



UiO • **University of Oslo**



LITERATURA LUSÓFONA

PARTE DA LITERATURA MUNDIAL

recursos para leitura distante em português

Diana Santos, Cláudia Freitas & João Marques Lopes



cost

EUROPEAN COOPERATION
IN SCIENCE & TECHNOLOGY

Contexto

- A **Linguateca** como um centro de recursos, distribuído, para a língua portuguesa
 - Linguateca como infraestrutura
- *A Literateca < Linguateca*
- Uma ação COST na qual estamos envolvidos e na qual queremos incluir a literatura lusófona (e não só portuguesa)

Apresentando a ação *Distant Reading for European Literary History* (CA16204)

- <https://www.distant-reading.net/>
- Dinamizar a história da literatura através da construção de corpos e metodologias que a possam estudar de formas inovadoras
- Financiada pela União Europeia; primeiro foco → literaturas europeias
- Por questões pragmáticas, a coleção a construir será de romances (e novelas) publicados entre 1840 e 1919

COST is supported by the EU Framework
Programme Horizon 2020



Literatura lusófona?

- Ótima oportunidade para estudar e aplicar esses métodos à literatura lusófona em geral
- Restrição (?) COST:
 - coleção de obras em português publicadas na Europa (1ª fase)
 - coleção sem o “filtro” europeu → coleção alargada (“extended collection”).
- Coleção alargada: Literaturas brasileira, portuguesa, e africanas de expressão portuguesa
 - *Diferenças? Semelhanças?*
 - Questão empírica a ser investigada com base no material compilado
 - Diferenças geográficas se sobrepõem às divisões de escola/estilo?
 - Perspectiva de literatura COMPARADA – não está em jogo a “unificação” das literaturas
- Linguateca: interesse na coleção alargada + colaboração estreita com representantes de Portugal na ação COST
 - *Ambas as coleções são importantes!*

Recursos incluídos na *Literateca* que podem ser usados nessa coleção

Literateca (e nossa proposta para o COST)

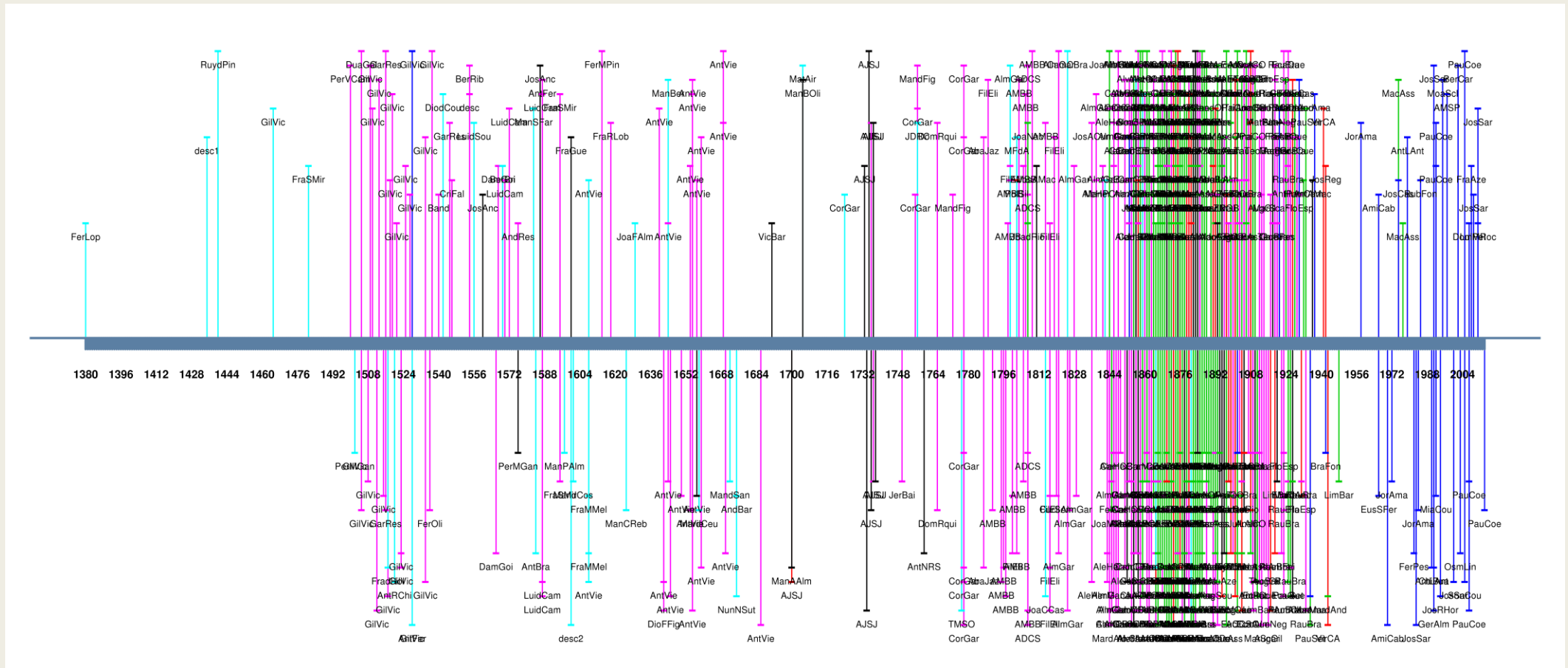
- Corpo Vercial
- Corpos OBRAS e NOBRE (criados pela Linguateca, Oslo e Rio de Janeiro)
- Corpo Tycho Brahe (<http://www.tycho.iel.unicamp.br/corpus/index.html>) (Galves et al., 2017)
- Corpo COLONIA (<http://corporavm.uni-koeln.de/colonia/>) (Zampieri & Becker, 2013)

Cerca de 700 obras literárias em português

O que conta como uma «obra literária»?

- Dilemas da seleção (o que é o canone? O que conta como obra literária?)
- Unidade: conto/ensaio ou a obra que o(s) contém?
- *Mais de 150 obras (romances ou novelas) cuja primeira publicação é entre 1840 e 1919*
- Até agora apenas 5 (das 151) escritas por (3) mulheres! Maria Firmino da Costa, Ana de Castro e Almeida e Júlia Lopes de Almeida

Descrição cronológica da *Literateca*

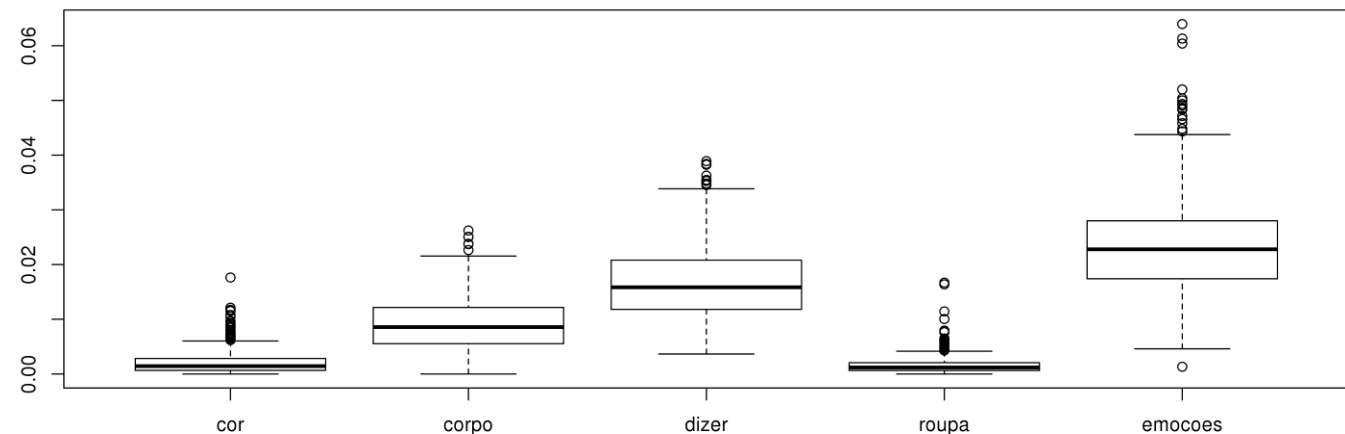


■ Conteúdo em 31 de março de 2018. Veja-se <https://www.linguateca.pt/Literateca>

Descrição da anotação dos textos

A vantagem de sermos linguistas com corpos (“corpus linguists”): Os textos são

- anotados sintaticamente pelo PALAVRAS (Bick, 2000), entre outros:
 - *Orações*
 - *Tempos*
 - *Argumentos de um verbo*
- anotados semanticamente usando uma colaboração entre anotadores e sistemas (Mota e Santos, 2010), englobando
 - *Referência a emoções*
 - *Corpo, roupa, cores*
 - *Discurso direto, indireto etc.*



O que se pode pesquisar com *leitura distanciada*, ou *leitura distante*?

- Influências literárias entre autores
- Temas da moda e referências que passam de moda
- Descrição das personagens
- Ligação entre elas
- Mapas cobertos por diferentes obras e/ou épocas
- Emoções presentes ou referidas pelos textos
- Comunicação com o leitor
- Intertextualidade
- Etc. etc.

Análise das descrições de personagens

Trabalho inicial sobre uma versão anterior do OBras... (Freitas, 2017)

femininas



masculinas

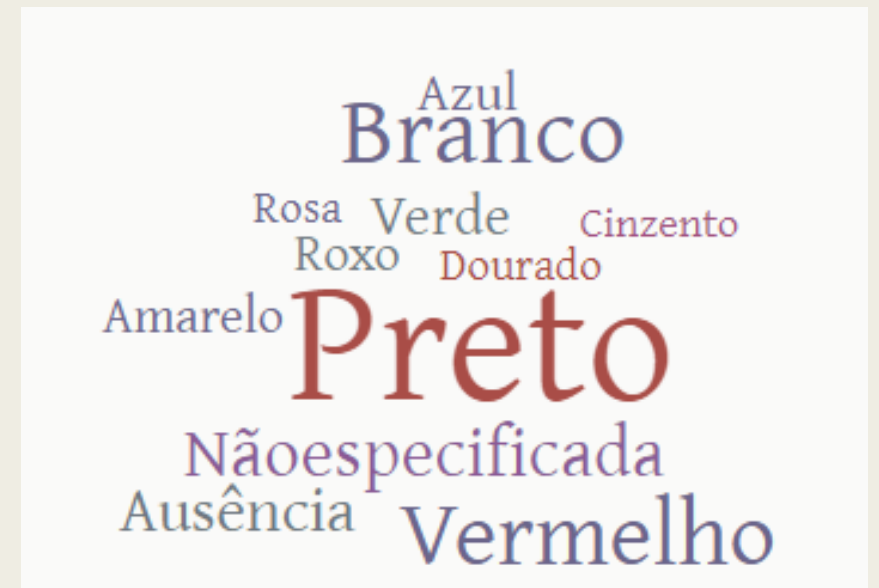


Comparação de obras com base em atributos semânticos – neste caso a cor

A Afilhada (Manuel de Oliveira Paiva, 1899)

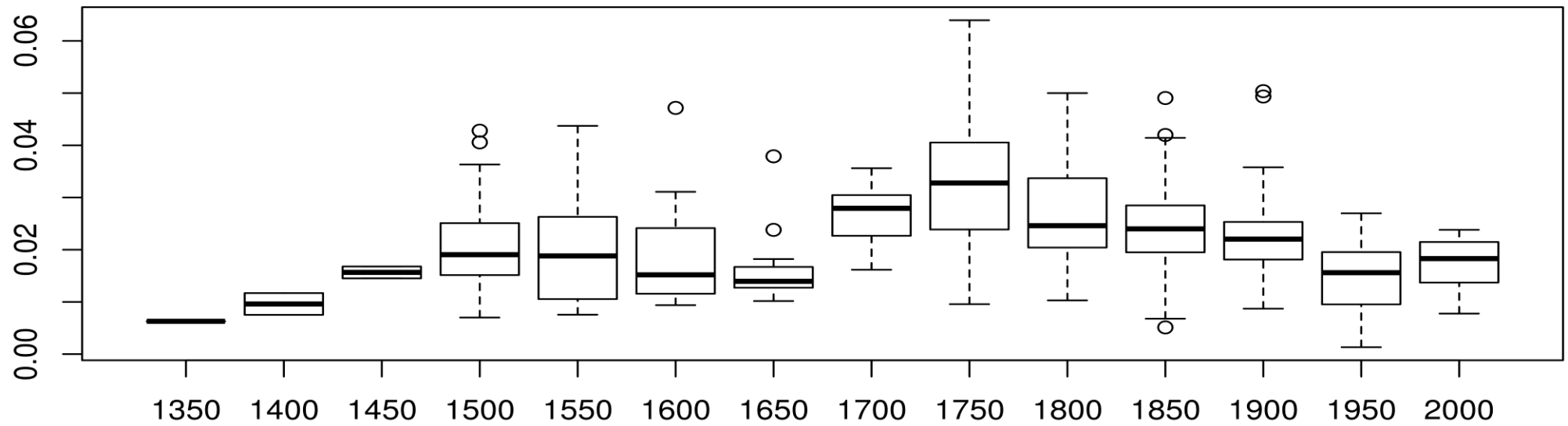


O Mulato (Aluísio Azevedo, 1881)

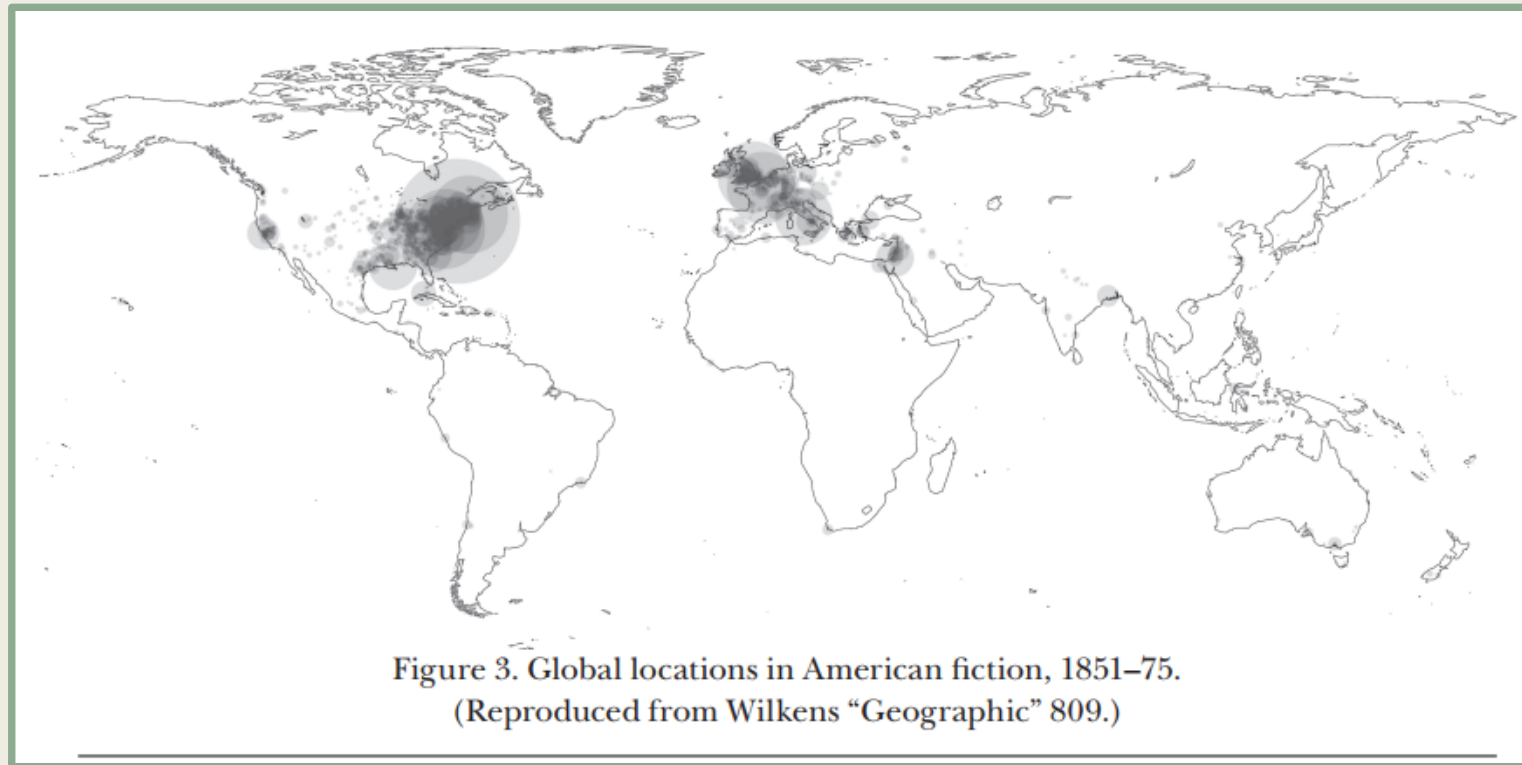


Ou as emoções...

- A distribuição da referência a emoções ao longo do tempo (~700 obras)

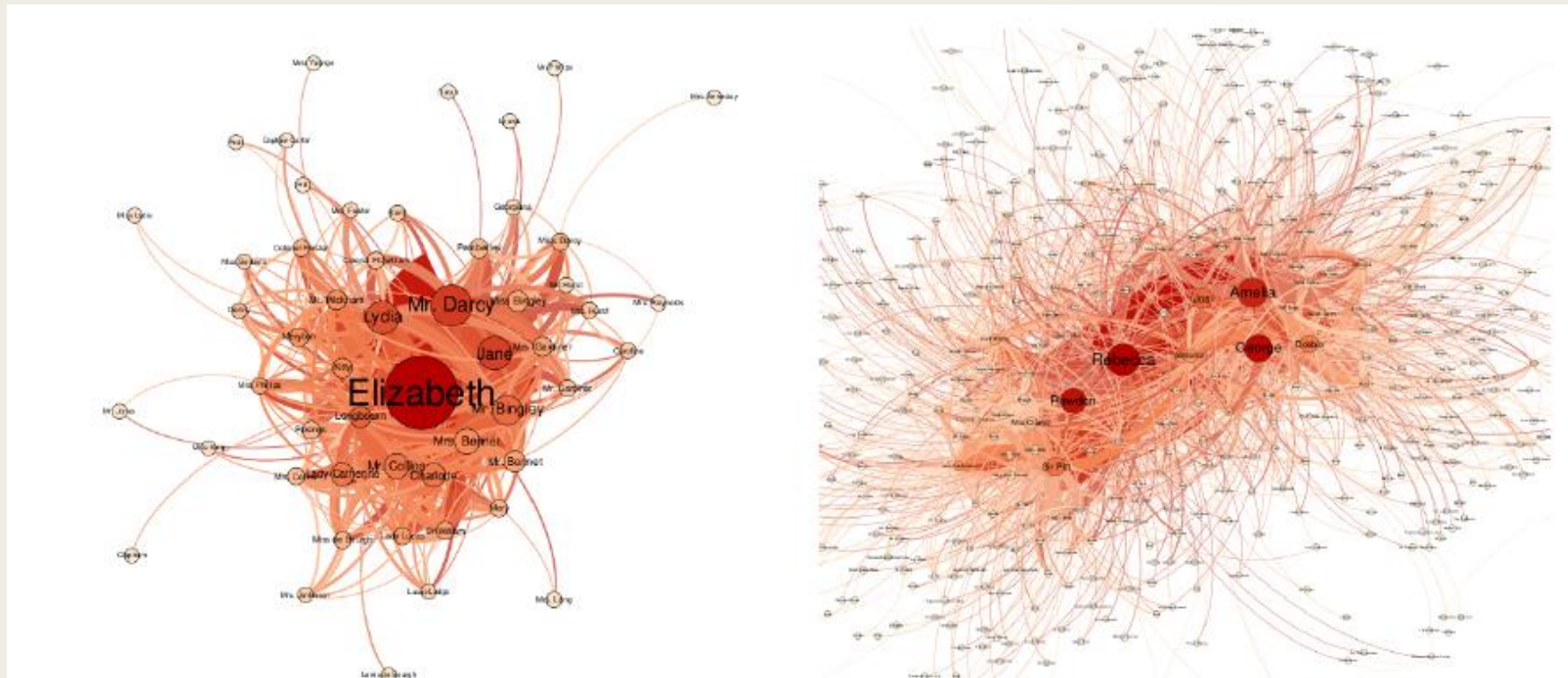


Fazer o “mapa” da literatura



Redes de personagens

De Ardanuy & Sporleder (2014)



Pride and Prejudice, Jane Austen

Vanity Fair, Thackeray

Que autores dialogam explicitamente com o leitor – ou a leitora?

- *the most reader-communicative authors are XIXth century novelists, namely Machado de Assis, Camilo, Júlio Dinis, Alexandre Herculano, Raul Pompéia and Almeida Garrett. Only the three first frequently addressed their female readers, though. (Santos 2018)*
- É preciso uma leitura atenta, depois, para ver quando é que eles recorrem aos leitores (vs personagens lendo)

NOBRE

leitor	96
leitores	34
leitora	27
Leitores	12
leitoras	6
Leitor	4

OBras

leitor	311
leitora	80
leitores	72
leitoras	8
Leitora	1

Vercial

leitor	1032
leitores	274
leitora	75
leitoras	63
Leitor	35
Leitores	27

Tycho Brahe

leitor	114
leitores	37
leitoras	9
Leitor	7
leitora	2
Leitores	2

Para terminar...

- Fase inicial: articulação e colaboração com o COST
- Interesse em ter ajuda e colaboração por parte da audiência, e de todos os interessados → é uma proposta de trabalho conjunto
- Há muito mais para fazer do que aquilo que nós conseguiríamos sozinhos!
- A literatura lusófona é de todos nós

- Comentários, sugestões, críticas?

Referências

- Ardanuy, Mariona Coll & Caroline Sporleder. "Structure-based Clustering of Novels". *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLfL) @ EACL 2014*, Gothenburg, Sweden, April 27, 2014, pp. 31–39, ACL, 2014.
- Bick, Eckhard. *The Parsing System "Palavras": Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press, 2000.
- Freitas, Cláudia. "Literoteca (OBRAS + VERCIAL +...) + COST Action Distant reading for European Literary History". Dez 2017.
- Galves, Charlotte; Andrade, Aroldo Leal de; and Faria, Pablo (2017, December). *Tycho Brahe Parsed Corpus of Historical Portuguese*.
- Mota, Cristina & Diana Santos. "Experiments in human-computer cooperation for the semantic annotation of Portuguese corpora". In Proc. LREC 2010, Malta, ELRA
- Santos, Diana. *Literature studies in Literoteca: between digital humanities and corpus linguistics*. In press.
- Wilkens, Matthew. "Digital Humanities and Its Application in the Study of Literature and Culture." *Empirical and Systemic Approaches to the Study of Literature and Culture*, a special issue of *Comparative Literature* 67.1. Ed. Steven Tötösy de Zepetnek (March 2015) 11–20.
- Zampieri, Marcos & Martin Becker. "Colonia: Corpus of Historical Portuguese". In Marcos Zampieri & Sascha Diwersy (eds.), *ZSM-Studien, Schriften des Zentrums Sprachenvielfalt und Mehrsprachigkeit der Universität zu Köln Shaker*, 5, 2013, pp. 77-84.