'Tópicos sintáticos', ou

Apontamentos sobre um experimento de anotação referencial de textos



Maria Clara Paixão de Sousa, 24 de maio de 2018



# Introdução

A partir do trabalho seminal de Galves (2001/2003) – *'Syntax and Style in Padre Antonio Vieira' – diversos* estudos sobre o Português Clássico fundados no conjunto de textos anotados do Corpus Tycho Brahe têm defendido que a gramática deste período é fortemente marcada pela relação entre sintaxe e discurso, e que esta relação seria um fator central para compreender os padrões de ordem atestados nos textos.

A hipótese do condicionamento discursivo da ordem no Português dos séculos XV a XVII está presente, sob diferentes formas e com diferentes implementações, em diversos trabalhos posteriores a Galves (2001), como Paixão de Sousa (2004), Galves e Paixão de Sousa (2005), Galves, Britto e Paixão de Sousa (2005), Cavalcante e Paixão de Sousa (2010), Gibrail (2010), Antonelli (2011), Galves, Cavalcante e Paixão de Sousa (2015), Galves e Gibrail (2016), e Galves e Paixão de Sousa (2017), entre outros.

De uma forma geral, as hipóteses abraçadas por estes trabalhos partem da observação de uma proeminência discursiva na periferia esquerda da sentença, que condicionaria o fronteamento de constituintes argumentais, e consequentemente, todo o padrão de ordenamento dos constituintes maiores e menores na frase.



Esses trabalhos estão solidamente ancorados em resultados de estudos quantitativos apresentados em sucessivos levantamentos sobre os dados do CTB, que mostram que, em textos representativos dos séculos XVI e XVII, a proporção de estruturas 'XV' (na qual o constituinte inicial não é um sujeito) é elevada em comparação ao que se pode medir em textos posteriores, e comparável ao que mostram os estudos sobre as chamadas 'línguas V2', por conta da elevada

sobre as chamadas 'línguas V2', por conta da elevada proporção de construções XVS (ou seja, com constituintes não-sujeitos pré-verbais, e sujeitos lexicais e pós-verbais) nos textos.

A metodologia e os pressupostos teóricos desses trabalhos sobre a ordem no Português Clássico – realizados a partir de grandes bases de dados, nos quais se buscam os reflexos da mudança gramatical nos padrões estatísticos de aspectos sintáticos nos textos (em particular, a ordem) – filiam-nos à linha de investigações em sintaxe histórica no campo gerativo fundada em Kroch (1989, 1994, 2001), e que tem resultado em um conjunto importante de debates sobre a mudança gramatical em diferentes línguas (na história do inglês, por exemplo, Kroch & Taylor 1997; Pintzuk & Kroch 1989; Pintzuk & Taylor 2008; Pintzuk 1995, 2014; Taylor & Pintzuk 2012a).

Entretanto, para além dos padrões de frequência de ocorrência dos dados, a hipótese geral do condicionamento discursivo da ordem nesses trabalhos não está apoiada em análises exaustivas que demonstrem, empiricamente, como esse condicionamento funciona nos textos.

Temos disponíveis algumas análises pontuais de ocorrências características, que indicam o acerto da hipótese, quando combinadas com os dados quantitativos solidamente pesquisados.

Essas análises pontuais se fundam em duas observações centrais: primeiro, a observação de Galves (2001) no sentido de que a colocação de clíticos no texto dos Sermões de A. Vieira (1604) dependia da interpretação do sujeito pré-verbal como um 'tópico' ou um 'foco'.



(1) Ênclises nos Sermões de Vieira (Galves, 2001/2003)

Eis aqui porque David queria que o julgasse Deus, e não os homens:

no Juiso de Deus perdoam-se os peccados como fraquezas:

no juiso dos homens castigam-se as valentias como peccados.

Deus julga como Juiz; os homens julgam como judiciários. Deus julga a cada um pelo que é, os homens julgam a cada um pelo que são.

Deus julga-nos a nós por nós; os homens julgam-nos a nós por si. 2) Próclises nos Sermões de Vieira (Galves, 2001/2003)

Que effeitos ha-de causar nos homens a vista d'aquelles signaes?

O Evangelhista o refere por bem extraordinarios termos: Arescentibus hominibus præ timore, et expectatione, quæ supervenient universo orbi.

O Evangelho o diz: Erunt signa in sole, et luna, et stellis.

O mesmo Texto o declara admiravelmente no que logo acrescenta:...

Segundo, a partir disso, a observação de Paixão de Sousa (2004) de que os sujeitos pós-verbais nos textos clássicos não podiam ser interpretados exclusivamente como 'focos' (como é o caso de VS no Português Europeu), e que os sujeitos pré-verbais podiam ser 'focos' ou 'tópicos' (definidos ali de modo muito amplo, a partir do critério tradicional do contraste 'informação nova/informação velha'). Essa ideia, junto com a de Galves (2001), se expande em trabalhos posteriores para a interpretação dos sujeitos em geral, não apenas aqueles combinados com a ocorrência de clíticos.

1) SV com 'alternância de referentes' (Galves e Paixão de Sousa, 2017)

# (a) Vieira, sermões:

Deus julga como Juiz; os homens julgam como judiciários.

Deus julga a cada um pelo que é, os homens julgam a cada um pelo que são.

Deus julga-nos a nós por nós; os homens julgam-nos a nós por si.

## (b) Vieira, Cartas:

Duarte Nunes me avisa tem comprado sessenta peças de boa artilharia para as duas naus...

Jerónimo Nunes me escreveu hoje tivera carta de V. Ex.a. com recado

# (c) Pinto, Peregrinação:

O capitão-mor lhe respondeu que os embaixadores tinham seguro para suas pessoas...

O brâmene lhe deu por isso seus agradecimentos

2) VS com 'continuidade de referentes' (Galves e Paixão de Sousa, 2017):

Pinto, 1510:

Ao mercador que me trouxe mandou Pero de Faria dar sessenta cruzados

A mim me mandou o capitão agasalhar em casa de um escrivão da feitoria 2) VS com 'continuidade de referentes' (Galves e Paixão de Sousa, 2017):

Pinto, 1510:

Ao mercador que me trouxe mandou Pero de Faria dar sessenta cruzados

A mim me mandou o capitão agasalhar em casa de um escrivão da feitoria

3) 'Proeminência à esquerda' (Paixão de Sousa, 2009, 2012):

Cardim, 1584:

Este papagaio he formosissimo, e nelle se achão quasi todas as côres em grande perfeição, sc, vermelho, verde, amarello, preto, azul, pardo, côr de rosmaninho, e

de todas estas cores tem o corpo salpicado, e espargido.

Estes tambem fallão, e têm mais huma vantagem que he criar em casa, e tirar seus filhos, pelo que são de grande estima.



(3) 'Proeminência à esquerda' (Paixão de Sousa, 2009, 2012):

Cardim, 1584:

Esta cobra he muito formosa,

a cabeça tem vermelha, branca e preta, e assi todo o corpo.

Esta he a mais peçonhenta de todas, anda de vagar, e vive em as gretas da terra

) 'Proeminência à esquerda' e diacronia das valências verbais (*Paixão de Sousa, 2008, 2009, 2012*):

Góis:

Uma chamada Dona Urraca casou com o Conde Dom Reymão de Tolosa



(3) 'Proeminência à esquerda' e diacronia das valências verbais (*Paixão de Sousa, 2008, 2009, 2012*):

Góis:

**Uma chamada Dona Urraca**i proi casou com o Conde Dom Reymão de Tolosa



(3) 'Proeminência à esquerda' e a interpretação de sujeitos nulos (*Paixão de Sousa, 2009, 2012*):

Cardim, 1584:

quando *pro* vão para os apanhar, *pro* botão-lhes aquella tinta diante dos olhos (3) 'Proeminência à esquerda' e a interpretação de sujeitos nulos (*Paixão de Sousa, 2009, 2012*):

Cardim, 1584:

Ora: tendo em vista o volume e a importância dos trabalhos que tem se dedicado a estudar este aspecto da sintaxe do Português Clássico, seria extremamente desejável ter-se à disposição uma metodologia e uma ferramenta que permitissem efetivamente pesquisar, da forma o mais objetiva possível, o estatuto deiscursivo dos constituintes nos textos – preferencialmente, uma ferramenta que trabalhasse em harmonia com as anotações linguísticas (morfossintática e sintática) já desenvolvidas no bojo das pesquisas junto ao CTB.

É a este objetivo que o experimento apresentado neste artigo se propôs.



Partindo da hipótese geral de Galves e Paixão de Sousa (2010, 2017) no sentido de que os constituintes argumentais, no Português Clássico, são fronteados de acordo com sua proeminência discursiva; e, mais particularmente, da ideia sugerida em Paixão de Sousa (2009, 2012), de que essa propriedade de proeminência pode ser analisada obervando-se a alternância de referentes em sequências narrativas e descritivas dos textos, desenvolvi uma anotação experimental que, basicamente, explicita a cadeia de referentes formada pelos constituintes num texto.

Para isso, a anotação codifica os diferentes referentes expressos por cada sintagma nominal de um texto, e indexa as co-referências formadas entre eles — sejam lexicais ou nulos.



Esta anotação é ainda bastante rudimentar, e não consititui uma anotação de estrutura informacional, nem uma anotação semântica: trata-se ainda, e de forma incipiente, simplesmente de uma anotação de cadeias referenciais.



Acredito, entretanto, que possa servir como base para refinamentos futuros, formando uma ferramenta com base na qual se poderá analisar objetivamente o ordenamento dos constituintes nos textos conforme seu encadeamento discursivo, com reflexos nos trabalhos que seguem as hipóteses resumidas acima.



Nas seções que seguem, são detalhados os princípios e a técnica básica da anotação experimentada, e os desafios colocados para a perspectiva de sua aplicação.



# Projetos envolvidos





https://humanidadesdigitais.org/HB



# 2015-2016:

Sintaxe e estrutura informacional da "História da província Sancta cruz", 1576: primeiros estudos para as histórias do Brasil Pesquisa realizada como estágio de pós-doutoramento na Universidade de York, Reino Unido, e abrigada por Susan Pintzuk



https://humanidadesdigitais.org/HB





https://humanidadesdigitais.org/HB



## Isabel de Lamas

Isabel de Lamas cristă velha mulher de Francisco Martins, natural deste Pernambuco ré presa que presente esta como quem tinha pouca reverência ao venerandissimo Tribunal do Santo Oficio e pouco temor de Deus e da condenação de sua alma vindo à mesa jurou nela falso em caso grave pertencente ao Santo Oficio

1594



Nomeada em processo inquisitorial, Olinda, 1594. Acusada de perjúrio. Incluída no mesmo processo que seu marido, acusado de sodomia.

Documento acessível, digitalizado Não há edição filológica conhecida Não há menção na bibliografia conhecida

## Documentos primários conhecidos

 $\label{eq:continuous} Tribunal \ do \ Santo \ Oficio \ (TSO) . \ Processo \ inquisitorial . \ Vila \ de \ Olinda \ Capitania \ de \ pernambuco \ , 1594 . \\ ANTT - PT/TT/TSO-IL/028/09480 \ . \ http://digitarq.arquivos.pt/details?id=2309626 \ . \ [NN | 018 | IL] \ . \\ NN | 108 | IL] \ .$ 

## Dados biográficos conhecidos

Idade: 18 anos Estatuto social: Cristă velha Estado civil: Casada Morada: Pernambuco, Brasil Naturalidade: Pernambuco, Brasil

#### Dados internos

Isabel de Lamas, Nomeada em documento primário [NN | 018 | IL]

Chave de pesquisa: Sodomia Catalogado por: MLS

# https://humanidadesdigitais.org/MAP/



# 2017-atual

Projeto P.U.B – USP com

Vanessa Martins do Monte

Beatriz Moreira de Souza Isabelle de Moura Vitorino Letícia Junqueira Sena Alves Mariana Lourenco Sturzeneker Tais Estéfanie Pacheco Ferreira

## Isabel de Lamas

Isabel de Lamas cristă velha mulher de Francisco Martins, natural deste Pernambuco ré presa que presente esta como quem tinha pouca reverência ao venerandissimo Tribunal do Santo Oficio e pouco temor de Deus e da condenação de sua alma vindo à mesa jurou nela falso em caso grave pertencente ao Santo Oficio

1594



Nomeada em processo inquisitorial, Olinda, 1594. Acusada de perjúrio. Incluída no mesmo processo que seu marido, acusado de sodomia.

Documento acessível, digitalizado Não há edição filológica conhecida Não há menção na bibliografia conhecida

## Documentos primários conhecidos

 $\label{eq:continuous} Tribunal \ do \ Santo \ Oficio \ (TSO) . \ Processo \ inquisitorial . \ Vila \ de \ Olinda \ Capitania \ de \ pernambuco \ , 1594 . \\ ANTT - PT/TT/TSO-IL/028/09480 \ . \ http://digitarq.arquivos.pt/details?id=2309626 \ . \ [NN | 018 | IL] \ . \\ NN | 108 | IL] \ .$ 

## Dados biográficos conhecidos

Idade: 18 anos Estatuto social: Cristă velha Estado civil: Casada Morada: Pernambuco, Brasil Naturalidade: Pernambuco, Brasil

#### Dados internos

Isabel de Lamas, Nomeada em documento primário [NN | 018 | IL]

Chave de pesquisa: Sodomia Catalogado por: MLS

# https://humanidadesdigitais.org/MAP/



## Isabel de Lamas

Isabel de Lamas cristà velha mulher de Francisco Martins, natural deste Pernambuco ré presa que presente esta como quem tinha pouca reverência ao venerandissimo Tribunal do Santo Oficio e pouco temor de Duse e da condenação de sua alma vindo à mesa jurou nela falso em caso grave pertencente ao Santo Oficio

1594



Nomeada em processo inquisitorial, Olinda, 1594. Acusada de perjúrio. Incluída no mesmo processo que seu marido, acusado de sodomia.

Documento acessível, digitalizado Não há edição filológica conhecida Não há menção na bibliografia conhecida

## Documentos primários conhecidos

 $\label{eq:continuous} Tribunal \ do \ Santo \ Oficio \ (TSO) . \ Processo \ inquisitorial . \ Vila \ de \ Olinda \ Capitania \ de \ pernambuco \ , 1594 . \\ ANTT - PT/TT/TSO-IL/028/09480 \ . \ http://digitarq.arquivos.pt/details?id=2309626 \ . \ [NN | 018 | IL] \ . \\ NN | 108 | IL] \ .$ 

## Dados biográficos conhecidos

Idade: 18 anos Estatuto social: Cristă velha Estado civil: Casada Morada: Pernambuco, Brasil Naturalidade: Pernambuco, Brasil

#### Dados internos

Isabel de Lamas, Nomeada em documento primário [NN | 018 | IL]

Chave de pesquisa: Sodomia Catalogado por: MLS

# https://humanidadesdigitais.org/MAP/









```
(NP-ACC (ID ID.3088.H-AC-NNC=5366.1904/000)
                (N-P=1904/000 Bugios)
                (QP *ICH*-3))
        (HV-P há)
       (PP (P n@)
            (NP (ID ID.3089.M-RR-DNN=0046/080) (D-F @a) (N=1905:0046/082 terra)))
        (QP-3 (QP (Q-P muitos))
              (CONJP (CONJ e)
                     (PP (P de)
                         (NP (ID ID.3090.H-RC-NNQ=8052.1248/000) (Q-F-P muitas) (N-P=1906:1248/003 castas)))))
        (CP-ADV (C como)
                (IP-SUB (NP-SBJ-2 *exp*)
                        (ADVP (ADV já))
                        (NP-SE-1-2 (ID COSE) (CL se))
                        (VB-P sabe)))
       (.:)
(ID G 008,23.359))
```

#### Numerando os nomes

Sub-etiquetas numéricas

( (IP-MAT (NP-SBJ \*exp\*)

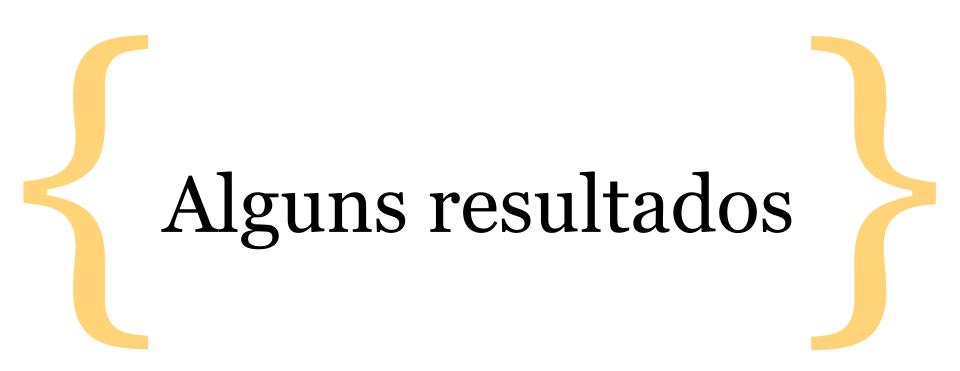
(adicionadas a N, NPR)

```
( (IP-MAT (NP-SBJ *exp*)
          (NP-ACC (ID ID.3088.H-AC-NNC=5366.1904/000)
                  (N-P=1904/000 Bugios)
                  (OP *ICH*-3))
          (HV-P há)
          (PP (P n@)
              (NP (ID ID.3089.M-RR-DNN=0046/080) (D-F @a) (N=1905:0046/082 terra)))
          (QP-3 (QP (Q-P muitos))
                (CONJP (CONJ e)
                       (PP (P de)
                           (NP (ID ID.3090.H-RC-NNQ=8052.1248/000)
                                                                     (Q-F-P muitas) (N-P=1906:1248/003 castas)))))
          (CP-ADV (C como)
                  (IP-SUB (NP-SBJ-2 *exp*)
                          (ADVP (ADV já))
                          (NP-SE-1-2 (ID COSE) (CL se))
                          (VB-P sabe)))
          (.:)
  (ID G 008,23.359))
```

#### Elementos ID

Contendo os ínices dos referentes (incluidos sob NP, WNP)





## em números

- **22.944** palavras =
  - **4.165** nomes
  - 8.033 sintagmas nominais
  - 1.265 referentes individuais

ıs,

es,

is,

is.



um exemplo



(ID G\_008,21.304)



{

Outros há também nestas partes muito para notar, e mais fora da comum semelhança dos outros animais (a meu juízo) que quantos até agora se tem visto. Chamam-lhes Tatus, e são quase tamanhos como leitões: tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que parece totalmente um cavalo armado. Tem um rabo comprido todo coberto do mesmo casco: o focinho é como de leitão, ainda que mais delgado algum tanto, e não bota mais fora do casco que a cabeça. Tem as pernas baixas, e criam-se em covas como coelhos. A carne destes animais é a melhor e a mais estimada que há nesta terra, e tem o sabor quase como de galinha.

Outros há também nestas partes muito para notar, e mais fora da comum semelhança dos outros animais (a meu juízo) que quantos até agora se tem visto. Chamam-lhes Tatus, e \_\_\_ são quase tamanhos como leitões: \_\_\_ tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que parece totalmente um cavalo armado. \_\_\_ Tem um rabo comprido todo coberto do mesmo casco: o focinho é como de leitão, ainda que mais delgado algum tanto, e \_\_\_ não bota mais fora do casco que a cabeça. \_\_\_ Tem as pernas baixas, e \_\_\_ criam-se em covas como coelhos.

A carne destes animais é a melhor e a mais estimada que há nesta terra, e tem o sabor quase como de galinha.

## cadeia referencial do 'tatu', 7520

```
(ID H-EC-000=7520#1473##5400/000) Outros
(ID M-EE-PPP=7520/001) - lhes
(ID H-BB-NOM=7520:1692/000) Tatus
(ID M-EE-PPP=7520/002) *pro*
(ID M-EE-PPP=7520/003) *pro*
(ID M-EE-PPP=7520/004) *pro*
(ID M-EE-PPP=7520/005) *pro*
(ID M-EE-PPP=7520/006) *pro*
(ID M-EE-PPP=7520/007) *pro*
(ID M-EE-PPP=7520/008) *pro*
(ID M-EE-PPP=7520/008) *pro*
(ID M-RR-ENN=7520/009) estes animais
```

## {funcionamentos de navegação

## funcionamento de navegação – texto:

Outros há também n<sub>0</sub> ▶ @estas partes muito para notar , e mais fora d<sub>0</sub> ▶ @a comum semelhança d<sub>0</sub> ▶ @os outros animais ( a ▶ meu juízo ) que ▶ ▶ quantos até agora se tem visto .

```
D G_008,21.301

➤ *pro* Chamam- ➤ -lhes ➤ Tatus ,

D G_008,21.302

e ➤ *pro* são quase tamanhos como ➤ leitões :
```

ID G 008,21.303

### funcionamento de navegação – blocos:

```
também
                           ne ▶ eestas partes
Outros OBI
muito para
            notar
 mais fora de ► ea comum semelhança de ► eos outros animais
                                                                      a ▶ meu juízo
                            até
             ▶ quantos
                                          se tem visto
   que
                                agora
                                                         SBJ
```

ID G\_008,21.301



### funcionamento de navegação - árvore:

```
( (IP-MAT (NP-SBJ *exp*)
          (NP-ACC (ID ID.2683.H-EC-000=7520#1473##5400/000) (OUTRO-P Outros))
          (HV-P há)
          (ADVP (ADV também))
          (PP (P n@)
              (NP (ID ID.2684.M-RR-ENN=0013/013) (D-F-P @estas) (N-P=1688:0013/021 partes)))
         (PP (PP (Q muito)
                  (P para)
                  (IP-INF (VB notar)))
             (, ,)
             (CONJP (CONJ e)
                     (ADVP (ADV-R mais)
                           (ADV fora)
                           (PP (P d@)
                               (NP (ID ID.2685.H-RC-DNC=5975.0053/000)
                                   (D-F @a)
                                   (ADJ-G comum)
                                   (N=1689:0053/005 semelhanca)
                                   (PP (P d@)
                                       (NP (ID ID.2686.H-RC-DNO=7521.1473/000) (D-P @os) (OUTRO-P outros) (N-P=1690:1473/008 animais)))))
                           ([paren] [paren])
                           (PP-PRN (P a)
                                   (NP (ID ID.2687.H-RC-NNS=1323$0000/000) (PRO$ meu) (N=1691:1323/001 juízo)))
                          ([paren] [$$paren])
                           (CP-CMP-2 (C que)
                                     (IP-SUB (NP-SBJ (ID ID.2688.H-BB-CLA=6914/000)
                                                     (CP-FRL (WNP-3 (ID ID.2689.H-RC-DNC=5522#1473/000) (WD-P quantos))
                                                             (IP-SUB (NP-SBJ-4 *T*-3)
                                                                     (PP (P até)
                                                                         (ADVP (ADV agora)))
                                                                     (NP-SE-4 (ID COSE) (CL se))
                                                                     (TR-P tem)
                                                                     (VB-PP visto))))))))
         (...)
  (ID G_008,21.301))
```

### aspectos sintáticos interessantes

- 1# os sujeitos lexicais, suas posições, e seus referentes
- 2# os sujeitos nulos e seus referentes

1# Os sujeitos lexicais, suas posições, e seus referentes

**Outros** há também nestas partes muito para notar, e mais fora da comum semelhança dos outros animais (a meu juízo) que quantos até agora se tem visto. \_\_\_ Chamam-lhes Tatus, e \_\_\_ são quase tamanhos como leitões: \_\_\_ tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que parece totalmente um cavalo armado. Tem um rabo comprido todo coberto do mesmo casco: o focinho é como de leitão, ainda que mais delgado algum tanto, e \_\_\_ não bota mais fora do casco que a cabeça. Tem as pernas baixas, e \_\_\_ criam-se em covas como coelhos.

A carne destes animais é a melhor e a mais estimada que há nesta terra, e tem o sabor quase como de galinha.

Outros (i) há também nestas partes muito para notar, e mais fora da comum semelhança dos outros animais (a meu juízo) que quantos até agora se tem visto. \_\_\_ Chamam-lhes Tatus, e \_\_\_ são quase tamanhos como leitões: \_\_\_ tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que parece totalmente um cavalo armado. \_\_\_ Tem um rabo comprido todo coberto do mesmo casco:

o focinho (ii) é como de leitão, ainda que mais delgado algum tanto, e \_\_\_ não bota mais fora do casco que a cabeça. \_\_\_ Tem as pernas baixas, e \_\_ criam-se em covas como coelhos.

A carne destes animais (iii) é a melhor e a mais estimada que há nesta terra, e tem o sabor quase como de galinha.

Outros (i) há também nestas partes muito para notar, e mais fora da comum semelhança dos outros animais (a meu juízo) que quantos até agora se tem visto. \_\_\_(iv) Chamam-lhes Tatus, e \_\_\_(i) são quase tamanhos como leitões: \_\_\_(i) tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que parece totalmente um cavalo armado. \_\_\_(i) Tem um rabo comprido todo coberto do mesmo casco:

o focinho (ii) é como de leitão, ainda que mais delgado algum tanto, e (i) não bota mais fora do casco que a cabeça. (ii) Tem as pernas baixas, e (iii) criam-se em covas como coelhos.

A carne destes animais (iii) é a melhor e a mais estimada que há nesta terra, e \_\_\_(iii) tem o sabor quase como de galinha.

Outros (i) há também nestas partes muito para notar, e mais fora da comum semelhança dos outros animais (a meu juízo) que quantos até agora se tem visto. \_\_\_(iv) Chamam-lhes Tatus, e \_\_\_(i) são quase tamanhos como leitões: \_\_\_(i) tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que parece totalmente um cavalo armado. \_\_\_(i) Tem um rabo comprido todo coberto do mesmo casco:

o focinho (ii) é como de leitão, ainda que mais delgado algum tanto, e (i) não bota mais fora do casco que a cabeça. (ii) Tem as pernas baixas, e (iii) criam-se em covas como coelhos.

A carne destes animais (iii) é a melhor e a mais estimada que há nesta terra, e \_\_\_(iii) tem o sabor quase como de galinha.

2# os sujeitos nulos e seus referentes



{

(1) Chamam-lhes Tatus, e são quase tamanhos como leitões:

(2) tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que parece totalmente um cavalo armado.

(3) o focinho é como de leitão, ainda que mais delgado algum tanto, e não bota mais fora do casco que a cabeça.

 $\left\{ 
ight.$ 

(1) <u>pro</u> Chamam-lhes Tatus, e <u>pro</u> são quase tamanhos como leitões:

(2) <u>pro</u> tem um casco como de cágado, o qual é repartido em muitas juntas como lâminas e proporcionado de maneira, que <u>pro</u> parece totalmente um cavalo armado.

(3) o focinho é como de leitão, ainda que mais delgado algum tanto, e não bota mais fora do casco que a cabeça.

{

#### Indexação na leitura imediata de um falante do PB:

- (1) <u>pro</u> (i) Chamam-<u>lhes</u>(ii) Tatus, e <u>pro</u> (i) são quase tamanhos como leitões:
- (2) <u>pro</u> (i) tem <u>um casco como de cágado</u> (iii), <u>o qual</u> (iii) é repartido em muitas juntas como lâminas e proporcionado de maneira, que <u>pro</u> (iii) parece totalmente um cavalo armado.
- (3) <u>o focinho (iv)</u> é como de leitão, ainda que mais delgado algum tanto, e pro (iv) não bota mais fora do casco que a cabeça.

{

#### Indexação autorizada pelo contexto (e refletida na anotação sintática)

- (1) <u>pro</u> (i) Chamam-<u>lhes</u>(ii) Tatus, e <u>pro</u> (ii) são quase tamanhos como leitões:
- (2) <u>pro</u> (ii) tem <u>um casco como de cágado</u> (iii), <u>o qual</u> (iii) é repartido em muitas juntas como lâminas e proporcionado de maneira, que <u>pro</u> (ii) parece totalmente um cavalo armado.
- (3) <u>o focinho (iv)</u> é como de leitão, ainda que mais delgado algum tanto, e pro (ii) não bota mais fora do casco que a cabeça.

\* buscas com Corpus Search



```
define: port.def
print indices: t
       IP*
node:
query: (IP-MAT* iDominates
                                     NP-SBJ*)
 AND (NP-SBJ* iDominates
                                      !*pro*|*exp*)
                                      !NP-SE*)
 AND
      (IP-MAT* iDominates
                                     VB-D|VB-I|VB-P|VB-R|VB-RA|VB-SD|VB-SP|VB-SR|
 AND
      (NP-SBJ* HasSister
                                     SR-D|SR-I|SR-P|SR-R|SR-RA|SR-SD|SR-SP|
                                      ET-SR|ET-D|ET-I|ET-P|ET-R|ET-RA|ET-SD|ET-SP|ET-SR|
                                     HV-D|HV-I|HV-P|HV-R|HV-RA|HV-SD|HV-SP|HV-SR|
                                     TR-D|TR-I|TR-P|TR-R|TR-RA|TR-SD|TR-SP|TR-SR)
                                     VB-D|VB-I|VB-P|VB-R|VB-RA|VB-SD|VB-SP|VB-SR|
 AND
      (NP-SBJ* precedes
                                     SR-D|SR-I|SR-P|SR-R|SR-RA|SR-SD|SR-SP|
                                      ET-SR|ET-D|ET-I|ET-P|ET-R|ET-RA|ET-SD|ET-SP|ET-SR|
                                     HV-D|HV-I|HV-P|HV-R|HV-RA|HV-SD|HV-SP|HV-SR|
                                     TR-D|TR-I|TR-P|TR-R|TR-RA|TR-SD|TR-SP|TR-SR)
AND (NP-SBJ* iDominates ID)
AND (ID iDominates H-AA-NNN*)
```

### \* padrões encontrados: posição e referência dos sujeitos lexicais

pre-verbal versus post-verbal subjects

0,33

0,67

vs/sv

sv/vs

0,67

0,00





pre-verbal subjects, heads contrasted with general rate

0,67 0,66

sv/vs

sv/vs, heads

0,00



#### padrões encontrados: posição e referência dos sujeitos lexicais

pre-verbal subjects mentions contrasted with general rate

0,67 0,70

vs/sv

0,00

0,67

# \* padrões encontrados: posição e referência dos sujeitos lexicais

pre-verbal subjects non-modified phrases, heads, contrasted with general rate

0,76

0,67

0,67

sv/vs

0,00



# padrões encontrados: posição e referência dos sujeitos lexicais

pre-verbal subjects phrases with demonstratives contrasted with general rate

0,80

0,67

0,67

vs/sv

with demonstratives

0,00



\* conclusão parcial



A primeira conclusão do experimento é a de que neste texto, até há uma correlação positiva entre a propriedade referencial de 'informação velha' e a posição préverbal do sujeito - mas esta correlação é muito tênue.

```
* conclusão parcial
```

67% de SV em geral; 70% de SV com sujeitos "velhos"; 66% de SV com sujeitos "novos".



{\\*

A única conclusão interessante até agora não parece depender do estatuto referencial dos sujeitos quanto a sua menção anterior no texto.



A única conclusão interessante até agora não parece depender do estatuto referencial dos sujeitos quanto a sua menção anterior no texto.



\*

A boa notícia é que ninguém vai precisar escarafunchar textos contando referentes.



A boa notícia é que ninguém vai precisar escarafunchar textos

contando referentes.

A má (?) notícia é que precisamos ser mais cuidadosos com a estrutura interna dos NPs sujeitos



A conclusão parcial é que o traço mais importante a diferenciar sujeitos pré-verbais e pós-verbais é sua especificidade



**{** \*

A conclusão parcial seria que o traço mais importante a diferenciar sujeitos pré-verbais e pós-verbais é sua especificidade, que não parece se reduzir ao estatuto de informação "nova" ou "velha".

\*

O sujeito pré-verbal típico deste texto é '+específico'.

Ou: quanto mais específico é o sintagma nominal sujeito, mais ele tende a ser pré-verbal.

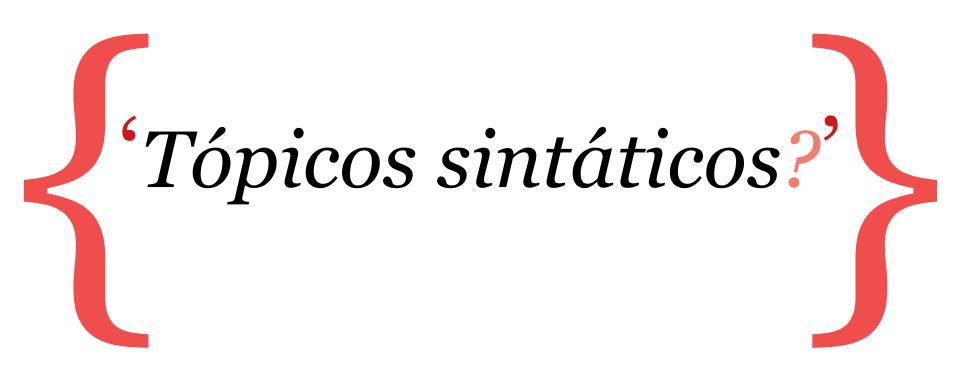
**{** \*

## Mas como assim específico



**\*** 

I do not believe this is a generalisation that would have any bearing on a more general definition of 'topic' in an informational or pragmatic perspective.



{ {

> Para um sintaticista, as definições disponíveis para o conceito "tópico" podem ser desafiantes, pois dependem de uma apreciação refinada do encadeamento argumentativo do texto que não é nosso objetivo primário de pesquisa – e nos parecem, em alguns momentos, muito pouco objetivas.



Para um sintaticista, as definições disponíveis para o conceito "tópico" podem ser desafiantes, pois dependem de uma apreciação refinada do encadeamento argumentativo do texto que não é nosso objetivo primário de pesquisa – e nos parecem, em alguns momentos, muito pouco objetivas.

<u>nii</u>

ne

<u>ne</u>



aquele muito católico e sereníssimo Príncipe el-Rei Dom Manuel a Índia o Oriente a Santa Cruz **Bugios** Onças Sapucaias

aquele muito católico e sereníssimo Príncipe el-Rei Dom Manuel

a Índia o Oriente

a Santa Cruz

Bugios

Onças

Mandiocas

Sapucaias

Informação "nova"? Informação "velha"?

Sapucaias

{

Afinal, o que é um tópico?

A definição mais geral, de 'informação velha' (um termo já tão gasto...), pode significar – em termos pragmáticos – uma informação compartilhada ou um referente já mencionado. Note-se que, evidentemente, o fato de um referente ter diso ou não ter sido mencionado não significa imediatamente que a informação que ele veicula seja "nova" ou "velha" (cf. *Índia, sapucaias*). ..

 $\left\{ 
ight.$ 

Mas é claro que há!

Pela estrutura dos sintagmas nominais...



 $\left\{ 
ight.$ 

Mas é claro que há!

Pela estrutura dos sintagmas nominais...

(... que, por sinal, foi o indicador mais seguro da probabilidade da posição do sujeitos no nosso experimento...)



{

**aquele** muito católico e sereníssimo Príncipe el-Rei Dom Manuel

**a** Índia

Oriente

a Santa Cruz

Bugios Onças



Mandioca Sapucaias Informação "velha"!

Informação "nova"? Informação "velha"?

```
'Esta provincia é a vista
muito deliciosa e fresca em
grande maneira'
```

muitagilhag aug a dividam

'Este rio tem na entrada

```
{
```

'Isto geralmente se costuma nestas partes'



```
E os índios da terra que ali
se ajuntaram ouviam tudo
com muita quietação...'
```

'As fontes que há na terra,

{

'Outros animais há nesta provincia muito feros...'



```
'Os louros têm um cabelo
muito fino'
'Os pardos se acham daí
para o Norte em todas as
```

autrag agnitaniag'

 $\left\{ 
ight.$ 

In this context, the 'profile' of phrases that tend to appear in the positions that I was calling 'prominent' is of high specificity, so that I could call them now syntactic topics. They do tend to be 'old information' ('mentions' in the terminology of my annotation) – but more than that: they tend to be constructed in such a way so that the indication of this 'given-ness' is explicit in the construction.



{

What is, after all, a topic?

The more general way to describe it is as 'old information' (a worn and torn term I was trying to avoid). In pragmatic terms, this may mean a piece of shared knowledge or an already mentioned referent – in any case, a 'given', something that is taken as 'understood' in the (textual or broader) context. The more theoretical ways to describe 'topics' are varied, as we know; but most of them do rely on this basic assumption – that 'topicality' is related to 'givenness'

{

Syntactic topics would be constructions that may be identified as carrying a strong component of 'given-ness' – even before any interpretation or resource to the broader context.



 $\left\{ 
ight.$ 

They would be classified as 'topics' for the simple reason that the common knowledge presupposed in the context of communication is translated into formal material that explicitly point out, make reference, and specify: materials such as determiners, modifiers, and complements.



 $\left\{ 
ight.$ 

They would be classified as 'topics' for the simple reason that the common knowledge presupposed in the context of communication is translated into formal material that explicitly point out, make reference, and specify: materials such as determiners, modifiers, and complements.





Combining the linguistic aspects and the technical aspects mentioned above, my main conclusion over these first empirical results is that it will be necessary to make this annotation better in the future by conducting further queries with Corpus Search and thereby trying to find the exact

factors that are at play; but that, nevertheless, the stage of the

technique up to this point is a good start.

Mainly, because it is allowing for patterns to be found in the text that are at the same time independent of intuitive interpretation, and do not entirely clash with it; rather, they make it more precise, and allow it to be reproduced for future texts (as was, in fact, the main idea in starting the

annotation).

## Obhigada!

## {Referências

Ferreira, M.B. *Sujeitos Nulos e Objetos Nulos no Português Brasileiro*. Tese de Mestrado. Universidade Estadual de Campinas. 2000.

Figueiredo e Silva, M.C. *A Posição do sujeito no Português Brasileiro: Frases finitas e infinitivas*. Campinas: Editora da Unicamp. 1996.

Modesto, M. *On the Identification of Null Arguments*. PhD Dissertation, University of South California. 2000.

Modesto, M. Topic Prominence and Null Subjects. In T. Biberauer (org), *The Limits of Syntactic Variation*. Amsterdam: J. Benjamins. 2008.

Negrão, E.V. O português brasileiro: uma língua voltada para o discurso. *Boletim da Associação Brasileira de Lingüística (ABRALIN)*, Fortaleza, v. 25, p. 183-199. 2000.

10

## A anotação em detalhe



REINANDO aquele muito católico e sereníssimo **Príncipe el-Rei Dom MANUEL**, fez-se uma **frota** para a **Índia** de que ia por **capitão** mór **Pedro Álvares Cabral**: que foi a segunda **navegação** que fizeram os **Portugueses** para aquelas **partes** do **Oriente**' [G\_008.6,1]

```
(NPR
            Príncipe)
(NPR
            el-Rei)
(NPR
            Dom)
(NPR
            MANUEL)
            frota)
(N
(NPR
            Índia)
            capitão)
(N
            Pedro)
(NPR
(NPR
            Álvares)
            Cabral)
(NPR
(N
            navegação)
(NPR-P
            Portugueses)
(N-P
            partes)
(NPR
            Oriente)
```

{

REINANDO aquele muito católico e sereníssimo **Príncipe el-Rei Dom MANUEL**, fez-se uma **frota** para a **Índia** de que ia por **capitão** mór **Pedro Álvares Cabral**: que foi a segunda **navegação** que fizeram os **Portugueses** para aquelas **partes** do **Oriente**' [G\_008.6,1]

```
(NPR=0001
            Príncipe)
(NPR=0002
            el-Rei)
(NPR=0003
            Dom)
(NPR=0004
            MANUEL)
(N=0005
            frota)
(NPR=0006
            Índia)
(N=0007
            capitão)
            Pedro)
(NPR=0008
(NPR=0009
            Álvares)
            Cabral)
(NPR=0010
(N=0011)
            navegação)
(NPR-P=0012 Portugueses)
(N-P=0013)
            partes)
(NPR=0014
            Oriente)
```

REINANDO aquele muito católico e sereníssimo **Príncipe el-Rei Dom MANUEL**, fez-se uma **frota** para a **Índia** de que ia por **capitão** mór **Pedro Álvares Cabral**: que foi a segunda **navegação** que fizeram os **Portugueses** para aquelas **partes** do **Oriente**' [G\_08.6,1]

(NPR=0001 Príncipe) (NPR=0002 el-Rei) (NPR=0003 Dom) (NPR=0004 MANUEL) (N=0005 frota) Índia) (NPR=0006 (N=0007 capitão) (NPR=0008 Pedro) (NPR=0009 Álvares) (NPR=0010 Cabral) (N=0011)navegação) (NPR-P=0012 Portugueses) (N-P=0013)partes) (NPR=0014 Oriente)

Nomes numerados automaticamente por expressões regulares focando as etiquetas morfológicas N(-P) e NPR(-P) REINANDO aquele muito católico e sereníssimo **Príncipe el-Rei Dom MANUEL**, fez-se uma **frota** para a **Índia** de que ia por **capitão** mór **Pedro Álvares Cabral**: que foi a segunda **navegação** que fizeram os **Portugueses** para aquelas **partes** do **Oriente**' [G\_08.6,1]

```
(NPR=0001
            Príncipe)
(NPR=0002
            el-Rei)
(NPR=0003
            Dom)
(NPR=0004
            MANUEL)
(N=0005
            frota)
            Índia)
(NPR=0006
(N=0007
            capitão)
(NPR=0008
            Pedro)
            Álvares)
(NPR=0009
(NPR=0010
            Cabral)
(N=0011)
            navegação)
(NPR-P=0012 Portugueses)
(N-P=0013)
            partes)
(NPR=0014
            Oriente)
```

```
\left\{ 
ight.
```

```
(a)
        (NP-SBJ
                                          (D-P os)(NPR-P=0012
 Portugueses))
(b)
        (NP
                                           (NPR-P=0151)
 Portugueses))
(c)
        (NP
                                           (NPR-P=0509)
 Portugueses))
(d)
        (NP-SBJ
                                          (D-P os)(NPR-P=0548)
 Portugueses))
(e)
        (NP
                                           (PRO eles))
                                          (CL -lhes)) ...
(f)
        (NP-DAT
```

```
\left\{ 
ight.
```

```
(NP-SBJ
                                          (D-P os)(NPR-P=0012
(a)
 Portugueses))
(b)
        (NP
                                          (NPR-P=0151:0012
 Portugueses))
(c)
        (NP
                                          (NPR-P=0509:0012
 Portugueses))
       (NP-SBJ
                                          (D-P os)(NPR-P=0548:0012
(d)
 Portugueses))
(e)
        (NP
                                          (PRO eles))
                                          (CL -lhes)) ...
(f)
        (NP-DAT
```

```
{
```

```
(NP-SBJ
                                          (D-P os)(NPR-P=0012/000
(a)
 Portugueses))
(b)
        (NP
                                          (NPR-P=0151:0012/001
 Portugueses))
(c)
        (NP
                                          (NPR-P=0509:0012/003
 Portugueses))
       (NP-SBJ
(d)
                                          (D-P os)(NPR-P=0548:0012/004
 Portugueses))
(e)
        (NP
                                          (PRO eles))
(f)
        (NP-DAT
                                          (CL -1hes)) ...
```

```
{
```

```
(NP-SBJ (ID
                                       (D-P os)(NPR-P=0012/000
(a)
 Portugueses))
(b)
       (NP
               (ID
                                       (NPR-P=0151:0012/001
 Portugueses))
(c)
       (NP
               (ID
                                       (NPR-P=0509:0012/003
 Portugueses))
                                      (D-P os)(NPR-P=0548:0012/004
(d)
    (NP-SBJ (ID
 Portugueses))
       (NP
                                       (PRO eles))
(e)
                                      (CL -lhes)) ...
(f)
       (NP-DAT (ID
```

```
{
```

(f) (NP-DAT (ID

```
Cadeia ID 0012, iniciada por 'os Portugueses' (apenas ocorrências 000 a 005)
                           0012/000) (D-P os)(NPR-P=0012/000
(a) (NP-SBJ (ID
 Portugueses))
(b) (NP
               (ID
                           0012/001) (NPR-P=0151:0012/001
 Portugueses))
(c) (NP
               (ID
                           0012/002) (NPR-P=0509:0012/003
 Portugueses))
(d) (NP-SBJ (ID
                           0012/003) (D-P os)(NPR-P=0548:0012/004
 Portugueses))
       (NP
                          0012/004) (PRO eles))
(e)
               (ID
```

0012/005) (CL -lhes)) ...

```
{
```

```
(a) (NP-SBJ (ID
                     DNN=0012/000) (D-P os)(NPR-P=0012/000
 Portugueses))
(b) (NP
            (ID
                     NNN=0012/001) (NPR-P=0151:0012/001
 Portugueses))
(c) (NP
             (ID
                     NNN=0012/002) (NPR-P=0509:0012/003
 Portugueses))
(d) (NP-SBJ (ID
                     DNN=0012/003) (D-P os)(NPR-P=0548:0012/004
Portugueses))
      (NP (ID PPP=0012/004) (PRO eles))
(e)
(f) (NP-DAT (ID
                     PPP=0012/005) (CL -lhes)) ...
```

```
{
```

Portugueses))

(e)

(NP (ID EE-PPP=0012/004) (PRO eles))

(f) (NP-DAT (ID EE-PPP=0012/005) (CL -lhes)) ...

```
{
```

```
{
```

```
Cadeia ID 0012, iniciada por 'os Portugueses' (apenas ocorrências 000 a 005)
(a) (NP-SBJ (ID H-AA-DNN=0012/000) (D-P os)(NPR-P=0012/000)
 Portugueses))
(b) (NP (ID M-RR-NNN=0012/001) (NPR-P=0151:0012/001
 Portugueses))
(c) (NP (ID M-RR-NNN=0012/002) (NPR-P=0509:0012/003)
 Portugueses))
       (NP-SBJ (ID M-RR-DNN=0012/003) (D-P os)(NPR-P=0548:0012/004
(d)
 Portugueses))
(e) (NP (ID M-EE-PPP=0012/004) (PRO eles))
(f) (NP-DAT (ID M-EE-PPP=0012/005) (CL -lhes)) ...
                Anotação automatizada
```

```
{
```

```
Cadeia ID 0012, iniciada por 'os Portugueses' (apenas ocorrências 000 a 005)
(a) (NP-SBJ (ID H-AA-DNN=0012/000) (D-P os)(NPR-P=0012/000)
 Portugueses))
(b) (NP (ID M-RR-NNN=0012/001) (NPR-P=0151:0012/001
 Portugueses))
(c) (NP (ID M-RR-NNN=0012/002) (NPR-P=0509:0012/003
 Portugueses))
(d)
       (NP-SBJ (ID M-RR-DNN=0012/003) (D-P os)(NPR-P=0548:0012/004)
 Portugueses))
(e) (NP (ID M-EE-PPP=0012/004) (PRO eles))
(f) (NP-DAT (ID M-EE-PPP=0012/005) (CL -lhes)) ...
                Anotação automatizada
                por aplicação de
                expressões regulares
```

```
{
```

```
Cadeia ID 0012, iniciada por 'os Portugueses' (apenas ocorrências 000 a 005)

(a) (NP-SBJ (ID H-AA-DNN=0012/000) (D-P os)(NPR-P=0012/000
Portugueses))
(b) (NP (ID M-RR-NNN=0012/001) (NPR-P=0151:0012/001
Portugueses))
(c) (NP (ID M-RR-NNN=0012/002) (NPR-P=0509:0012/003
Portugueses))
(d) (NP-SBJ (ID M-RR-DNN=0012/003) (D-P os)(NPR-P=0548:0012/004 Portugueses))
(e) (NP (ID M-EE-PPP=0012/004) (PRO eles))
(f) (NP-DAT (ID M-EE-PPP=0012/005) (CL -lhes)) ...
```

# Anotação automatizada usando *Corpus Search*

```
{
```

```
Cadeia ID 0012, iniciada por 'os Portugueses' (apenas ocorrências 000 a 005)
(a) (NP-SBJ (ID H-AA-DNN=0012/000) (D-P os)(NPR-P=0012/000)
 Portugueses))
(b) (NP (ID M-RR-NNN=0012/001) (NPR-P=0151:0012/001
 Portugueses))
(c) (NP (ID M-RR-NNN=0012/002) (NPR-P=0509:0012/003)
 Portugueses))
(d)
       (NP-SBJ (ID M-RR-DNN=0012/003) (D-P os)(NPR-P=0548:0012/004)
 Portugueses))
(e) (NP (ID M-EE-PPP=0012/004) (PRO eles))
(f) (NP-DAT (ID M-EE-PPP=0012/005) (CL -lhes)) ...
                Anotação não-automatizada
```

```
Cadeia ID 0012, iniciada por 'os Portugueses' (apenas ocorrências 000 a 005)
     (NP-SBJ (ID H-AA-DNN=0012/000) (D-P os)(NPR-P=0012/000)
(a)
 Portugueses))
(b) (NP (ID M-RR-NNN=0012/001) (NPR-P=0151:0012/001
 Portugueses))
(c) (NP (ID M-RR-NNN=0012/002) (NPR-P=0509:0012/003
 Portugueses))
(d)
       (NP-SBJ (ID M-RR-DNN=0012/003) (D-P os)(NPR-P=0548:0012/004)
 Portugueses))
(e) (NP (ID M-EE-PPP=0012/004) (PRO eles))
(f) (NP-DAT (ID M-EE-PPP=0012/005) (CL -lhes)) ...
                Anotação não-automatizada:
```

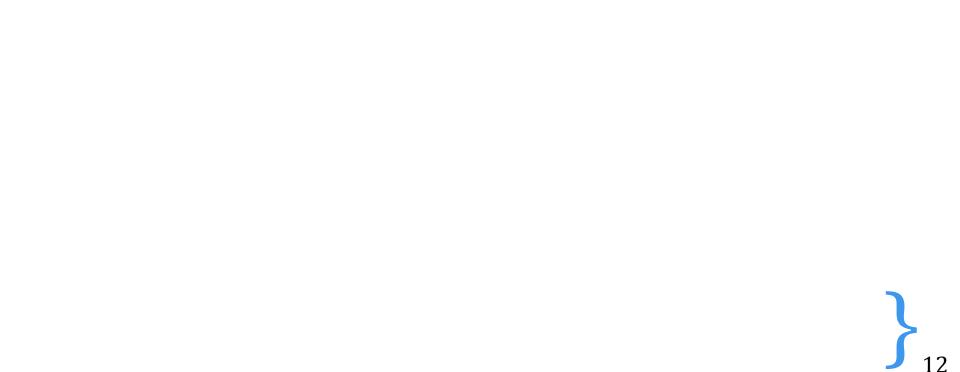
- identificação dos participantes da cadeia
- contagem dos elementos da cadeia

```
\left\{ 
ight.
```

```
Cadeia ID 0012, iniciada por 'os Portugueses' (apenas ocorrências 000 a 005)
     (NP-SBJ (ID H-AA-DNN=0012/000) (D-P os)(NPR-P=0012/000
(a)
 Portugueses))
(b) (NP (ID M-RR-NNN=0012/001) (NPR-P=0151:0012/001
 Portugueses))
(c) (NP (ID M-RR-NNN=0012/002) (NPR-P=0509:0012/003)
 Portugueses))
(d)
       (NP-SBJ (ID M-RR-DNN=0012/003) (D-P os)(NPR-P=0548:0012/004
 Portugueses))
       (NP (ID M-EE-PPP=0012/004) (PRO eles))
(e)
(f) (NP-DAT (ID M-EE-PPP=0012/005) (CL -lhes)) ...
                Anotação não-automatizada:
                - identificação dos participantes da cadeia
```

- contagem dos elementos da cadeia

```
{
```



**\** 

1# Mesmo nome, diferentes referentes

Exemplo (a): 'Índia'

```
{
```

```
ʻÍndia', NPR=0006
```

Ocorrências de 'Índia': 7

```
(NPR=0006/000 Índia)

(NPR=1187:0006/001 Índia)

(NPR=1295:0006/002 Índia)

(NPR=2599:0006/003 Índia)

(NPR=2627:0006/004 Índia)

(NPR=2640:0006/005 Índia)

(NPR=2694:0006/006 Índia)

(NPR=3599:0006/007 Índia)
```



```
\left\{ \right.
```

```
ʻÍndia', NPR=0006
```

Ocorrências de 'Índia': 7 Cadeias nucleadas por 'Índia': 4

```
(NPR=0006/000  (NPR=1187:0006/000  (NPR=1187:0006/001  (NPR=1295:0006/001  (NPR=1295:0006/002  (NPR=1295:0006/002  (NPR=2599:0006/003  (NPR=2599:0006/003  (NPR=2627:0006/004  (NPR=2627:0006/004  (NPR=2640:0006/005  (NPR=2640:0006/005  (NPR=2694:0006/006  (NPR=2694:0006/006  (NPR=3599:0006/007  (NPR=3590)  (NPR=3590)
```



```
\left\{ 
ight.
```

```
ʻÍndia', NPR=0006
```

Ocorrências de 'Índia': 7 Cadeias nucleadas por 'Índia': 4

```
(NPR=0006/000  findia) \rightarrow ID 0006/000  (NPR=1187:0006/001 findia) \rightarrow ID 6060/000  (NPR=1295:0006/002 findia) \rightarrow ID 0006/001  (NPR=2599:0006/003 findia) \rightarrow ID 6153/000  (NPR=2627:0006/004 findia) ID 6153/001  (NPR=2640:0006/005 findia) ID 6153/002  (NPR=2694:0006/006 findia) ID 6153/003  (NPR=3599:0006/007 findia) \rightarrow ID 6227
```



 $\left\{ 
ight.$ 

Contexto de ID 0006/000:

"REINANDO aquele muito católico e sereníssimo Príncipe el-Rei Dom MANUEL, fez-se uma frota para **a Índia** de que ia por capitão mór Pedro Álvares Cabral: que foi a segunda navegação que fizeram os Portugueses para aquelas partes do Oriente", G\_008,6,1.

```
Cadeia ID 0006, iniciada por 'a Índia'

(a) (NP (ID H-AA-DNN=0006/000) (D-F a) (NPR=0006/000 Índia)

(b) (NP (ID M-RR-DNN=0006/001) (D-F @a)(NPR=1295:0006/002 Índia)
```

Contexto de ID 0006/001:

"Umas árvores há também nestas partes muito altas a que chamam Sapucaias : nas quais se criam uns vasos tamanhos como grandes cocos , quase da feição de jarras d**a Índia**", G\_008,17.214.

```
ʻÍndia', NPR=0006
```

Ocorrências de 'Índia': 7 Cadeias nucleadas por 'Índia': 4

```
(NPR=0006/000  Índia) \rightarrow ID 0006/000 (NPR=1187:0006/001  Índia) \rightarrow ID 6060/000 (NPR=1295:0006/002  Índia) \rightarrow ID 0006/001 (NPR=2599:0006/003  Índia) \rightarrow ID 6153/000 (NPR=2627:0006/004  Índia) ID 6153/001 (NPR=2640:0006/005  Índia) ID 6153/002 (NPR=2694:0006/006  Índia) ID 6153/003 (NPR=3599:0006/007  Índia) \rightarrow ID 6227
```



Contextos da cadeia ID 6060:

"E depois de assim a terem curada desta maneira põem um alguidar sobre o fogo em que a lançam , a qual está mexendo **uma Índia** até que o mesmo fogo lhe acabe de gastar aquela umidade e fique enxuta e disposta para se poder comer , que será por espaço de meia hora pouco mais ou menos", ID  $G_{008,16.182}$ 

```
Cadeia ID 6060, composta por 'uma Índia'

(NP-SBJ (ID H-RR-UNN=6060.0006/000) (D-UM-F uma) (NPR=1187:0006/001 Índia))
```

```
ʻÍndia', NPR=0006
```

Ocorrências de 'Índia': 7 Cadeias nucleadas por 'Índia': 4



### Contextos da cadeia ID 6153:

"Na capitania de São Vicente, sendo já alta noite a horas em que todos começavam de se entregar ao sono, acertou de sair fora de casa uma Índia escrava do capitão: a qual lançando os olhos a uma várzea que está pegada com o mar, e com a povoação da mesma capitania, \_\_\_ viu andar nela este monstro, movendo-se de uma parte para outra, com passos e meneios desusados, e dando alguns urros de quando em quando tão feios, que \_\_\_ como pasmada e quase fora de si, se veio ao filho do mesmo capitão, cujo nome era Baltesar Ferreira", G\_008,29.538; "E obedecendo a Índia a seu mandado foi", G\_008,31.542; "e \_\_\_ tornou mais espantada, afirmando-lhe e repetindo-lhe uma vez e outra, que andava ali uma coisa tão feia , que não podia ser senão o demônio", G\_008,31.543; "Então se levantou ele muito depressa", "e lançou mão a uma espada que tinha junto de si, com a qual botou somente em camisa pela porta fora, tendo para si ( quando muito ) que seria algum Tigre, ou outro animal da terra conhecido com a vista do qual se desenganasse do que a Índia lhe queria persuadir", G\_008,31.544, G\_008,31.545; "E pondo os olhos naquela parte que **ela** lhe assinalou, viu confusamente o vulto do monstro ao longo da praia , sem poder divisar o que eram por causa da noite lho impedir e o monstro também ser coisa não vista, e fora do parecer de todos os outros animais", G\_008,31.546; "Neste tempo acodiram alguns escravos aos gritos da Índia que estava em vela", G 008,31.557.

Cadeia ID 6153, iniciada por 'uma Índia escrava do capitão'

```
(a)
           (NP-SBJ (ID H-RC-UNC=6153.0006/000)
                                                    (D-UM-F uma)(NPR=2599:0006/003 Índia)(NP-PRN escrava do capitão))
(b)
           (WNP
                   (ID M-EE-DPP=6153/001)
                                                    (D-F a) (WPRO qual))
(c)
           (NP-SBJ (ID M-EE-PPP=6153/002)
                                                    (*pro* *pro*))
(d)
           (NP
                   (ID M-EE-PPP=6153/003)
                                                    (PRO si))
(e)
           (NP-SE (ID M-EE-PPP=6153/004)
                                                    (CL se))
(f)
           (NP-SBJ (ID M-EE-PPP=6153/005)
                                                    (*pro* *pro*))
(g)
           (NP-SBJ (ID M-EE-PPP=6153/006)
                                                    (*pro* *pro*))
(h)
           (NP-DAT (ID M-EE-PPP=6153/007)
                                                    (CL -lhe))
(i)
           (NP-DAT (ID M-EE-PPP=6153/008)
                                                   (CL 1h@))
(j)
           (NP
                   (ID M-EE-PPP=6153/009)
                                                    (CL a))
(k)
           (NP-DAT (ID M-EE-PPP=6153/010)
                                                   (CL -lhe))
(1)
           (NP-SBJ (ID M-EE-PPP=6153/011)
                                                   (*pro* *pro*))
           (NP-SBJ (ID M-EE-PPP=6153/012)
(m)
                                                   (*pro* *pro*))
(n)
           (NP-SBJ (ID M-RR-DNN=6153/013)
                                                   (D-F a) (NPR=2627:0006/004 Índia))
(o)
           (NP-SBJ (ID M-EE-PPP=6153/014)
                                                   (*pro* *pro*))
(p)
           (NP-SBJ (ID M-RR-DNN=6153/015)
                                                   (D-F a) (NPR=2640:0006/005 Índia))
(q)
           (NP-SBJ (ID M-EE-PPP=6153/016)
                                                   (PRO ela))
(r)
           (NP
                   (ID M-RC-DNC=6153/017)
                                                   (D-F @a) (NPR=2694:0006/006 Índia (...) )
(s)
           (WNP-1
                   (ID M-EE-PPP=6153/018)
                                            (WPRO que))
```

```
ʻÍndia', NPR=0006
```

Ocorrências de 'Índia': 7 Cadeias nucleadas por 'Índia': 4



Contextos da cadeia de ID 6227:

"Está **uma Índia velha** prestes com um cabaço grande na mão", "e como ele cai, \_\_\_ acode muito depressa a meter-lho na cabeça para tomar nele os miolos e o sangue", G\_008,42.776 e G\_008,42.777.

```
Cadeias nucleadas por 'Índia': 4

(NPR=0006/000 Índia) → ID 0006/000

(NPR=1187:0006/001 Índia) → ID 6060/000

(NPR=1295:0006/002 Índia) → ID 0006/001

(NPR=2599:0006/003 Índia) → ID 6153/000

(NPR=2627:0006/004 Índia) ID 6153/001

(NPR=2640:0006/005 Índia) ID 6153/002

(NPR=2694:0006/006 Índia) ID 6153/003

(NPR=3599:0006/007 Índia) → ID 6227
```

ʻIndia', NPR=0006

Ocorrências de 'Índia': 7

```
Cadeias ID 0006, 6153, 6227 - cabeças:
```

```
(a) (NP (ID H-AA-DNN=0006/000) (D-F a) (NPR=0006/000 Índia) (b) (NP-SBJ (ID H-RR-UNN=6060.0006/000) (D-UM-F uma) (NPR=1187:0006/001 Índia)) (C) (NP-SBJ (ID H-RC-UNC=6153.0006/000) (D-UM-F uma) (NPR=2599:0006/003 Índia) (NP-PRN escrava do capitão)) (D-UM-F uma) (NPR=3599:0006/007 Índia) (ADJ-F velha))
```

Cadeias ID 0006, 6153, 6227 - cabeças:

```
(D-F a) (NPR=0006/000)
(a) (NP
        (ID H-AA-DNN=0006/000)
 Índia)
(b) (NP-SBJ (ID H-RR-UNN=6060.0006/000)
                                                          (D-UM-F uma) (NPR=1187:0006/001
 Índia))
(c) (NP-SBJ (ID H-RC-UNC=6153.0006/000)
                                                          (D-UM-F uma)(NPR=2599:0006/003
 Índia)
                                                (NP-PRN escrava do capitão))
(d) (NP-SBJ (ID H-RC-UNC=6227.0006/000)
                                                          (D-UM-F uma)(NPR=3599:0006/007
 Índia)
        (ADJ-F velha))
 Automático índices derivados dos NPR +
                      códigos de estrutura interna
```

Cadeias ID 0006, 6153, 6227 - cabeças:

```
(D-F a) (NPR=0006/000)
(a) (NP
        (ID H-AA-DNN=0006/000)
 Índia)
(b) (NP-SBJ (ID H-RR-UNN=6060.0006/000)
                                                        (D-UM-F uma) (NPR=1187:0006/001
 Índia))
(c) (NP-SBJ (ID H-RC-UNC=6153.0006/000)
                                                        (D-UM-F uma)(NPR=2599:0006/003
 Índia)
                                               (NP-PRN escrava do capitão))
(d) (NP-SBJ (ID H-RC-UNC=6227.0006/000)
                                                        (D-UM-F uma)(NPR=3599:0006/007
 Índia)
        (ADJ-F velha))
 Automático índices derivados dos NPR +
                     códigos de estrutura interna
  Manual
                     identificação das mudanças de referente
                     (e contagem no interior de cada cadeia)
```

1# Mesmo nome, diferentes referentes

```
Exemplo (b): 'Portugueses'
```

Contexto da cadeia de 8006, iniciada por 'alguns Portugueses':

"Aqui se metem dois rios nele que vem do sertão, por um dos quais entraram **alguns Portugueses** quando foi do descobrimento que \_\_\_ foram fazer no ano de 35 e \_\_\_ navegaram por ele acima duzentas e cinquenta léguas, até que \_\_\_ não puderam ir mais por diante por causa da água ser pouca e o rio se ir estreitando de maneira, que não podiam já por ele caber as embarcações", G\_008,9.47.

```
Cadeia ID 8006, iniciada por 'alguns Portugueses'

(NP-SBJ (ID H-RC-NNQ=8006.0012/000) (Q-P alguns) (NPR-P=0409:0012/002 Portugueses))

(NP-SBJ (ID M-EE-PPP=8006/001) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/002) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/003) (*pro* *pro*))
```

```
Cadeia ID 8006, iniciada por 'alguns Portugueses'

(NP-SBJ (ID H-RC-NNQ=8006.0012/000) (Q-P alguns) (NPR-P=0409:0012/002 Portugueses))

(NP-SBJ (ID M-EE-PPP=8006/001) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/002) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/003) (*pro* *pro*))
```

```
Cadeia ID 8006, iniciada por 'alguns Portugueses'

(NP-SBJ (ID H-RC-NNQ=8006.0012/000) (Q-P alguns) (NPR-P=0409:0012/002 Portugueses))

(NP-SBJ (ID M-EE-PPP=8006/001) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/002) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/003) (*pro* *pro*))
```

```
Cadeia ID 8006, iniciada por 'alguns Portugueses'

(NP-SBJ (ID H-RC-NNQ=8006.0012/000) (Q-P alguns) (NPR-P=0409:0012/002 Portugueses))

(NP-SBJ (ID M-EE-PPP=8006/001) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/002) (*pro* *pro*))

(NP-SBJ (ID M-EE-PPP=8006/003) (*pro* *pro*))
```

Outros casos de 'cabeças de cadeia' iniciados por sintagmas com demonstrativos

(a) (ID H-AC-ENC=5000.0001/000) aquele muito católico e sereníssimo Príncipe el-Rei Dom MANUEL (b) (ID H-AC-ENC=5100.0089/000) Aquele grande e inefável mistério do Santíssimo Sacramento (c) (ID H-AC-ENC=5095.0200/000) aquela grande América, uma das quatro partes do mundo (d) (ID H-AC-ENC=5142.0317/000) aquela temperança da primavera que cá nos oferece Abril e Maio (e) (ID H-AC-ENC=5614.2559/000) aquelas grandes minas que promete a mesma terra (f) (ID H-AC-ENC=5322.2576/000) aquele fero e espantoso monstro marinho que nesta província se matou no ano de 1564

2# Nomes diferentes, mesmo referente

```
Exemplo (a):
'bálsamo'
'óleo'
'licor'
```

### A anotação: alguns casos interessantes Contextos da cadeia ID 5200:

"Um certo gênero de árvores há também pelo mato dentro na capitania de Paranambuco a que chamam Copahíbas de que se tira bálsamo muito salutífero e proveitoso em extremo para enfermidades de muitas maneiras, principalmente nas que procedem de frialdade \_\_\_ causa grandes efeitos e \_\_\_ tira todas as dores por graves que sejam em muito breve espaço. Para feridas ou quaisquer outras chagas, \_\_\_ tem a mesma virtude: as quais tanto que com **ele** lhe acodem, saram muito depressa, e \_\_\_ tira os sinais de maneira, que de maravilha se enxerga onde estiveram, e nisto faz vantagem a todas as outras medicinas. Este óleo não se acha todo ano perfeitamente nestas árvores, nem procuram ir buscá-lo, senão no estio, que é o tempo em que assinaladamente o criam. E quando querem tirá-lo, dão certos golpes ou furos no tronco delas, pelos quais pouco a pouco estão estilando do âmago este licor precioso..."

```
Cadeia do referente ID 5200 (iniciado por 'bálsamo muito salutífero...')
                                              (N=1439/000 bálsamo) muito
(a)
        (NP-SBJ (ID H-AC-NNC=5200.1439/000)
 salutífero...)
                                              (*pro* *pro*))
(b)
        (NP-SBJ (ID M-EE-PPP=5200/001)
                                              (*pro* *pro*))
(c)
        (NP-SBJ (ID M-EE-PPP=5200/002)
(d)
        (NP-SBJ (ID M-EE-PPP=5200/003)
                                              (*pro* *pro*))
(e)
        (NP (ID M-EE-PPP=5200/004)
                                          (PRO ele))
        (NP-SBJ (ID M-EE-PPP=5200/005) (*pro* *pro*))
(f)
        (NP-SBJ (ID M-EE-PPP=5200/006) (*pro* *pro*))
(g)
        (NP-SBJ (ID M-AA-ENN=5200/007) (D Este) (N=1245/000 óleo))
(h)
(i)
        (NP-ACC (ID M-EE-PPP=5200/008) (CL -lo))
        (NP-ACC (ID M-EE-PPP=5200/009) (CL o))
(j)
(k)
        (NP-ACC (ID M-EE-PPP=5200/010) (CL -lo))
(1)
        (NP-ACC (ID M-AC-ENC=5200/011) (D este) (N=1463/000 licor) precioso )
```

```
Cadeia do referente ID 5200 (iniciado por 'bálsamo muito salutífero...')
        (NP-SBJ (ID H-AC-NNC=5200.1439/000)
                                              (N=1439/000 bálsamo) muito
(a)
 salutífero...)
        (NP-SBJ (ID M-EE-PPP=5200/001)
                                               (*pro* *pro*))
(b)
                                              (*pro* *pro*))
(c)
        (NP-SBJ (ID M-EE-PPP=5200/002)
(d)
        (NP-SBJ (ID M-EE-PPP=5200/003)
                                              (*pro* *pro*))
(e)
        (NP (ID M-EE-PPP=5200/004)
                                              (PRO ele))
(f)
        (NP-SBJ (ID M-EE-PPP=5200/005) (*pro* *pro*))
(g)
        (NP-SBJ (ID M-EE-PPP=5200/006) (*pro* *pro*))
(h)
        (NP-SBJ (ID M-AA-ENN=5200/007) (D Este) (N=1245/000 óleo))
(i)
        (NP-ACC (ID M-EE-PPP=5200/008) (CL -lo))
        (NP-ACC (ID M-EE-PPP=5200/009) (CL o))
(j)
(k)
        (NP-ACC (ID M-EE-PPP=5200/010) (CL -lo))
(1)
        (NP-ACC (ID M-AC-ENC=5200/011) (D este) (N=1463/000 licor) precioso )
```

```
Cadeia do referente ID 5200 (iniciado por 'bálsamo muito salutífero...')
        (NP-SBJ (ID H-AC-NNC=5200.1439/000)
                                               (N=1439/000 bálsamo) muito salutífero...
(a)
                                               (*pro* *pro*))
(b)
        (NP-SBJ (ID M-EE-PPP=5200/001)
(c)
                                               (*pro* *pro*))
        (NP-SBJ (ID M-EE-PPP=5200/002)
(d)
        (NP-SBJ (ID M-EE-PPP=5200/003)
                                              (*pro* *pro*))
(e)
        (NP (ID M-EE-PPP=5200/004)
                                              (PRO ele))
        (NP-SBJ (ID M-EE-PPP=5200/005) (*pro* *pro*))
(f)
(g)
        (NP-SBJ (ID M-EE-PPP=5200/006) (*pro* *pro*))
(h)
        (NP-SBJ (ID M-AA-ENN=5200#1245~1439/007) (D Este) (N=1245/000 óleo))
        (NP-ACC (ID M-EE-PPP=5200/008) (CL -lo))
(i)
(j)
        (NP-ACC (ID M-EE-PPP=5200/009) (CL o))
(k)
        (NP-ACC (ID M-EE-PPP=5200/010) (CL -lo))
(1)
        (NP-ACC (ID M-AC-ENC=5200#1463~1439/011) (D este) (N=1463/000 licor) precioso )
```

Detalhe da anotação completa, incluindo indicadores suplementares de sinonímia:

3# Sintagmas nominais... sem nomes

```
Exemplo (a):
  'a do Sul'
  'a do Oriente'
  'a do Ocidente'
```

Contextos das cadeias ID 5099, 5005, 5508:

"Está formada esta província à maneira de uma harpa : cuja costa pela banda do Norte corre do Oriente ao Ocidente e está olhando direitamente a Equinocial . E pela do Sul confina com outras províncias da mesma América povoadas e possuídas de povo gentílico com que ainda não temos comunicação. E pela do Oriente confina com o mar Oceano Áfrico, e olha direitamente os Reinos de Congo e Angola até o cabo de Boa Esperança que é o seu oposto. E pela do Ocidente confina com as altíssimas serras dos Andes e fraldas do Peru , as quais são tão soberbas em cima da terra , que se diz terem as aves trabalho em as passar", G\_008,8.22

(ID ID.0000.A-BB-CCC=0000/000)

(ID ID.0000.A-BB-CCC=0000/000)

ID.0000

Identificador único

#### (ID ID.0000.A-BB-CCC=0000/000)

ID.0000 Identificador único

Classificadores mnemônicos:

A Cabeça de cadeia ou menção posterior
BB Estatuto do nome contido no NP

CCC Estrutura interna

### (ID ID.0000.A-BB-CCC=0000/000)

ID.0000 Identificador único

Classificadores mnemônicos:

A Cabeça de cadeia ou menção posterior

BB Estatuto do nome contido no NP
CCC Estrutura interna

Estrutura interna

Índices de referência:

0000 Índice do referente

17

```
Classificadores básicos da estrutura interna
(com as etiquetas morfossintáticas correspondentes (*):
N - Contém um nome
                                           (N; NPR)
D - Contém um determinante definido
                                           (D)
U - Contém um determinante indefinido
                                          (D-UM)
E - Contém um demonstrativo
                                          (DEM; D(**))
Q - Contém um quantificador
                                           (Q; NUM)
0 - Contém 'outro(s)/a(s)'
                                           (OUTRO)
C - Contém um complemento ou modificador (PP; CP-REL; ADJ/ADJ-G/ADJ-S)
S - Contém um pronome possessivo
                                           (PRO$)
P - Contém um pronome pessoal ou relativo (PRO, *pro*, CL, SE, WPRO)
```

#### Exemplos

```
Exemplos
```

Note-se que os classificadores BBB são **redundantes** frente à anotação da estrutura interna dos NPs

```
Exemplos
```

Os classificadores BB são **redundantes** frente à combinação entre os índices de referente e os índices de cada N ou NPR

```
Exemplos
```

Os classificadores A são **redundantes** frente à combinação entre os índices de referente e o número da ocorrência

#### Exemplos

Ou seja, os classificadores mnemônicos A-BB-CCC são inteiramente dispensáveis na anotação.

Seu objetivo é apenas facilitar as buscas caso não se deseje usar o Corpus Search.

```
Exemplos
```

```
(ID
             0012/001)
                                      (NPR-P
                 Portugueses)
             0012/000)
                                      (D-P os)(NPR-P
(ID
       Portugueses)
                                                                             Índia))
             6060.0006/000)
                                      (D-UM-F uma) (NPR
(ID
                                      (D-UM-F uma)(NPR
                                                                             Índia)(ADJ-F
             6227.0006/000)
(ID
 velha)
             0012/004)
                                      (PRO eles)
(ID
```

Um sistema final mais limpo poderia ser composto **exclusivamente dos índices de referentes** (no limite, excluíndo-se também os índices dos NP – NPRs).

- (i) Indexação das instâncias individuais de 'nomes': N, N-P, NPR, NPR-P (=xxxx)
- (ii) Adição do índice da primeira ocorrência a casos de nomes idênticos subsequentes (yyyy:xxxx)
- (iii) Numeração de cada ocorrência de palavras idênticas (/zzz)
- (iv) Identificação de NPs com referentes inéditos;
- (v) Indexação dos referentes inéditos, derivada dos nomes(=xxxx) ou aleatória (=wwww.xxxx).
- (vi) Identificação de NPs com referentes repetidos;
- (vii) Co-indexação dos referentes repetidos com o índice da cadeia (=xxxx ou =wwww)
- (viii) Numeração de cada sintagma em uma cadeia (/zzz)
- (ix) Aplicação do código de letras inicial para estatuto do referente (H-/M-)
- (x) Aplicação do índice medial para estatuto do nome (-AA-/-AC-/-RR-/-RC-/-EE-/-EC-/-BB-)
- (xi) Aplicação do código de letras final para estrutura interna (-NNN/-NNC/-DNN/-DNC/...)

- (i) Indexação das instâncias individuais de 'nomes': N, N-P, NPR, NPR-P (=xxxx)
   (ii) Adição do índice da primeira ocorrência a casos de nomes idênticos subsequentes (yyyy:xxxx)
- (iii) Numeração de cada ocorrência de palavras idênticas (/zzz)
- (iv) Identificação de NPs com referentes inéditos;
- (v) Indexação dos referentes inéditos, derivada dos nomes(=xxxx) ou aleatória (=wwww.xxxx).
- (vi) Identificação de NPs com referentes repetidos;
- (vii) Co-indexação dos referentes repetidos com o índice da cadeia (=xxxx ou =wwww)
- (viii) Numeração de cada sintagma em uma cadeia (/zzz)
- (ix) Aplicação do código de letras inicial para estatuto do referente (H-/M-)
- (x) Aplicação do índice medial para estatuto do nome (-AA-/-AC-/-RR-/-RC-/-EE-/-EC-/-BB-)
- (xi) Aplicação do código de letras final para estrutura interna (-NNN/-NNC/-DNN/-DNC/...)

Etapas preparatórias, automatizadas

# A anotação (etapas resumidas)

- (i) Indexação das instâncias individuais de 'nomes': N, N-P, NPR, NPR-P (=xxxx)
- (ii) Adição do índice da primeira ocorrência a casos de nomes idênticos subsequentes (yyyy:xxxx)
- (iii) Numeração de cada ocorrência de palavras idênticas (/zzz)
- (iv) Identificação de NPs com referentes inéditos;
- (v) Indexação dos referentes inéditos, derivada dos nomes(=xxxx) ou aleatória (=wwww.xxxx).
- (vi) Identificação de NPs com referentes repetidos;
- (vii) Co-indexação dos referentes repetidos com o índice da cadeia (=xxxx ou =wwww)
- (viii) Numeração de cada sintagma em uma cadeia (/zzz)
- (ix) Aplicação do código de letras inicial para estatuto do referente (H-/M-)
- (x) Aplicação do índice medial para estatuto do nome (-AA-/-AC-/-RR-/-RC-/-EE-/-EC-/-BB-)
- (xi) Aplicação do código de letras final para estrutura interna (-NNN/-NNC/-DNN/-DNC/...)

#### Etapas automatizáveis?

#### A anotação: em resumo

- (i) Indexação das instâncias individuais de 'nomes': N, N-P, NPR, NPR-P (=xxxx)
- (ii) Adição do índice da primeira ocorrência a casos de nomes idênticos subsequentes (yyyy:xxxx)
- (iii) Numeração de cada ocorrência de palavras idênticas (/zzz)
- (iv) Identificação de NPs com referentes inéditos;
- (v) Indexação dos referentes inéditos, derivada dos nomes(=xxxx) ou aleatória (=wwww.xxxx).
- (vi) Identificação de NPs com referentes repetidos;
- (vii) Co-indexação dos referentes repetidos com o índice da cadeia (=xxxx ou =wwww)
- (viii) Numeração de cada sintagma em uma cadeia (/zzz)
- (ix) Aplicação do código de letras inicial para estatuto do referente (H-/M-)
- (x) Aplicação do índice medial para estatuto do nome (-AA-/-AC-/-RR-/-RC-/-EE-/-EC-/-BB-)
- (xi) Aplicação do código de letras final para estrutura interna (-NNN/-NNC/-DNN/-DNC/...)

#### Etapas talvez dispensáveis

#### A anotação: em resumo

- (i) Indexação das instâncias individuais de 'nomes': N, N-P, NPR, NPR-P (=xxxx)
- (ii) Adição do índice da primeira ocorrência a casos de nomes idênticos subsequentes (yyyy:xxxx)
- (iii) Numeração de cada ocorrência de palavras idênticas (/zzz)
- (iv) Identificação de NPs com referentes inéditos;
- (v) Indexação dos referentes inéditos, derivada dos nomes(=xxxx) ou aleatória (=wwww.xxxx).
- (vi) Identificação de NPs com referentes repetidos;
- (vii) Co-indexação dos referentes repetidos com o índice da cadeia (=xxxx ou =wwww)
- (viii) Numeração de cada sintagma em uma cadeia (/zzz)
- (ix) Aplicação do código de letras inicial para estatuto do referente (H-/M-)
- (x) Aplicação do índice medial para estatuto do nome (-AA-/-AC-/-RR-/-RC-/-EE-/-EC-/-BB-)
- (xi) Aplicação do código de letras final para estrutura interna (-NNN/-NNC/-DNN/-DNC/...)

Etapas centrais e não-automatizáveis



# Perspectivas

```
Perspectivas
(algumas perguntas prementes)
```



#### Perspectivas (algumas perguntas prementes)

- 1. Os padrões de relação entre a expressão argumental e a estrutura informacional se repetirão em textos portugueses coetâneos?
- 1. Os padrões de relação entre a expressão argumental e a estrutura informacional se repetirão em textos brasileiros coetâneos?



#### Perspectivas (algumas perguntas prementes)

- 3. Será possível analisar, com esta metodologia, a relação entre os padrões de expressão argumental, a estrutura informacional, **e a estrutura dos documentos**?
- 4. Qual os impactos dessa análise para o nosso objetivo de formar uma agenda para os estudos sintáticos nas *Vésperas Brasilianas*?



#### Perspectivas (alguns caminhos para as respostas)



## Perspectivas (alguns caminhos para as respostas, no longo prazo)

Consolidação do experimento inicial e aprofundamento da análise:

- Replicação da anotação de cadeias de referentes em textos portugueses coetâneos e em textos brasileiros coetâneos;
- Aprofundamento das análises a partir dos novos textos anotados com a metodologia consolidada



### Perspectivas (algundação directo presimento inscraspostas, no curto prazo)

- Análise preliminar comparada entre o texto português quinhentista já estudado e um texto português coetâneo, a partir dos caminhos já possibilitados pelo experimento até este ponto.
- Análise preliminar comparada entre o texto português quinhentista já estudado e um texto brasileiro seiscentista, a partir dos caminhos já possibilitados pelo experimento até este ponto.



#### Perspectivas

(alguns caminhos para as respostas, no curto prazo) - Esta ampliação do experimento inicial na forma de análises

- Esta ampliação do experimento inicial na forma de análises comparadas preliminares pode ter como consequência, para o sub-projeto *Vésperas Brasilianas*, a consolidação de um dos aspectos sintáticos relevantes para a análise dos textos brasileiros dos primeiros séculos — justamente, os padrões de co-referência dos sujeitos nulos nos textos. Este pode ser um dos pontos da nossa *agenda*.



A única conclusão interessante até agora não parece depender do estatuto referencial dos sujeitos quanto a sua menção anterior no texto



## Primeiros experimentos (visão geral da anotação de cadeias)

```
(IP-MAT (IP-GER (VB-G REINANDO)
                        (NP-SBJ
                                 (ID ID.0001.M-AC-ENC=5000.0001/000)
                                   (D aquele)
                                   (ADJP (Q muito)
                                            (ADJ católico)
                                             (CONJP (CONJ e)
                                                    (ADJX (ADJ-S sereníssimo))))
                                   (NPR=0001/000 Principe)
                                   (NP-PRN (ID ID.0002.H-BB-NOM=9000/000) (NPR (ID ID.NPR.0001.H-BB-
NOM=9000/000) (NPR=0002/000 elRei) (NPR=0003/000 Dom) (NPR=0004/000 MANUEL)))))
             (,,)
             (VB-D fez-)
             (NP-SE-4 (ID COSE) (CL -se))
             (NP-SBJ-4 (ID ID.0003.H-AC-UNC=6000.0005/000)
                          (D-UM-F uma)
                          (N=0005/000 frota)
                          (PP (P para)
                                 (NP (ID ID.0004.H-AA-DNN=0006/000) (D-F a) (NPR=0006/000 Índia)))
                          (CP-REL (WPP-2 (P de)
                                               (NP (ID ID.0005.M-EE-PPP=6000/001) (WPRO que)))
                                     (IP-SUB (VB-D ia)
                                                (PP (P por)
                                                           (ID ID.0006.H-AC-NNC=5002.0007/000)
                                                            (N=0007/000 capitão)
F a)
                                                            (ADJ mór)
                                   (NPR=0006/0/0006 Índia)(NP-GEN *T*-2)))
```

(ND, CD)  $(TD, TD, QQQ7, II, DD, NQM_QQQ1, QQQ1, QQ1, QQQ1, QQQ$ 

#### Primeiros experimentos (visão geral da anotação de cadeias)

```
( (IP-MAT
           (NP-SBJ (ID 5706.0282/000)
                                  (D-FA)
                                  (N=0282/000 viração)
                                  (PP
                                             (P d@)
                                             (NP (ID 5066/007) (D-P @estes) (N-P=0283:0272/001
ventos))))
                                 (VB-P entra)
                                  (PP (P a@)
                                  (NP (ID 5127.0041/000) (D @o) (ADJ meio) (N=0284:0041/005 dia)))
                                  (ADVP (Q pouco)
                                  (ADV-R mais)
                                  (CONJP (CONJ ou)
                                  (ADVX (ADV-R menos))))
             (,,)
  (ID G 008,8.28))
```

## Primeiros experimentos (visão geral da anotação de cadeias)

```
((IP-MAT
           (NP-SBJ (ID ID.0429.H-AC-DNC=5706.0282/000)
                                 (D-FA)
                                  (N=0282/000 viração)
                                  (PP
                                             (P d@)
                                             (NP (ID ID.0430.M-RR-ENN=5066/007) (D-P @estes) (N-
P=0283:0272/001 ventos))))
                                 (VB-P entra)
                                  (PP (P a@)
                                  (NP (ID ID.0431.H-RC-DNC=5127.0041/000) (D @o) (ADJ meio)
(N=0284:0041/005 dia)))
                                  (ADVP (Q pouco)
                                  (ADV-R mais)
                                  (CONJP (CONJ ou)
                                  (ADVX (ADV-R menos))))
  (ID G 008,8.28))
```

