhentos, seiscentos, setecentos, oitocentos, novecentos

Revisão de Decisões de Anotação - NUMERAIS

Alguns fatos sobre o comportamento dos números em português:

- dois mil carrinhos (NUM, NUM, NOUN)
- duas mil bolinhas (NUM, NUM, NOUN)

Números não flexionam em número

Porque estamos anotando MILHÃO como NOUN

Quando chegamos à casa do milhão, as coisas mudam sintaticamente

fem

fem

- duas mil bolinhas (NUM, NUM, NOUN)

masc

fem

- dois milhões **de** bolinhas (NUM, NOUN, **ADP**, NOUN)

O numeral que flexiona em gênero (dois) não está concordando com a coisa (semanticamente) quantificada (bolinhas), mas sim com a palavra “milhões”, que é do gênero masculino.

A palavra “milhão”, que semanticamente entendemos como um número, flexiona em número, apresentando uma forma plural, “milhões”, quebrando o comportamento dos números, que não flexionam em número.

Porque estamos anotando MILHÃO como NOUN

Quando observamos o inglês, vemos que o comportamento sintático da palavra “million” motiva sua anotação:

Two **million** balls.

como os demais numerais, million não varia em número

MILLION = **NUM**

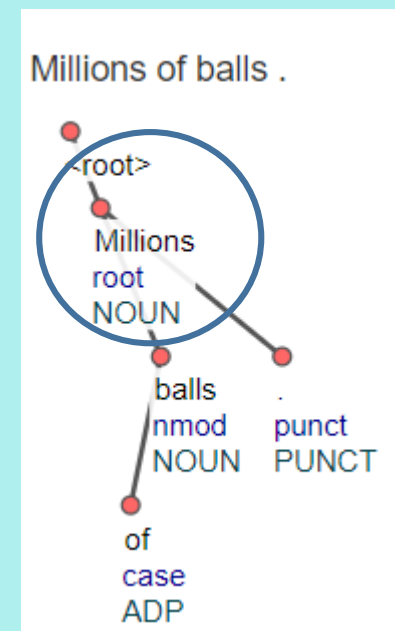
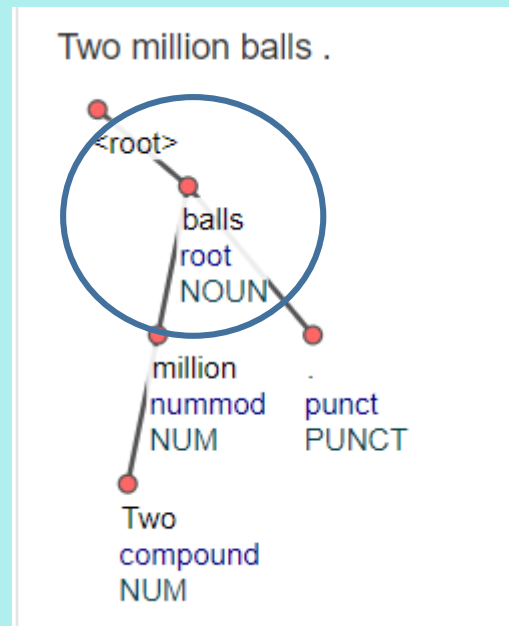
Millions of balls.

million varia em número e separa-se da coisa (semanticamente) quantificada pela preposição “of”.

MILLION = **NOUN**

Porque estamos anotando MILHÃO como NOUN

O “of” impede que, sintaticamente, milhão seja um quantificador no segundo exemplo e inverte-se a relação head-dependent:



Porque estamos anotando MILHÃO como NOUN

CONCLUSÃO: milhão, milhões, bilhão, trilhão, etc. são NOUN

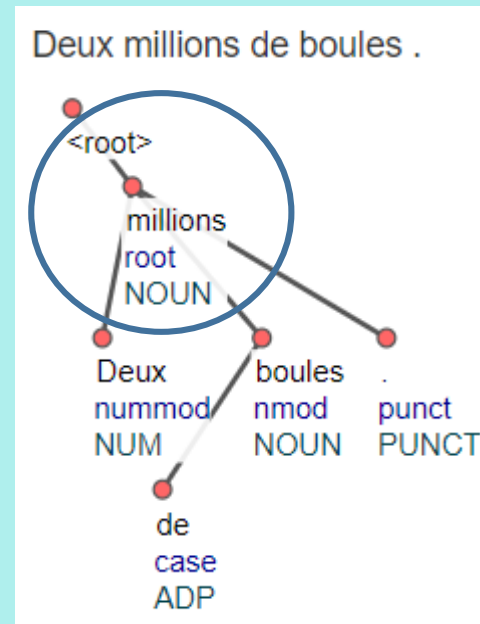
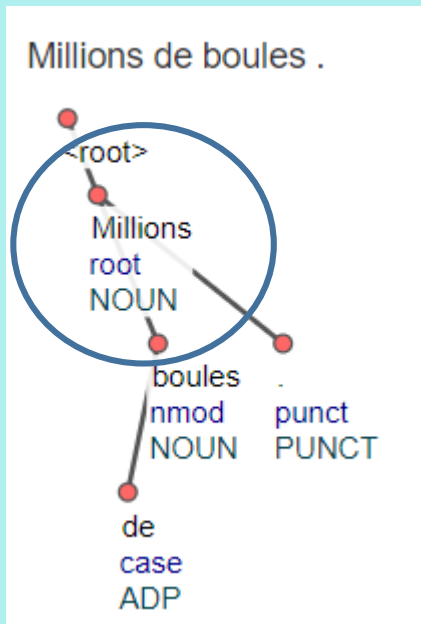
Porque estamos anotando MILHÃO como NOUN

No português, “milhão” (e “bilhão”, “trilhão”, etc.) sempre se comporta como NOUN sintaticamente:

- Sempre varia em número, concordando com a quantidade expressa pelo NUM que o precede
- Sempre exige a preposição “de” para se ligar ao NOUN que (semanticamente) quantifica

Porque estamos anotando MILHÃO como NOUN

Se o comportamento da palavra “milhão” não varia em português, por que deveria variar sua anotação? Vejamos o caso do francês:



Porque estamos anotando MILHÃO como NOUN



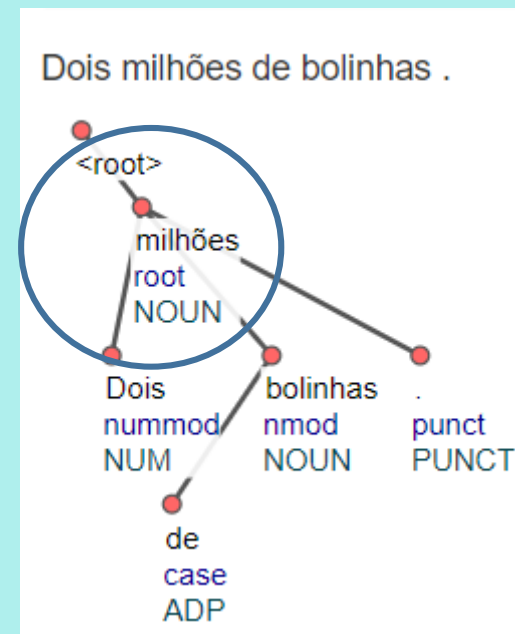
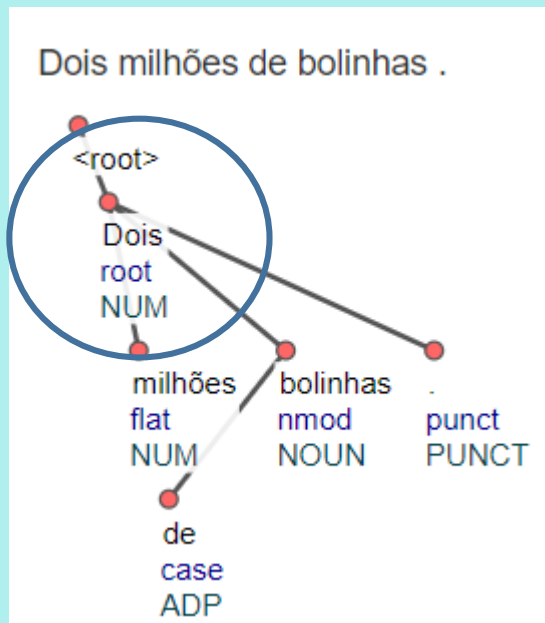
Tinha um “de” no meio do caminho.

No meio do caminho tinha um “de”.

Porque estamos anotando MILHÃO como NOUN

Tudo por causa de um “de”...

Seja “milhão” anotado como NUM ou NOUN, a coisa (semanticamente) quantificada será DEPENDENT da deprel nmod e não HEAD da deprel nummod



Porque estamos anotando MILHÃO como NOUN

FINALMENTE,

o maior argumento para anotar “milhão” como NOUN é que a UD nos orienta a privilegiar o sintático e não o semântico quando há conflito entre os dois níveis.

o padrão NUM, NOUN, ADP, NOUN

NUM = quantidade da unidade de medida

NOUN = unidade de medida

ADP = liga a unidade de medida à coisa (semanticamente) quantificada

NOUN = coisa quantificada

Dois milhões de vidas .

NUM

NOUN

ADP

NOUN

o padrão NUM, NOUN, ADP, NOUN

- Duas **dúzias** de banana
- Dez **metros** de tecido
- Meio **saco** de farinha
- Duas **gotas** de baunilha
- Cinco **fardos** de feno .
- Mil **quilômetros** de distância

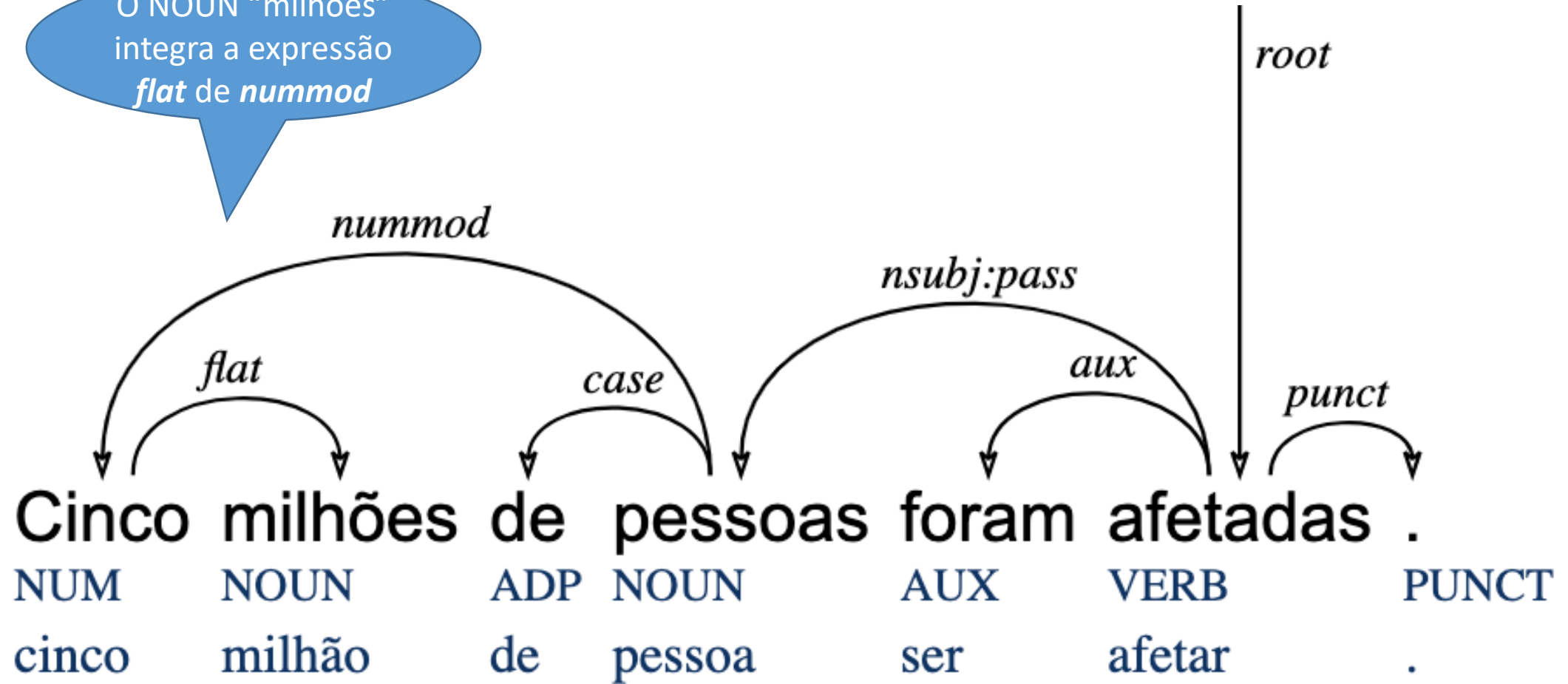
Milhões é o único **NOUN** que aceita combinação com outro **NOUN** unidade de medida:

- Cinco **milhões** de **toneladas** de grãos

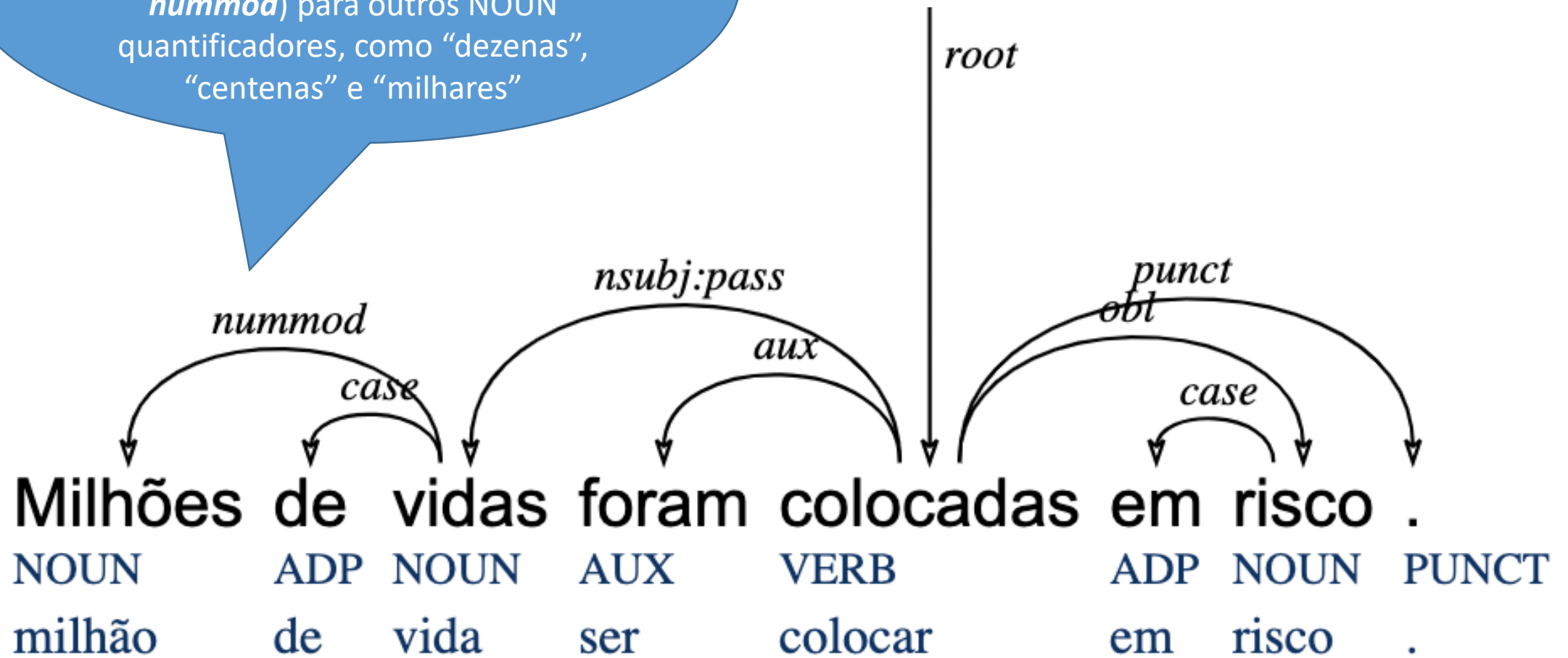
PROPOSTA DO GRUPO APÓS DISCUSSÃO

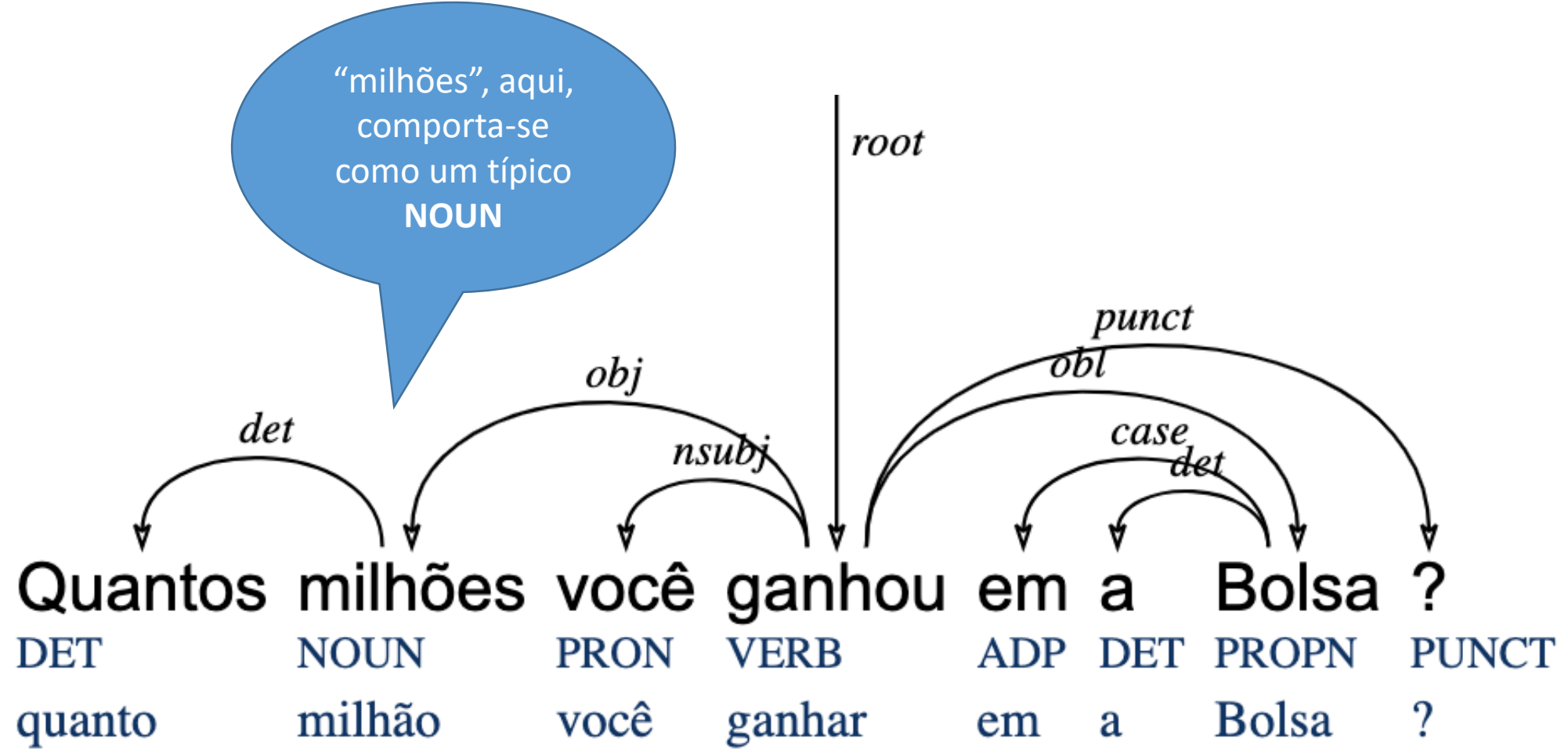
- Anotar “milhão” e “milhões” como NOUN. (Cláudia disse que, se integrar expressão flat, é indiferente para ela se é NUM ou NOUN)
- Nos próximos slides apresento as árvores anotadas pela Adriana com quatro exemplos de uso da palavra “milhões”.
- O Thiago testou a árvore que apresenta NOUN como dependente da relação ***nummod*** e ela passou no programa de validação da UD.

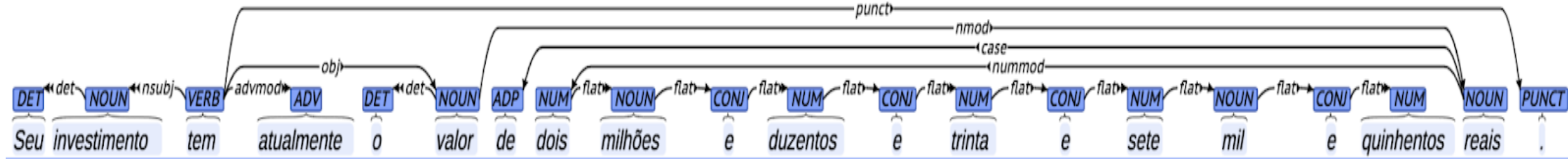
O NOUN “milhões”
integra a expressão
flat de *nummod*



“milhões” aqui se comporta como NOUN quantificador. Teríamos que estender esse tratamento (como *nummod*) para outros NOUN quantificadores, como “dezenas”, “centenas” e “milhares”







O NOUN “milhões”
integra a expressão
flat de *nummod*

