

Anotação de POS tags segundo o contexto

Magali Sanches Duran

20/05/2021

Diretrizes da UD: não leva em conta o contexto

- “... **adjectives** that exceptionally head a nominal phrase (as in the sick, the healthy) are still tagged ADJ.”
- “A **numeral** is a word, functioning most typically as a determiner, adjective or pronoun, that expresses a number and a relation to the number, such as quantity, sequence, frequency or fraction.”
- We understand the **DET class** as pro-adjectives, which is a slightly broader sense than what is usually regarded as determiners in English. In particular, it is possible that one nominal is modified by more than one determiner. They (DET) are more likely to be used attributively (modifying a noun phrase) than substantively (replacing a noun phrase). They may occur alone, though. If they do, it is either because of ellipsis, or because the hypothetical modified noun is something unspecified and general, as in All [visitors] must pay.

Diretrizes do **Bosque**: anotar segundo o contexto

Diretrizes do **Poetisa**: anotar segundo o contexto

Atribuir à palavra a POS tag que traduza sua função no contexto:

- Promover DET a PRON sempre que estiver funcionando como um nominal
- Promover ADJ a NOUN sempre que estiver funcionando como um nominal
- Promover NUM a PRON sempre que estiver funcionando como um nominal
- Promover VERB a NOUN sempre que estiver funcionando como um nominal

Motivações: ao contrário de outras línguas, o português tem muitas elipses e usa muitos adjetivos e pronomes indefinidos como nominal.

Anotação das POS tags em contexto

Cada palavra é anotada com a POS tag da função que exerce na oração:

Vamos observar o **andar** de a carruagem.

AUX, VERB, DET, **NOUN**, ADP, DET, NOUN

O **bonito** de a história é que ninguém se feriu.

DET, **NOUN**, ADP, DET, NOUN, VERB, SCONJ, PRON, VERB, PUNCT

Cada **um** sabe onde aperta o sapato.

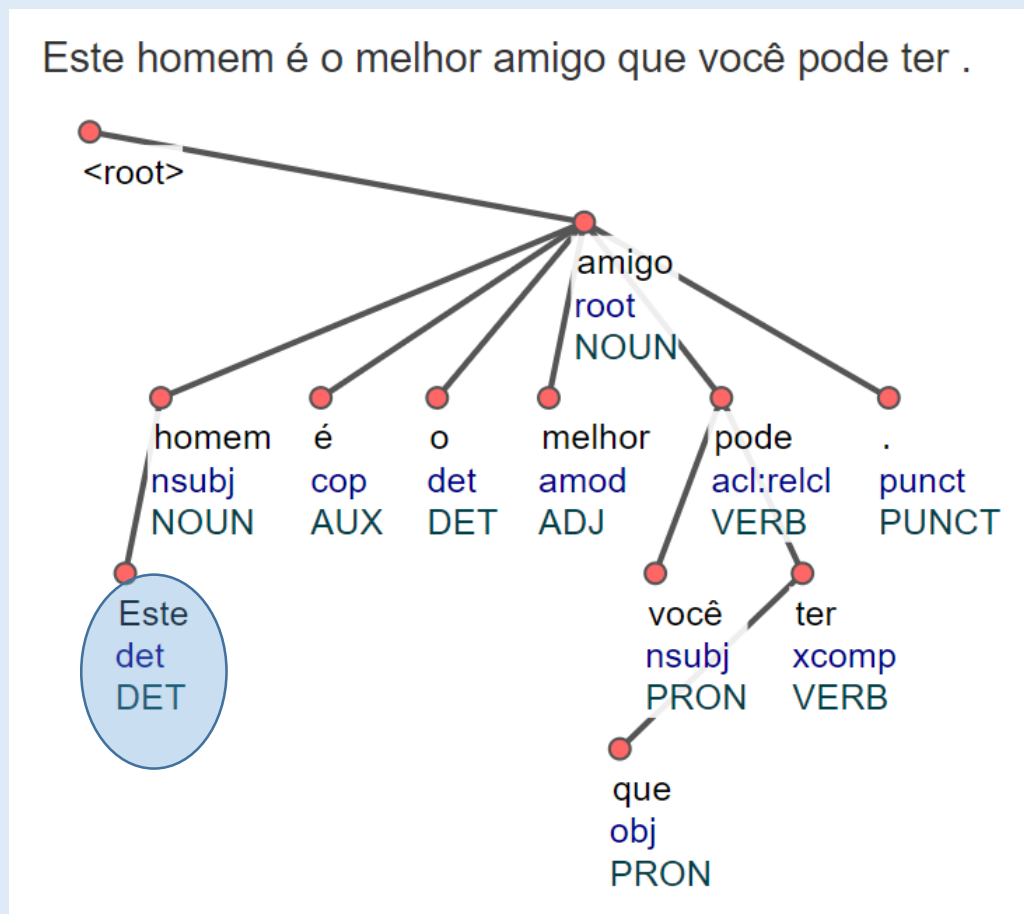
DET, **PRON**, VERB, ADV, VERB, DET, NOUN, PUNCT.

Eu me referia a **a** de a direita.

PRON, PRON, VERB, ADP, **PRON**, ADP, DET, NOUN, PUNCT

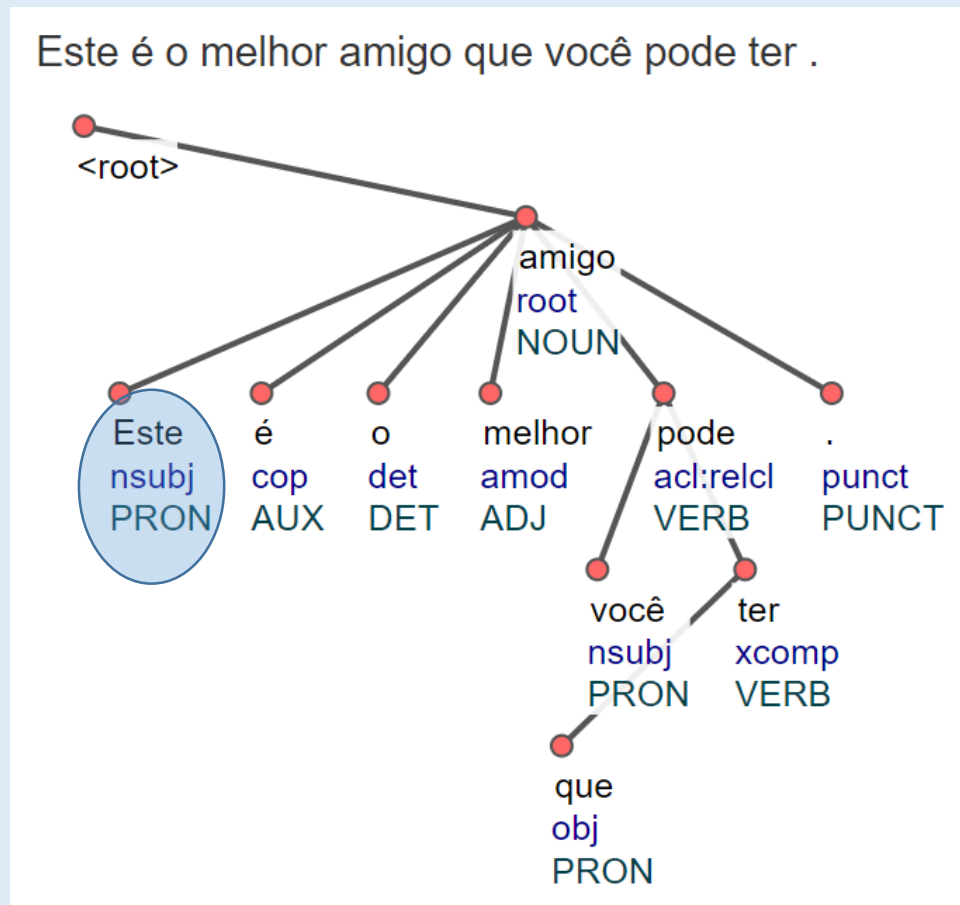
Anotação da POS tags seguindo contexto

Pronomes não substantivos são anotados como **DET** sempre que especificarem um nominal



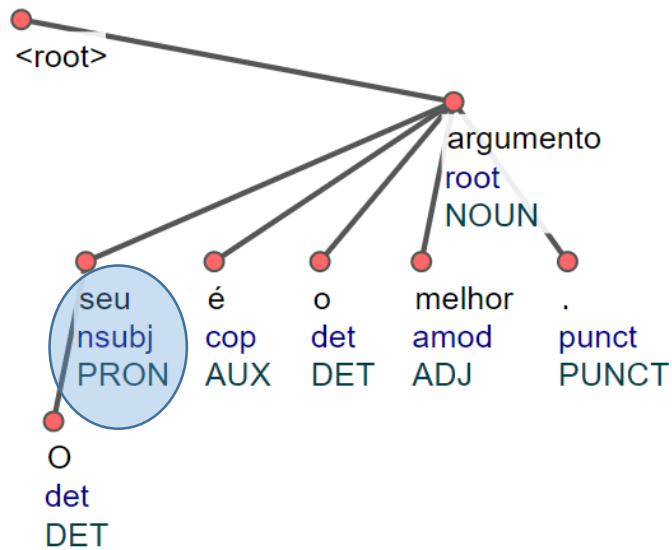
Anotação da POS tag seguindo contexto

Se um pronome **DET** estiver atuando como nominal (não como modificador de nominal), será promovido a **PRON**.

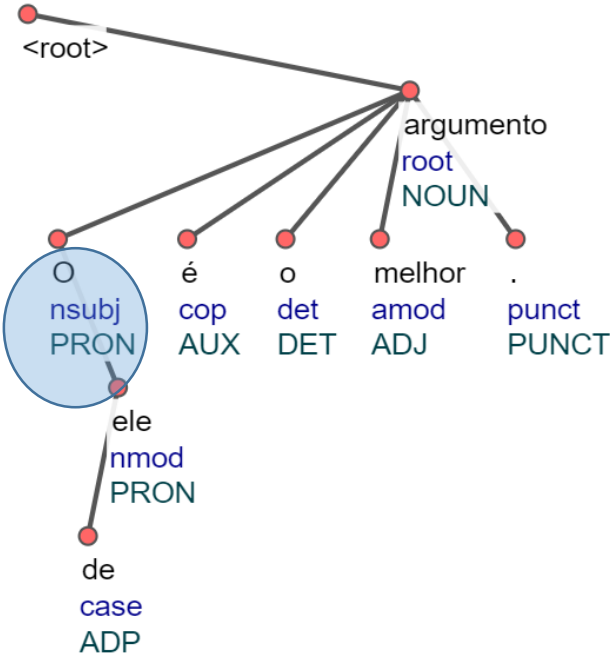


Qual **DET** promover a **PRON** quando há dois **DET** “órfãos” de nominal?

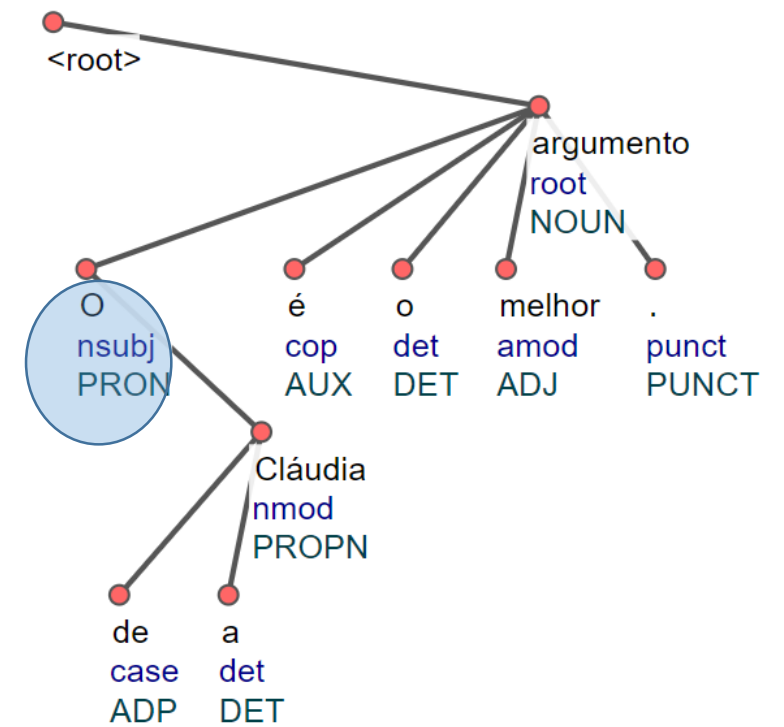
O seu é o melhor argumento .



O de ele é o melhor argumento .

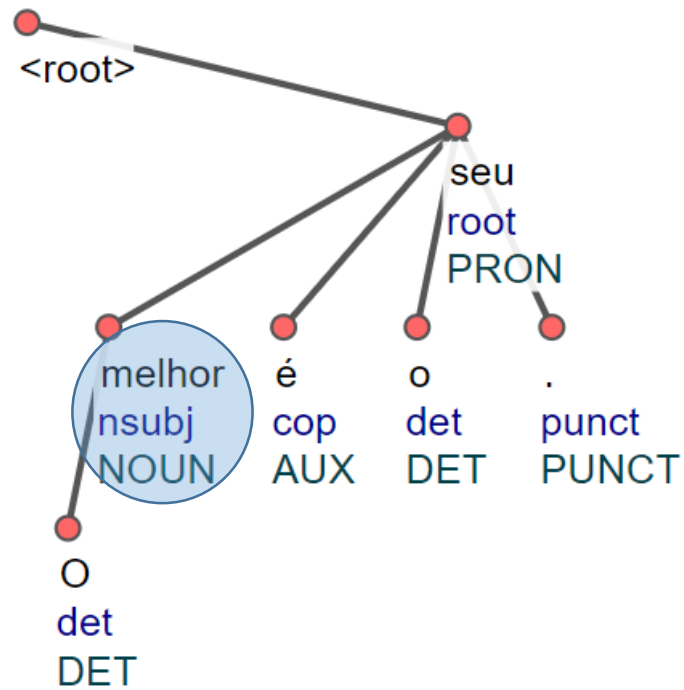


O de a Cláudia é o melhor argumento .

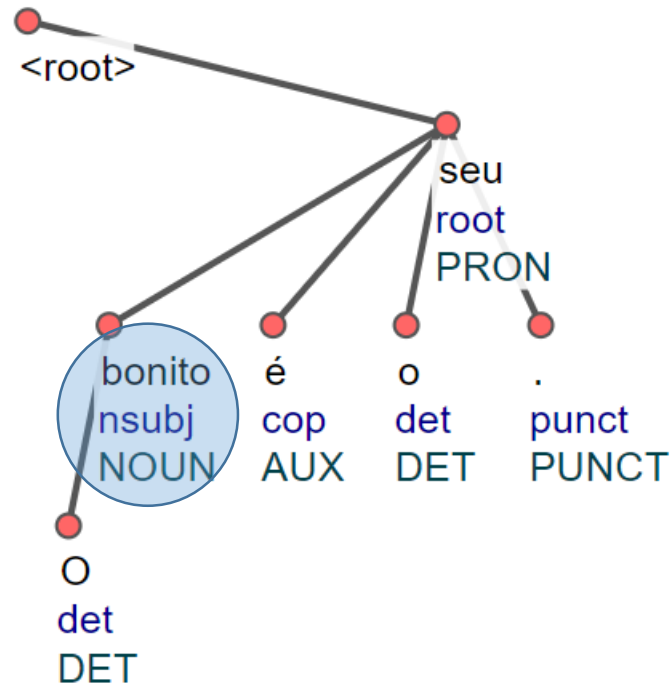


Se um **DET** e um **ADJ** estão “órfãos”, o **ADJ** é promovido a **NOUN**

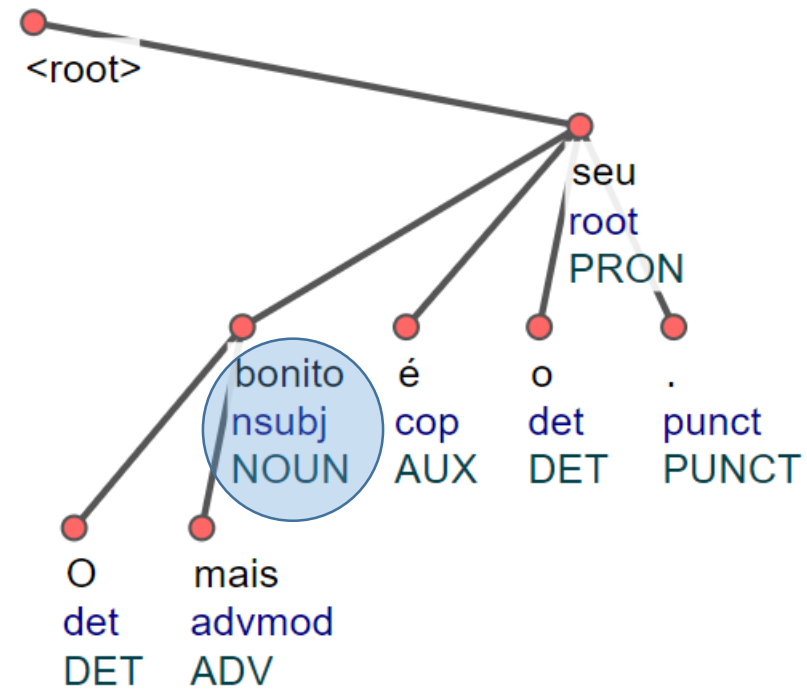
O melhor é o seu .



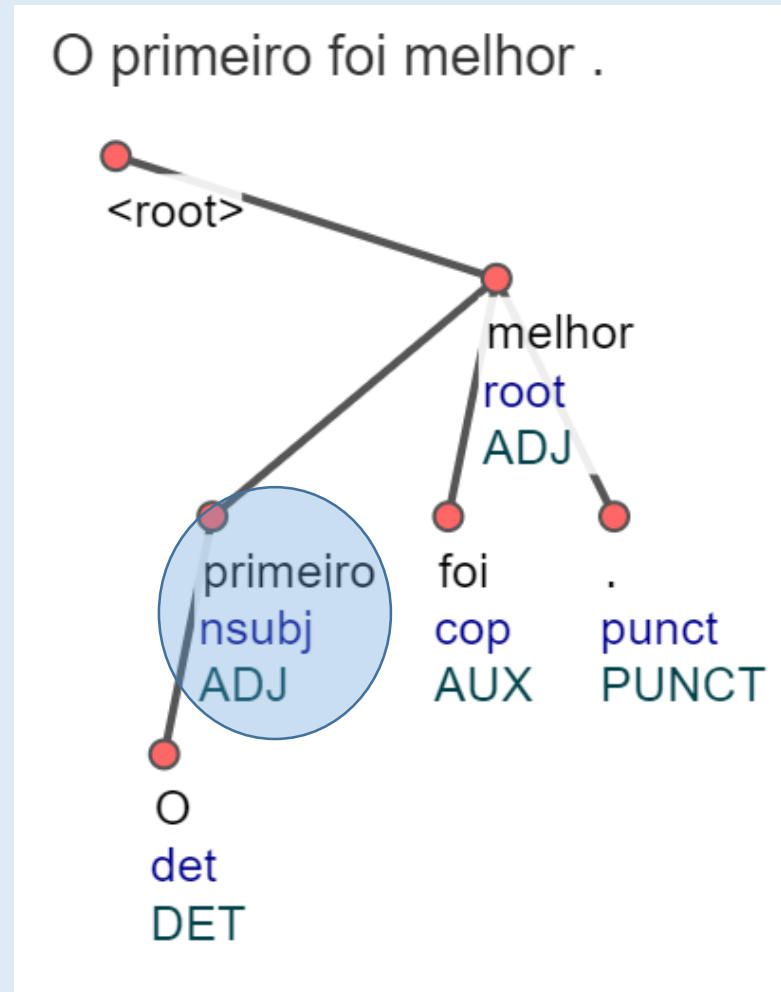
O bonito é o seu .



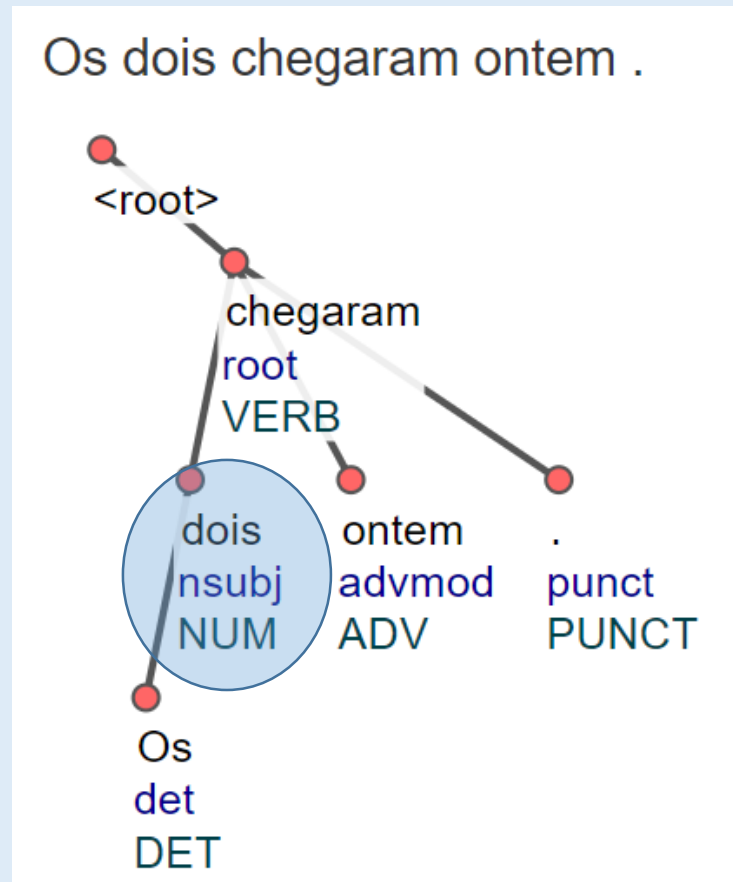
O mais bonito é o seu .



Se o **ADJ** “órfão” for um numeral ordinal, não é promovido a **NOUN**



No Bosque, os numerais cardinais são sempre anotados como **NUM**, independentemente de contexto



Que etiqueta atribuir para cada um desses numerais cardinais se o critério for contexto?

Tenho **cinco** sobrinhos: **três** deles fazem engenharia e **dois** não estudam. Os **cinco** têm menos de **20** anos, mas todos têm mais de **15**.

Sempre que **NUM** não estiver quantificando um **NOUN** deve ser promovido a **PRON**?

Tenho

(NUM) **cinco** sobrinhos:

(PRON) **três** deles fazem engenharia e

(PRON) **dois** não estudam. Os

(PRON) **cinco** têm menos de

(NUM) **vinte** anos, mas todos têm mais de

(PRON) **quinze**.