



PROPN e afins

Atualmente na UD

- **Nome próprio** é um substantivo que é “name” de um indivíduo, lugar ou objeto específico (→ SEMÂNTICO)
- **PROPN** é usada para os substantivos que funcionam como *names* e que geralmente exibem propriedades sintáticas especiais
 - * não modificação por um artigo singular em inglês (→ SINTÁTICO)
 - *I went at **Peter's**.* (PROPN)
 - *Did you see what **Peter** did?* (PROPN)
 - *I went at the **university**.* (NOUN)
 - *I went to the **shop**.* (NOUN)

Atualmente na UD

**Proper
nouns**
(single-words)

- **Nome próprio** é um substantivo que é “name” de um indivíduo, lugar ou objeto específico (→ SEMÂNTICO)
 - **PROPN** é usada para os substantivos que funcionam como *names* e que geralmente exibem propriedades sintáticas especiais
 - * não modificação por um artigo singular em inglês (→ SINTÁTICO)
- | | |
|--|---------|
| ◦ <i>I went at Peter's.</i> | (PROPN) |
| ◦ <i>Did you see what Peter did?</i> | (PROPN) |
| ◦ <i>I went at <u>the</u> university.</i> | (NOUN) |
| ◦ <i>I went to <u>the</u> shop.</i> | (NOUN) |

Atualmente na UD

- **PROPN** é usada para **expressões** que funcionam como “name”, sendo seus elementos constitutivos anotados com suas *tags* originais
 - *Cat on a Hot Tin Roof* (**NOUN, ADP, DET, ADJ, NOUN, NOUN**)
- Ressalva:
 - Como herança da anotação do Penn TreeBank, palavras “originalmente” ADJ e particípio são anotadas como PROPN
 - *Yellow Pages, United States* (**PROPN, PROPN**)

Atualmente na UD

*Proper
names*

- **PROPN** é usada para **expressões** que funcionam como “name”, sendo seus elementos constitutivos anotados com suas *tags* originais
 - *Cat on a Hot Tin Roof* (**NOUN, ADP, DET, ADJ, NOUN, NOUN**)
- Ressalva:
 - Como herança da anotação do Penn TreeBank, palavras “originalmente” ADJ e particípio são anotadas como PROPN
 - *Yellow Pages, United States* (**PROPN, PROPN**)

Atualmente na UD

- **PROPN** é usada para acrônimos de nome próprios
 - **NATO** (**PROPN**) (*“North Atlantic Treaty Organization”*)
 - **Covid-19** (**PROPN**)
 - *Windows 7* (**PROPN**, **NUM**)
 - **John Paul II** (**PROPN**, **PROPN**, **NUM**)

Discussão nos *Issues*

- Questionamento sobre “**properness**” com base no italiano
 - *I went at **Peter’s** / Sono andato da **Pietro**.* (PROPN)
 - *Did you see what **Peter** did? / Hai visto cos'ha fatto il **Pietro**?* (PROPN)
 - *I went at the **university** / Sono andato in **università**.* (NOUN)
 - *I went to the **shop** / Sono andato al **negozio**.* (NOUN)
- Em português
 - Eu fui a o **McDonald’s**.
 - Eu fui a a **universidade**.
 - Eu vi (a) **Magali**.

Discussão nos *Issues*

- Questionamento com base no italiano
 - *I went at **Peter's** / Sono andato da **Pietro**.* (PROPN)
 - *Did you see what **Peter** did? / Hai visto cos'ha fatto il **Pietro**?* (PROPN)
 - *I went at the **university** / Sono andato in **università**.* (NOUN)
 - *I went to the **shop** / Sono andato al **negozio**.* (NOUN)
- Em português
 - Eu fui a o **McDonald's**.
 - Eu fui a a **universidade**.
 - Eu vi (a) **Magali**.

*Properness diz
respeito aos
nomes em geral*

Pelos *corpora* em inglês

EWT *corpus* (blogs, reviews, e-mails) → PoS praticamente manual)

- Parece haver bem menos inconsistências quanto aos ***proper nouns***
- Inconsistência maior quanto aos ***proper names***

[the] **District of Columbia**

(NOUN, ADP, **PROPN**)

[the] **District of Columbia**

(**PROPN**, ADP, **PROPN**)

[The] **United States** [goes into ...]

(ADJ, **PROPN**)

[Can I text from Canada to the] **united states?**

(ADJ, **PROPN**)

Proposta 1 nos *Issues*

- Não usar a PoS **PROPN** (mas NOUN...)
- Mover a informação para o nível morfológico (lexical)
 - Criar uma *feature* (como *abbr*, *foriegn*) para PROPN
- * Similaridade com a estratégia do *Google* (Amanda)

Adriana mora em Belo Horizonte.

TEXT	LEMMA	POS		TAG	DEP	HEAD
Adriana	adriana	NOUN		NNP__feminine__singular__proper	NSUBJ	mora
mora	morar	VERB	VBC__imperfective__indicative__singular__third__not_proper__present		ROOT	mora
em	em	ADP		IN__not_proper	PREP	mora
Belo	belo	ADJ		ADJ__masculine__singular__proper	AMOD	Horizonte
Horizonte	horizonte	NOUN		NNP__masculine__singular__proper	POBJ	em
.	.	PUNCT		PUNCT__not_proper	P	mora

Prós e contras

- Prós
 - Uniformização dos casos
 - Tratamento mais “fácil” de casos “longos” (como títulos)
- Contras
 - Tradição linguística e do PLN
 - Impacto na compatibilidade de ferramentas e *corpora*
 - *Isso não deveria ser um argumento contrário à revisão de decisões pela comunidade, sobretudo porque inconsistências/incompatibilidades já existem*
 - Ferramentas e aplicações teriam de considerar o nível dos traços (*features*)

Proposta 2 nos *Issues*

- Deixar as PoS *tags* como estão: PROPN, NOUN...
- Criar uma *feature* adicional ***name:yes***
- Prós
 - Evitaria os contras da outra proposta

Casos do Folha-Kaggle e DANTEStocks

1. Nomes próprios em minúsculas

zika vírus

h. pylori (“Helicobacter pylori”)

bovespa, bradesco

Iren-nm

vale5

2. Palavras funcionais, ora em maiúsculas, ora em minúsculas, em nomes ou títulos

Banco **De** o Brasil

Banco **de** o Brasil

Casos do Folha-Kaggle e DANTEStocks

3. Palavras funcionais em língua estrangeira, em minúsculas, dentro de um título de filme ou nome de música

*Cat **on a** Hot Tin Roof*

4. Palavras plenas e funcionais, em minúsculas, dentro de um título de filme ou nome de música

Harry Potter **e o cálice sagrado**

5. Nomes próprios com *modificadores (em maiúsculo ou não)

Banco Bradesco / **banco** Bradesco

Lojas Renner

Sistema Cantareira

Casos do Folha-Kaggle e DANTEStocks

5. Acrônimos de língua estrangeira como palavras comuns no português (ou no domínio)

VIP, CEO, CD, DVD

PB (*“price-to-book ratio”*)

SAR (*“stop and reverse”*)

6. Acrônimos (em pt), ora em maiúsculas, ora em minúsculas

jcp (“juro sobre capital próprio”)

JCP

7. Abreviações em maiúsculas

TV

TX (taxa)

CP (compra)

Em direção a uma proposta...

PROPN

- Antropônimos (*propn* de pessoa) e topônimos (*propn* de lugar)
 - Lúcio Funaro
PROPN, PROP
 - Cunha, Eike
PROPN
 - Brasil, Lisboa, Araquarara
PROPN
 - Belo Horizonte
PROPN, PROP
- Mesmo que composto por nome comum OU palavras funcionais
 - Marília Taboada
PROPN, PROP
 - Arian Di Felippo
PROPN, PROP, PROP
 - Thiago Salgueiro Pardo
PROPN, PROP, PROP

Em direção a uma proposta...

PROPN

- Nome que se refere a uma entidade e que não seja traduzível
 - Friboi

PROPN

- Renner

PROPN

- *Macktrader*

PROPN

Em direção a uma proposta...

- Quando uma entidade for nomeada por uma palavra comum (mesmo em maiúscula), esta recebe a *tag* “original”
 - Ofertas, Supermercados, Câmara → **NOUN**
 - Extra ?
- Quando uma entidade for nomeada por **expressão multipalavra**, os elementos constitutivos são anotados com suas *tags* originais
 - Harry Potter e o cálice de fogo
PROPN, PROPN, CCONJ, DET, NOUN, ADP, NOUN
 - Lojas Renner
NOUN, PROPN
 - Remem para o Money
VERB, ADP, DAT, X
 - Boi Gordo
NOUN, ADJ
 - Macktrader Investimentos
PROPN, NOUN

Em direção a uma proposta...

- Acrônimos de PROPN são PROPN
 - Neymar Jr
 - NJ

Em direção a uma proposta...

- Acrônimos de expressões multipalavras são comumente NOUN
 - Partido do Movimento Democrático Brasileiro
NOUN, ADP, NOUN, ADJ, ADJ
 - PMDB
NOUN
 - Índice da Bolsa de Valores de São Paulo
NOUN, ADP, NOUN, ADP, NOUN, ADP, PROPN, PROPN
 - IBOV, Ibovespa
NOUN
 - Assembleia geral extraordinária
NOUN, ADJ, ADJ
 - Ago
NOUN
 - Usinas Siderúrgicas de Minas Gerais
NOUN, ADJ, ADP, NOUN, ADJ
 - USIMINAS
NOUN

A definir...

Demais casos dependentes do *corpus* Kaggle

- Júpiter
- Virgem

Demais casos dependentes do *corpus* DANTE

- Ações
 - **OGX3** , **petr4**
- Empresas?