Investigar a literatura lusófona através dos tempos usando a Literateca (/pt/licoes/investigar-literatura-lusofona-literateca)

Diana Santos (https://orcid.org/0000-0002-3108-7706)

Esta lição ensina a utilizar o projeto Acesso a corpos / Disponibilização de corpos (AC/DC) para analisar textos literários em português, apresentando os resultados da pesquisa através de vários tipos de visualização produzidos com a linguagem R.

Avaliada por pares (https://github.com/programminghistorian/ph-submissions/issues/599)

editado por

- Eric Brasil (https://orcid.org/0000-0001-5067-8475)

revisto por

- Suemi Higuchi
- Larissa Freitas

publicado

| 2025-05-07
modificado

| 2025-05-02
dificuldade

| Baixo
https://doi.org/10.46430/phpt0054

Conteúdos🔗

- Introdução
- Apresentação do AC/DC
    - A sintaxe de procura
    - O corpo Literateca
    - Outras formas de pesquisa
- O uso da linguagem R
    - Roupa na literatura
    - Diferenças entre as personagens femininas e masculinas
    - A localização na literatura portuguesa
    - O helenismo na literatura brasileira
- Observações finais
- Notas de fim

Introdução🔗

Esta lição ensina a utilizar o projeto Acesso a corpos / Disponibilização de corpos (https://www.linguateca.pt/ACDC/) (AC/DC), mais especificamente a Literateca, para analisar textos literários em português. Usando a Literateca, é possível estudar, por exemplo, diferenças entre autores, escolas, e géneros literários ao longo do tempo. Além disso, ensina a apresentar os resultados da pesquisa por meio de vários tipos de visualização utilizando a linguagem R.

Para seguir a lição, tem de saber o que são folhas de registo (em inglês, "dataframes") em R e estar familiarizado com as formas de produzir gráficos de barras ("bar plots") e gráficos de caixa ("boxplots") no R. Além de consultar as Noções básicas de R com dados tabulares (/pt/licoes/nocoes-basicas-R-dados-tabulares), também pode seguir a lição Visualização básica de dados tabulares com R (/pt/licoes/visualizacao-basica-dados-tabulares-r).

Os casos específicos que servirão de exemplo nesta lição são os seguintes:

- A roupa na literatura lusófona
- Diferenças entre a caracterização de personagens femininas e masculinas ao longo do tempo
- A localização na literatura portuguesa
- O helenismo na literatura brasileira

Após concluir esta lição, estará:

- Familiarizado com o AC/DC para estudos literários
- Mais familiarizado com as ferramentas básicas de visualização do R

Apresentação do AC/DC🔗

O AC/DC é um projeto, já antigo, cujo objetivo é tornar disponíveis corpos para o português. Nesta lição vamos utilizar apenas o corpo Literateca (https://www.linguateca.pt/acesso/corpus.php?corpus=LITERATECA), que contém mais de 900 obras escritas por mais de 280 escritores de língua portuguesa (o AC/DC faz parte de um projeto maior, a Linguateca (https://perma.cc/7D24-UMGW)).

Um corpo é um conjunto de textos (neste caso, obras literárias) compilado com um objetivo específico (neste caso, o estudo da língua na literatura em português) e

classificado (a que, geralmente, se chama metadadados). Além disso, os corpos do AC/DC são anotados pelo analisador PALAVRAS (https://perma.cc/SGX6-GEX5)[1] e enriquecidos com anotação semântica adicional, como descrito em Santos (2014).[2] Em Santos (2021) também são apresentados vários exemplos de uso do AC/DC.[3]

A procura nos corpos é feita usando o sistema Open CWB (https://perma.cc/A295-JQ5G),[4] que permite gerir e interrogar grandes corpos anotados (contendo até dois biliões de palavras).

A sintaxe de procura🔗

Ao criar um corpo, define-se um conjunto de atributos para cada unidade – seja palavra, número ou sinal de pontuação. No processo de anotação, preenchem-se os valores desses atributos, que depois servirão como critérios de procura.

O primeiro atributo é a própria unidade. De seguida, há os atributos morfossintáticos, como o lema, a categoria gramatical e o género morfológico. Também existem atributos semânticos, como o campo semântico e o grupo ao qual a unidade pertence. Por último, há os atributos extralinguísticos, que incluem informações como o autor, o seu sexo, a data de publicação e a variedade do português.

Isto é exemplificado na tabela seguinte:

word	**lema**	**pos**	**temcagr**
Estou	estar	V	PR_IND
sem	sem	PRP	0
pilhas	pilha	N	0
!	!	PU	0

Tabela 1: Esta tabela contém um exemplo do texto “Estou sem pilhas!” no formato AC/DC, com quatro atributos: word, lema, pos, temcagr (este último indica, conforme a categoria gramatical, o tempo verbal, o caso pronominal ou o grau adjetival). A Literateca contém 27 atributos.

Um sistema de interrogação de corpos tem dois modos:

- A identificação do que se procura em contexto (que no AC/DC é, em geral, uma frase), a que se dá o nome de concordância
- A distribuição quantitativa dos resultados do que se procurou

Assim, para a mesma procura é possível escolher como resultado uma concordância ou a distribuição segundo um dos muitos atributos.

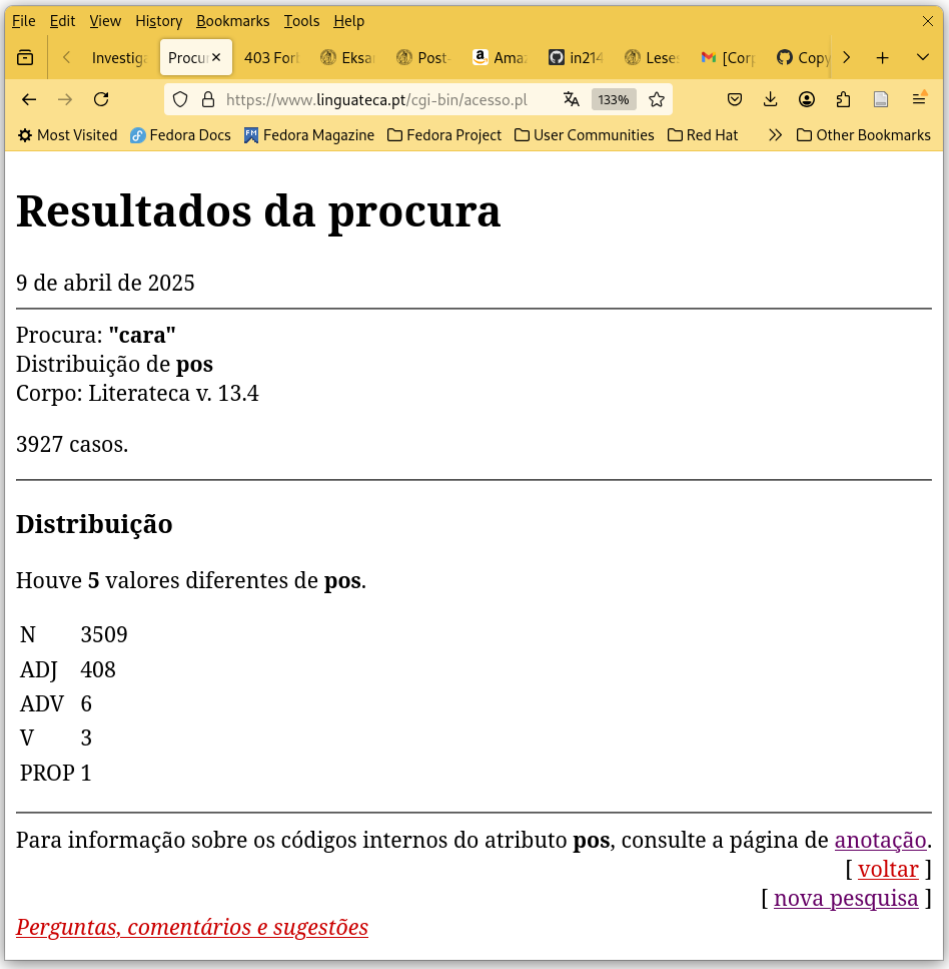
par=MdA-Dívida_extinta-299: É muito **cara** ?

par=MdA-D._Mônica-170: Facilmente se crê que Gaspar não saísse dali com a **cara** alegre .

par=MdA-D._Mônica-306: D. Mônica louvou os sentimentos do sobrinho e prometeu fazer por ele tudo o que fosse possível fazer por... por um neto, é o que ela devia dizer: mas ficou na vaga expressão -- por uma pessoa **cara** .

par=MdA-D._Mônica-335: -- Que **cara** é essa tão espantada ?

Figura 1. Pesquisa pela palavra ‘cara’, pedindo uma concordância (a figura exibe apenas 4 dos 4889 resultados da pesquisa.)



Resultados da procura

9 de abril de 2025

Procura: **"cara"**
Distribuição de **pos**
Corpo: Literateca v. 13.4

3927 casos.

Distribuição

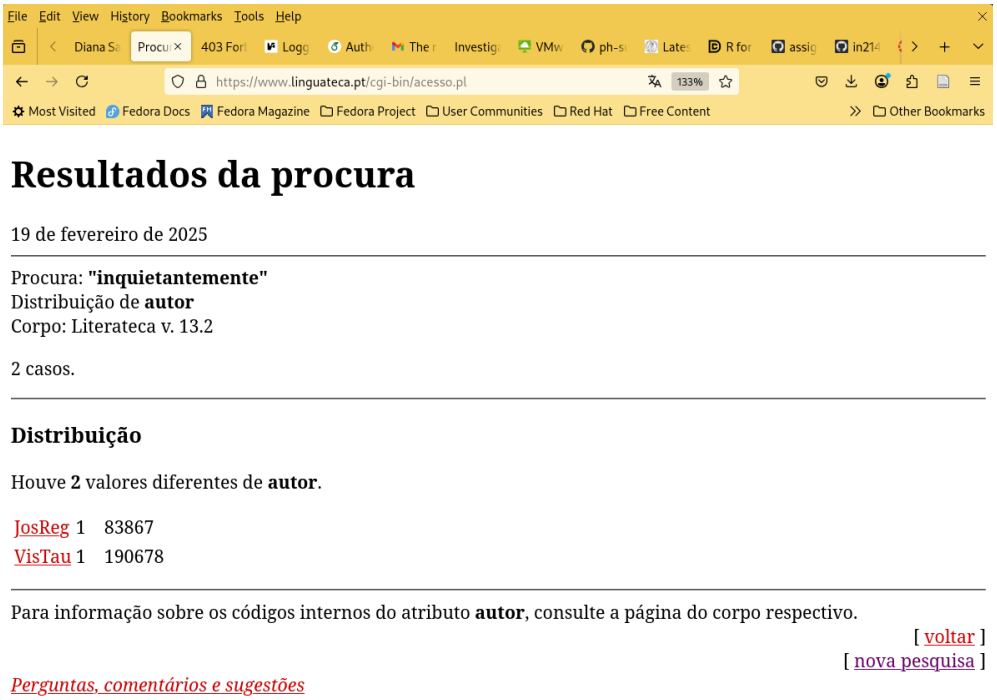
Houve **5** valores diferentes de **pos**.

N	3509
ADJ	408
ADV	6
V	3
PROP	1

Para informação sobre os códigos internos do atributo **pos**, consulte a página de anotação.
[voltar]
[nova pesquisa]

Perguntas, comentários e sugestões

Figura 2. Pesquisa exibindo a distribuição por categoria gramatical da palavra ‘cara’.



Resultados da procura

19 de fevereiro de 2025

Procura: **"inquietantemente"**
Distribuição de **autor**
Corpo: Literateca v. 13.2

2 casos.

Distribuição

Houve **2** valores diferentes de **autor**.

JosReg	1	83867
VisTau	1	190678

Para informação sobre os códigos internos do atributo **autor**, consulte a página do corpo respectivo.
[voltar]
[nova pesquisa]

Perguntas, comentários e sugestões

Figura 3. Pesquisa exibindo uma distribuição por autor da palavra ‘inquietantemente’.

Assim, o AC/DC permite fazer buscas nos textos. Tanto permite identificar em

contexto o resultado (para leitura próxima) como produzir um resumo quantitativo (a chamada leitura distante).

A sintaxe da procura é muito mais poderosa – além de buscar por palavras, permite pesquisar em todos os atributos e utilizar expressões regulares, quer nos valores dos atributos, quer sobre as próprias unidades.

Alguns exemplos:

- `[lema=".*ver" & pos="V.*"]` procura casos de verbos que terminem em *ver*
- `[pos="N"] [pos="ADJ.*" & word="re.*"]` procura casos de substantivos imediatamente seguidos de adjetivos iniciados por *re*
- `[lema="de"] [pos="DET.*"] [pos="N" & pessnum="S"]` procura casos da preposição *de* seguida por um determinante e por um ou mais adjetivos e um substantivo no singular
- `[lema="gostar"] [pos!="[NV].*"]* [func="<PIV"] [func=">N"]* @[func="P<"]` procura casos do verbo *gostar* até obter o núcleo do seu objeto de preposição

Para uma descrição mais completa da sintaxe do AC/DC, ver o texto Santos (2012),[5] assim como os exemplos (https://perma.cc/TFY4-EFNN) e as perguntas já respondidas (https://perma.cc/X5EW-FYVF) no website do AC/DC.

O corpo Literateca🔗

Enquanto a informação morfossintática e semântica é a mesma para todos os corpos do AC/DC, cada corpo contém uma informação extralinguística própria, que depende da informação que foi possível – e pertinente – obter sobre cada texto.

Para a Literateca, temos os seguintes atributos, identificados pelo pedido de distribuição na Figura 4.

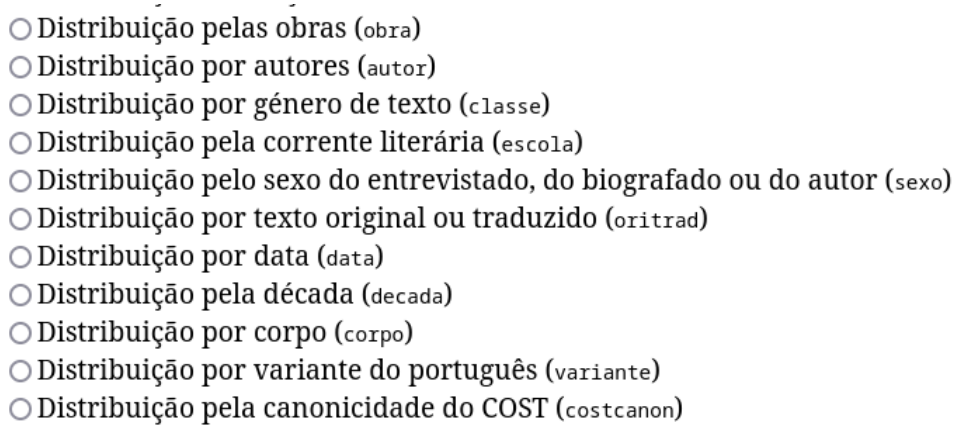


Figura 4. Distribuições de atributos extralinguísticos possíveis na Literateca.

O género de texto (atributo classe) está dividido entre Teatro, Prosa e Poesia. No caso da Prosa, pode assumir um dos seguintes valores: romance, novela, contos (livro de contos), conto, ensaio, cronica, historia, viagens, memorias, sermao, narrativaBiblica, autobiografia e cartas.

Para saber que obras ou autores foram incluídos, assim como a forma de os procurar, consulte a página lista de autores (https://perma.cc/5MJM-4SYE).

Outras formas de pesquisa🔗

Além da interface de pesquisa direta, existem outras formas de pedir informação de distribuição ao AC/DC, nomeadamente, através do Comparador (https://www.linguateca.pt/comparador/) e do Distribuidor (https://www.linguateca.pt/distribuidor/).

Enquanto o Comparador permite comparar duas distribuições com um único comando, o Distribuidor produz os resultados numa tabela que é facilmente utilizada em R.

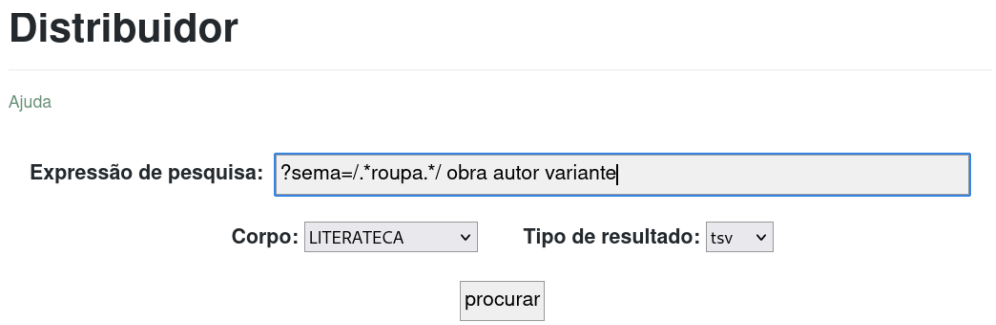


Figura 5. Interface de pesquisa no Distribuidor.

Em primeiro lugar, é preciso escolher o corpo que se quer pesquisar. Neste caso, a Literateca.

Se, por exemplo, quisermos saber a quantidade de menções a roupa distribuídas pelas obras, autores e variante, basta pedir:

```
?sema=/.*roupa.*/ obra autor variante
```

e, escolhendo a opção `tsv` (do inglês, "tab-separated values") para **Tipo de resultado**, obtém-se um ficheiro com dados tabulares que pode ser lido depois diretamente pelo R. Chamamos-lhe distribuicaoRoupa.tsv (/assets/investigar-literatura-lusofona-literateca/distribuicaoRoupa.tsv). Para mais informação sobre a exploração do vestuário na literatura em português, consulte o artigo Santos (2021).[6]

Mostro aqui o princípio desse ficheiro:

Os_Maias	1341	EcaQue	PT	1341	100.00
O_Primo_Basílio	942	EcaQue	PT	942	100.00
Gomes_Freire	866	RocMar	PT	866	100.00
A_Capital	728	EcaQue	PT	728	100.00
A_Relíquia	703	EcaQue	PT	703	100.00
Peregrinação	677	FerMPin	PT	677	100.00
A_Tragédia_da_Rua_das_Flores	665	EcaQue	PT	665	100.00
O_Crime_do_Padre_Amaro	663	EcaQue	PT	663	100.00
A_semana	533	MacAss	BR	533	100.00

A primeira coluna indica o nome da obra; a segunda o número de vezes que uma palavra marcada como sendo do campo semântico de roupa foi encontrada nessa obra; a terceira coluna contém o nome do autor; e a quarta o nome da variante. As colunas seguintes, relativas à frequência parcial, não são relevantes, visto que não há

variação da obra em relação ao autor ou à variante (por exemplo, a obra *O Primo Basílio* é totalmente escrita por Eça de Queirós, na variante de português de Portugal).

Para obter informação extralinguística sobre todas as obras da Linguateca, basta pedir no Distribuidor essa informação da seguinte forma:

```
obra autor variante data decada

```

e guardá-la num ficheiro com um nome apropriado. Escolhemos `distribuicaoObra.tsv` (/assets/investigar-literatura-lusofona-literateca/distribuicaoObra.tsv).

É importante esclarecer que algumas obras não possuem uma data específica, apenas o século a que pertencem. Nesse caso, pode-se atribuir uma data aproximada (como o ano 1650 para representar o século XVII) ou remover essas obras do ficheiro antes de processar os dados no R.[7]

Convém também converter os ficheiros para o formato UTF-8.[8]

O uso da linguagem R🔗

Para descrever e visualizar os resultados, podemos importar os ficheiros produzidos pelo Distribuidor no ambiente de programação R.

Roupa na literatura🔗

Primeiro, vamos observar quais os autores que têm mais menções a roupa.

```
roupa<-read.table("distribuicaoRoupa.tsv")
names(roupa)<-c("obra","roupa", "autor","variante",
"lixo","lixo2")
obras<-read.table("distribuicaoObra.tsv")
names(obras)<-c("obra","tamanho", "autor","variante","data",
"decada", "lixo","lixo2")
roupaObras<-merge(roupa, obras,
by=c("obra","autor","variante"))
roupaObras$rouparel<-roupaObras$roupa/roupaObras$tamanho
roupaObrasOrdenada<-roupaObras[order(roupaObras$rouparel,
decreasing=TRUE),]

```

De forma resumida, as quatro primeiras linhas leem os ficheiros e atribuem nomes às colunas. A quinta combina a informação das duas folhas de registo numa só. A sexta calcula a frequência relativa de roupa por número de unidades, criando uma nova coluna chamada `rouparel`. Por fim, a sétima linha obtém uma nova folha de registo ordenada pelo peso relativo do vestuário, que está na coluna `rouparel`.

Com os próximos comandos, podemos visualizar o resultado num gráfico de barras (Figura 6) e num gráfico de caixa (Figura 7), neste caso para dez autores que têm várias obras na Literateca.

```
par(mar=c(4,24,2,2)+0.1)
barplot(roupaObrasOrdenada$rouparel[1:25],
names=paste(roupaObrasOrdenada$autor[1:25],"-",roupaObrasOrdenada$c
horiz=TRUE)
par(mar=c(4,4,2,2)+0.1)
dezautores<-subset(roupaObras,roupaObras$autor=="JulDin"|
roupaObras$autor=="EcaQue"|roupaObras$autor=="RauBra"|
roupaObras$autor=="CoeNet"|roupaObras$autor=="MacAss"|
roupaObras$autor=="CamCBra"|roupaObras$autor=="AMBB"|
roupaObras$autor=="JosdAle"|roupaObras$autor=="AlmGar"|
roupaObras$autor=="AluAze",)
dezautores$autor<-dezautores$autor[drop=TRUE]
boxplot(dezautores$rouparel~dezautores$autor,
xlab="",ylab="",las=2)
```

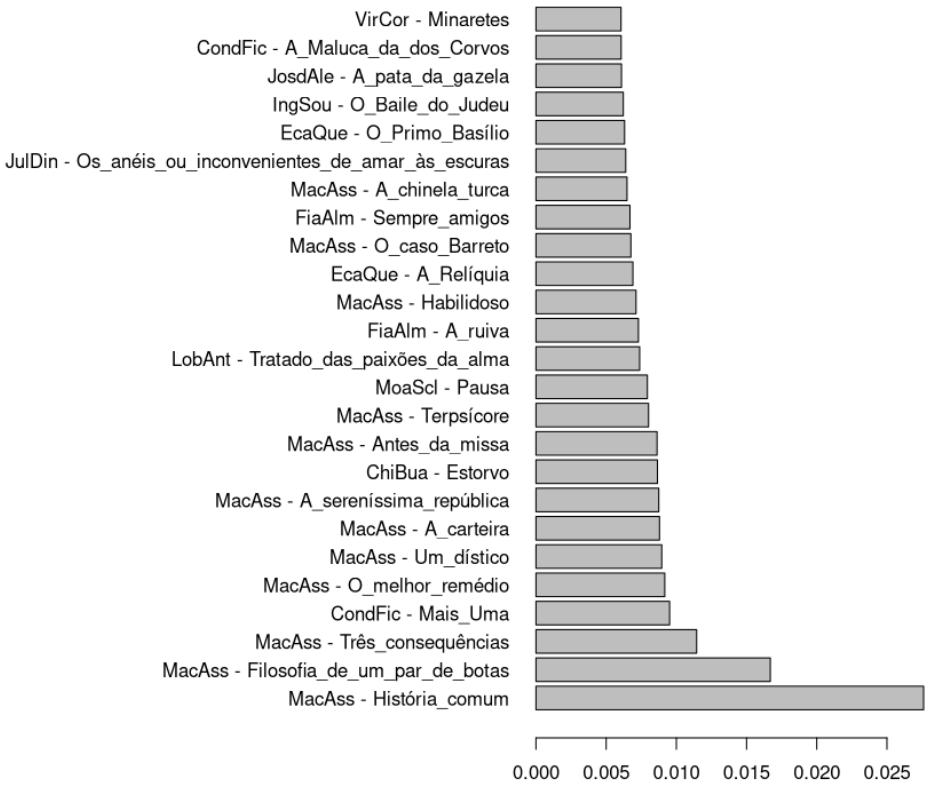


Figura 6. As vinte e cinco obras que mais referem roupa na Literateca, num gráfico de barras.

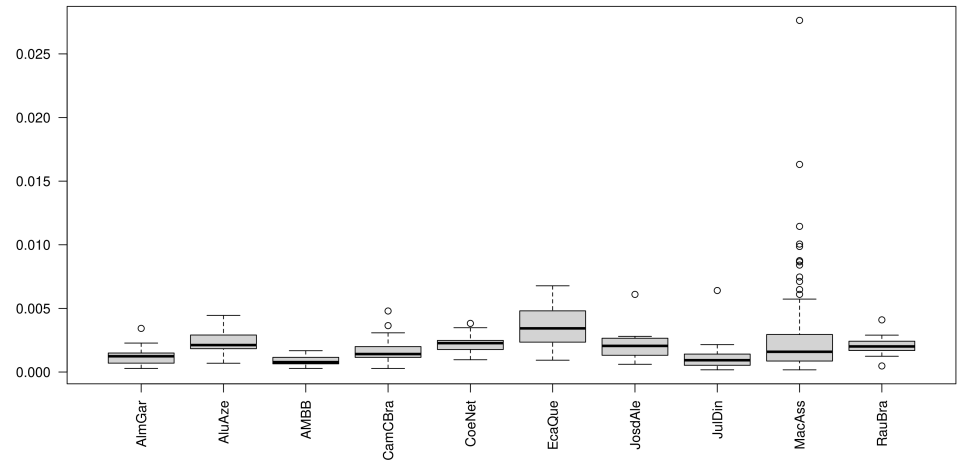


Figura 7. A distribuição de roupa por dez autores na Literateca, num gráfico de caixa.

Vemos pelas duas visualizações que, embora as obras com mais menção relativa a roupa fossem contos de Machado de Assis (Figura 6), ao considerar o conjunto das obras (Figura 7) é Eça de Queirós quem dá mais importância a esse campo semântico (a mediana de EcaQue é significativamente mais elevada do que a de MacAss).

Também podemos observar a menção ao campo semântico de vestuário ao longo do tempo. Para isso, usamos a data ou a década a que cada obra pertence (Figura 8).

```
boxplot(roupa0bras$rouparel~roupa0bras$decada,las=2,xlab="",ylab="
```

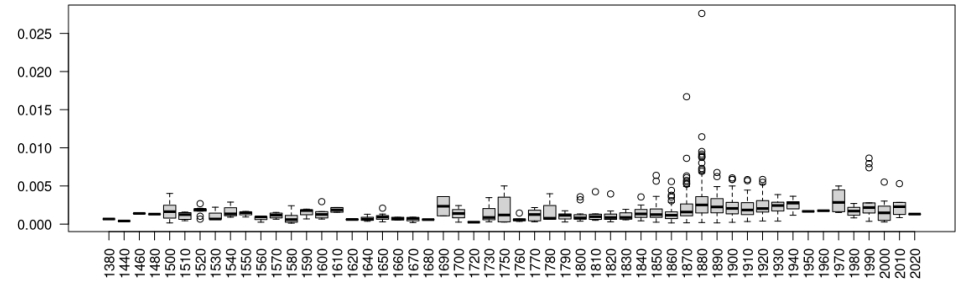


Figura 8. Distribuição de roupa por década na Literateca, num gráfico de caixa.

Pela figura podemos observar que a partir de 1870 existem muito mais referências ao vestuário na literatura lusófona do que anteriormente.

Diferenças entre as personagens femininas e masculinas🔗

No AC/DC, marcamos todas as caracterizações como pertencendo a uma de quatro classes:

- Emoção
- Carácter
- Aparência
- Social

Para explicação destas categorias e da forma de anotação, ver Freitas e Santos (2023).[9]

Vamos agora ver que casos femininos e masculinos estão marcados com

`pred:aparencia` ao longo do tempo.

No Distribuidor, pedimos a distribuição dos casos de aparência, selecionando apenas as obras literárias em prosa:

```
?sema=/.*pred:aparencia.*/ ?classe=/Prosa:.*/ ?decada=/[12].../
decada gen
```

e dos casos de predicação, seja ela qual for, em que se descreve uma pessoa (também selecionando as obras literárias em prosa):

```
?sema=/.*pred.*/ ?classe=/Prosa:.*/ ?decada=([12].../ decada
gen
```

Relembrando que:

- Escolhemos o corpo Literateca
- Escolhemos a opção tsv

e temos de guardar os ficheiros com nomes descritivos. No caso em questão, chamei-lhes `distribuicaoAparenciaDecadaGen.tsv` (/assets/investigar-literatura-lusofona-literateca/distribuicaoAparenciaDecadaGen.tsv) e `distribuicaoPredDecadaGen.tsv` (/assets/investigar-literatura-lusofona-literateca/distribuicaoPredDecadaGen.tsv).

No R, juntamos as duas informações, calculamos o peso relativo da aparência e depois produzimos uma figura que nos mostra a evolução ao longo do tempo:

```
apargen<-read.table("distribuicaoAparenciaDecadaGen.tsv")
names(apargen)<-
c("decada","aparencia","gen","tamapargen","lixo")
predgen<-read.table("distribuicaoPredDecadaGen.tsv")
names(predgen)<-c("decada","pred","gen","tampredgen","lixo")
aparencia<-merge(apargen,predgen,by=c("decada","gen"))
aparencia$genrel<-aparencia$tamapargen/aparencia$tampredgen
barplot(xtabs(aparencia$genrel~aparencia$gen+aparencia$decada),besi
```

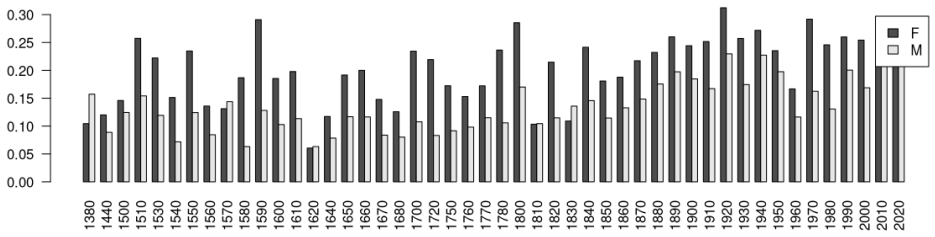


Figura 9. Caracterização da aparência feminina e masculina por década, num gráfico de barras.

Na Figura 9, vemos que as mulheres têm quase sempre mais caracterização de aparência do que os homens, o que não deve constituir uma surpresa. Para obter mais informação sobre este tema e sobre a construção social do género, consulte o artigo Freitas & Santos (2023).[9]

A localização na literatura portuguesa🔗

Também está em curso um projeto de anotação de lugares, que distingue palavras que podem ser locais em alguns contextos e não noutros. No caso da sua associação efetiva a espaços geográficos é indicado qual o tipo e granularidade (cidade, país, rio, etc.) e, no caso de essas localidades serem reais, as suas coordenadas geográficas. Para mais informação, veja Viagem (https://perma.cc/43E4-EW89), assim como Santos & Bick (2021).[10].

Assim, podemos identificar quais as cidades mais mencionadas na literatura portuguesa, usando simplesmente esta procura no AC/DC:

```
[sema="Local:cidade.*" & variante="PT"]
```

Também podemos investigar qual a cidade mais mencionada por obra, usando o Distribuidor e guardando o resultado, por exemplo, em distribuicaoCidadesObra.tsv (/assets/investigar-literatura-lusofona-literateca/distribuicaoCidadesObra.tsv).

```
?variante=/PT/ sema=/Local:cidade/ obra lema
```

É naturalmente possível fazer um gráfico de barras que represente este resultado. Aqui, vamos comparar o número de locais empregues por autores diferentes, em romances e novelas, usando mais uma vez o Distribuidor e guardando o resultado em distribuicaoLocaisObra.tsv (/assets/investigar-literatura-lusofona-literateca/distribuicaoLocaisObra.tsv):

```
?variante=/PT/ ?classe=/Prosa:(romance|novela)/ ?sema=/
Local:.*/ obra autor
```

Vamos visualizar essa questão através de um gráfico de caixa no R. De notar que reutilizaremos o ficheiro distribuicaoObra.tsv (/assets/investigar-literatura-lusofona-literateca/distribuicaoObra.tsv) que obtivemos anteriormente. Além disso, conforme mencionado acima, editamos as datas marcadas com "séc…" e convertemos para UTF-8 antes de invocar o R.

```
locais<-read.table("distribuicaoLocaisObra.tsv")
names(locais)<-c("obra","num","autor","lixo","lixo2")
obras<-read.table("distribuicaoObra.tsv")
names(obras)<-c("obra","tamanho", "autor","variante","data",
"decada", "lixo","lixo2")
locaisObras<-merge(locais, obras, by=c("obra","autor"))
locaisObras$localrel<-locaisObras$num/locaisObras$tamanho
barplot(locaisObras[order(locaisObras$localrel,decreasing=TRUE),]$l
```

As primeiras quatro linhas apenas leem e identificam as colunas das folhas de registo. A quinta junta ambas as informações. Já a sexta calcula o número relativo de locais por número de palavras, para ser possível comparar obras de diferentes tamanhos.

Escolhi apresentar na Figura 10 o gráfico das cinquenta obras com mais locais relativos, calculado na sétima linha (depois de ordenar, pedi os casos de 1 a 50).

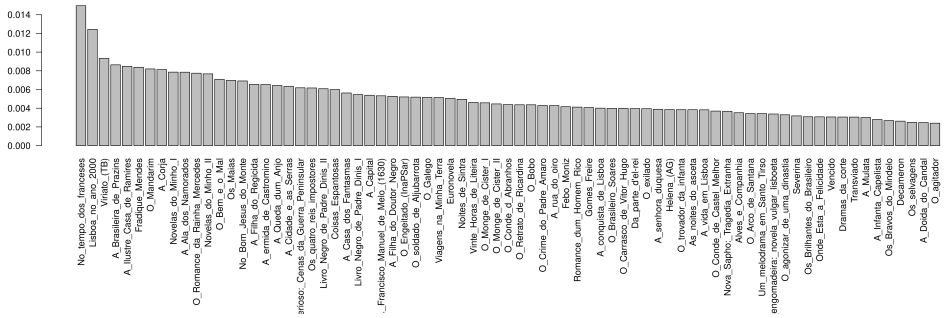


Figura 10. Distribuição de locais por obra (romances e novelas portugueses) na Literateca, num gráfico de barras.

É interessante constatar que são os romances históricos e de ficção científica os que mais dão nome a lugares.

Sugiro que faça também uma análise semelhante por autores, para ver (grandes) diferenças entre estes:

```
attach(locaisObras)
barplot(sort(tapply(num,autor,sum)/
tapply(tamanho,autor,sum),decreasing=TRUE)[1:25],las=2)
```

A primeira linha apenas instrui o R para se considerar "dentro" da folha de registo `locaisObras`, para não ser preciso estar sempre a preceder o nome da coluna pelo nome da folha de registo.

Na segunda linha, `tapply` é um comando no R que aplica uma função repetidamente. Neste caso é a função `sum` (soma), porque queremos somar todos os locais de um mesmo autor, sem interessar a obra, e todas as palavras escritas pelo autor (segundo `tapply`).

O helenismo na literatura brasileira🔗

Finalmente, apresento um estudo feito no âmbito da dissertação de mestrado de Marcus Vinicius Sousa Correia, que estudou o helenismo na literatura brasileira.[11]

O seu trabalho é um bom exemplo de como simples tarefas de anotação, em colaboração com o AC/DC, são fáceis de executar e produzem resultados interessantes.

De facto, Marcus apenas mandou um conjunto de lexemas que, segundo ele, estavam associados à cultura grega, e anotámos o corpo OBras, o principal corpo de literatura brasileira no AC/DC, com essa informação (com a marcação `helen`). Assim, tornou-se muito fácil medir o peso destas palavras num conjunto de autores brasileiros com obras no OBras.

O leitor é convidado a reproduzir as figuras da dissertação, visto que todos os comandos são apresentados na secção Anexos. Deixa-se para aperitivo a Figura 11, correspondente ao gráfico 3 da página 99.

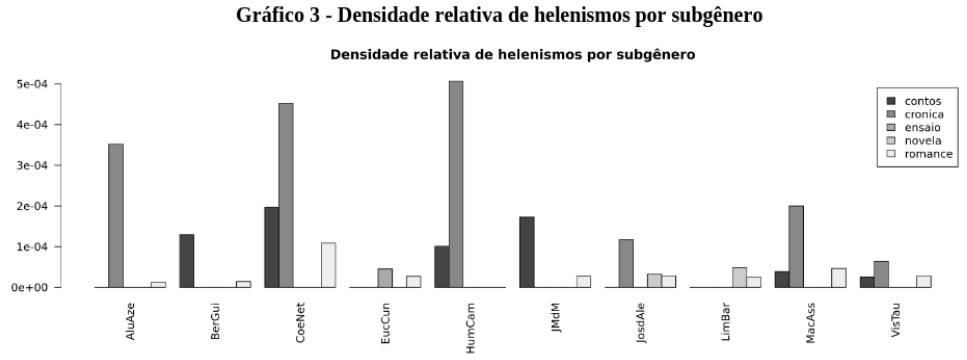


Figura 11. Distribuição de helenismos por autor e género de texto, num gráfico de barras.

Observações finais🔗

Nesta lição tentei apresentar o AC/DC, com duas formas de interação, a Procura e o Distribuidor e, depois, a partir dos resultados, obter gráficos agregadores usando o R.

O objetivo foi demonstrar diversas possibilidades de estudo da história da literatura lusófona, por meio da leitura distante, usando o AC/DC e, em particular, o corpo Literateca, que reúne textos literários em português e está em constante expansão.[12]

Notas de fim🔗

1. Bick, Eckhard, "PALAVRAS, a Constraint Grammar-based Parsing System for Portuguese", in *Working with Portuguese Corpora*, ed. Tony Berber Sardinha and Thelma de Lurdes São Bento Ferreira (Bloomsbury Academic, 2014), 279-302. ↩

2. Santos, Diana, "Corpora at Linguateca: Vision and roads taken", in *Working with Portuguese Corpora*, ed. Tony Berber Sardinha and Thelma de Lurdes São Bento Ferreira (Bloomsbury Academic, 2014), 219-236. ↩

3. Santos, Diana Maria de Sousa Marques Pinto dos, "A Gramateca e a Literateca como macroscópios linguísticos", *Domínios da Linguagem* 16, no. 4 (2022): 1242-1265. (https://doi.org/10.14393/DL52-v16n4a2022-2) ↩

4. Evert, Stefan and Hardie, Andrew, "Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium", in *Proceedings of the Corpus Linguistics 2011 conference* (University of Birmingham, 2011). pdf (https://perma.cc/3ZCG-3N24) ↩

5. Santos, Diana, "A sintaxe do AC/DC: apresentação do CWB e das opções tomadas", notas para a disciplina de POR2102, outono de 2012. pdf (https://perma.cc/U2PR-BLKU) ↩

6. Santos, Diana, "Explorando o vestuário na literatura em português", *TradTerm* 37, no. 2 (2021): 622-643. (https://doi.org/10.11606/issn.2317-9511.v37p622-643) ↩

7. Outra maneira ainda é especificar no Distribuidor que apenas pretende obter obras com data válida: `?data=/^[12].../ obra autor variante data decada .` ↩

8. O mais simples é ler, através de `read.table`, usando a opção `encoding="latin1"`. ↩

9. Freitas, Cláudia and Santos, Diana, "Gender Depiction in Portuguese: Distant reading Brazilian and Portuguese literature", *Journal of Computational Literary Studies* 2, no. 1 (2023). (https://doi.org/10.48694/jcls.3576) ↩ ↩[2]

10. Santos, Diana and Bick, Eckhard, "Distant reading places in Portuguese literature", *NorLit2021*, (Trondheim, 14-16 June 2022). pdf (https://perma.cc/5WBR-LFBL) ↩

11. Correia, Marcus Vinicius Sousa, "Helenismo nos trópicos: Análise da presença do Helenismo na literatura brasileira pelo viés da leitura distante" (Dissertação de mestrado, Universidade Estadual do Maranhão, 2023). pdf (https://perma.cc/G9QQ-M2TD) ↩

12. Agradeço sinceramente a Suemi Higuchi e a Larissa Freitas a sua revisão aturada desta lição, e as variadas sugestões de melhoria. ↩

SOBRE O(A) AUTOR(A)

Diana Santos é professora catedrática no Departamento de Literatura, Estudos de Área e Línguas Europeias da Universidade de Oslo, Noruega. (https://orcid.org/0000-0002-3108-7706)

CITAÇÃO SUGERIDA

Diana Santos, "Investigar a literatura lusófona através dos tempos usando a Literateca", *Programming Historian em português* 5 (2025), https://doi.org/10.46430/phpt0054.

ISSN 2397-2068 (inglês) (/)

ISSN 2517-5769 (espanhol) (/es)

ISSN 2631-9462 (francês) (/fr)

ISSN 2753-9296 (português) (/pt)

Concordância das traduções (/translation-concordance)