



COST is supported by the EU Framework Programme Horizon 2020

A ação COST e o que representa e representou para a leitura distante em português

Isabel Araújo Branco

Diana Santos

Paulo Silva Pereira

Raquel Amaro

História e apresentação dos autores

- Proposta inicial em 2013 de Christof Schöh, enviada por Christian-Emil Ore a Diana Santos por pertencerem ambos ao grupo de DH-Oslo
- Participação muito modesta de Diana Santos na proposta, mas **insistência no português**
- Após várias modificações e melhorias (três etapas, e a necessidade de incluir “early career members”, um deles o João Marques Lopes, pela UiO em 2016), a ação foi aprovada pela EU em 2016
- Diana Santos ficou como representante da Noruega, junto com Christian-Emil Ore

Participação de Portugal

- Raquel Amaro (linguista computacional) foi a primeira representante de Portugal pela Universidade Nova de Lisboa
- Isabel Araújo Branco (estudos literários, da mesma universidade) foi contactada pela Raquel para também pertencer à ação – ficou como membro substituto
- Paulo Silva Pereira (estudos literários), da Universidade de Coimbra, tornou-se o outro representante de Portugal

(um país pode ter no máximo dois representantes plenos e dois substitutos)

Isabel Araújo Branco

- Professora Auxiliar na Universidade Nova de Lisboa, onde dá aulas de Literaturas Comparadas, Literaturas Hispânicas e Tradução de Espanhol para Português.
- Investigadora integrada do CHAM-Centro de Humanidades (NOVA FCSH-UAc), onde coordena o Grupo «Cultura, História e Pensamento Ibéricos e Ibero-Americanos».
- Doutorada em Estudos Literários Comparados (tese: «A recepção das literaturas hispano-americanas na literatura portuguesa contemporânea: edição, tradução e criação literária»).
- Colabora regularmente com a Casa da América Latina.



Raquel Amaro



- Professora Auxiliar na Universidade Nova de Lisboa, Departamento de Linguística.
- Investigadora do Centro de Linguística da
- Universidade Nova de Lisboa, membro da direção
- Doutorada em Linguística Computacional, pela Universidade de Lisboa
- Áreas de interesse e investigação: semântica lexical computacional, interface sintaxe-semântica, linguística de corpus, *parsing*.

Paulo Silva Pereira

- Professor Auxiliar da Universidade de Coimbra (Faculdade de Letras/DLLC)
- Doutorado pela mesma universidade, leciona nas áreas de Literatura Portuguesa; Teoria da Literatura; História e Periodização da Literatura Portuguesa; Estudos Interartes; Literatura, Memória e História (Doutoramento em LLP); Literatura, Artes e Média (Doutoramento FCT em Materialidades da Literatura); Introdução às Humanidades Digitais
- Diretor do Programa de Doutoramento em Literatura de Língua Portuguesa (DLLP)
- Membro Integrado do Centro de Literatura Portuguesa (CLP) e coordenador do Projeto "Ex Machina: Inscrição e Literatura"
- Unidades curriculares com matéria relativa a Leitura Distante: Teoria da Literatura; Literatura, Artes e Média; Introdução às Humanidades Digitais

Diana Santos

- Engenheira da linguagem
- Processamento de linguagem natural
- Linguística com corpos
- Áreas preferidas: Semântica, avaliação, tradução
- Líder da Linguateca desde 1998
- Professora na Univ. de Oslo desde 2011
(português, métodos estatísticos, tradução)

Literatura lusófona, ou portuguesa?

- Várias posições
- Várias questões políticas
- Várias questões político-literárias

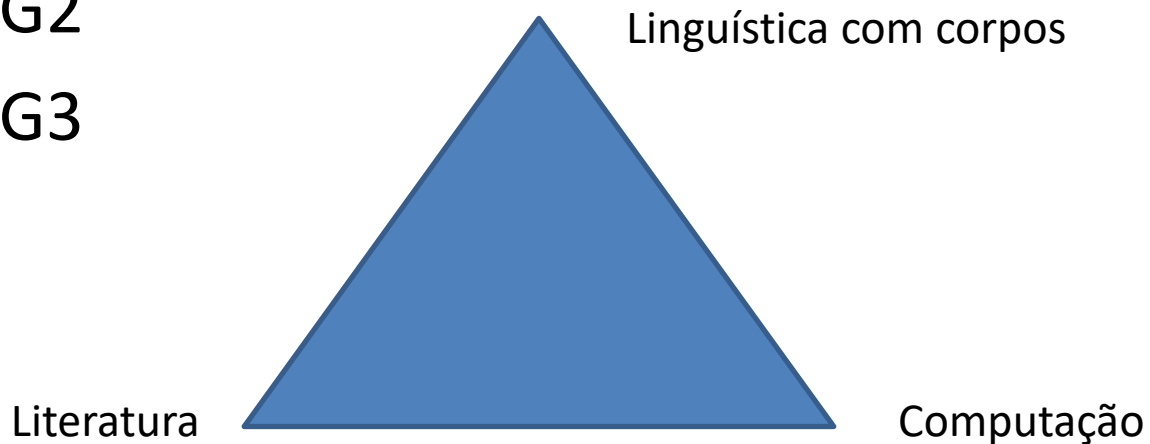
- De momento (mas em discussão em cima da mesa):
 - Coleção de literatura portuguesa
 - Coleção alargada incluindo literatura brasileira
 - Brasil (PUC-Rio) como observador oficial no COST, junto com a Austrália e a Ucrânia

Ação COST CA16204

- *Distant Reading for European Literary History*
- COST Site oficial do COST:
<https://www.cost.eu/actions/CA16204/>
- Site do projeto: <https://www.distant-reading.net/>
- Período: 4 anos, novembro de 2017 – outubro de 2021
- 29 países, 13 coleções

Três vertentes

- Correspondentes a três áreas na ação
 - WG1
 - WG2
 - WG3



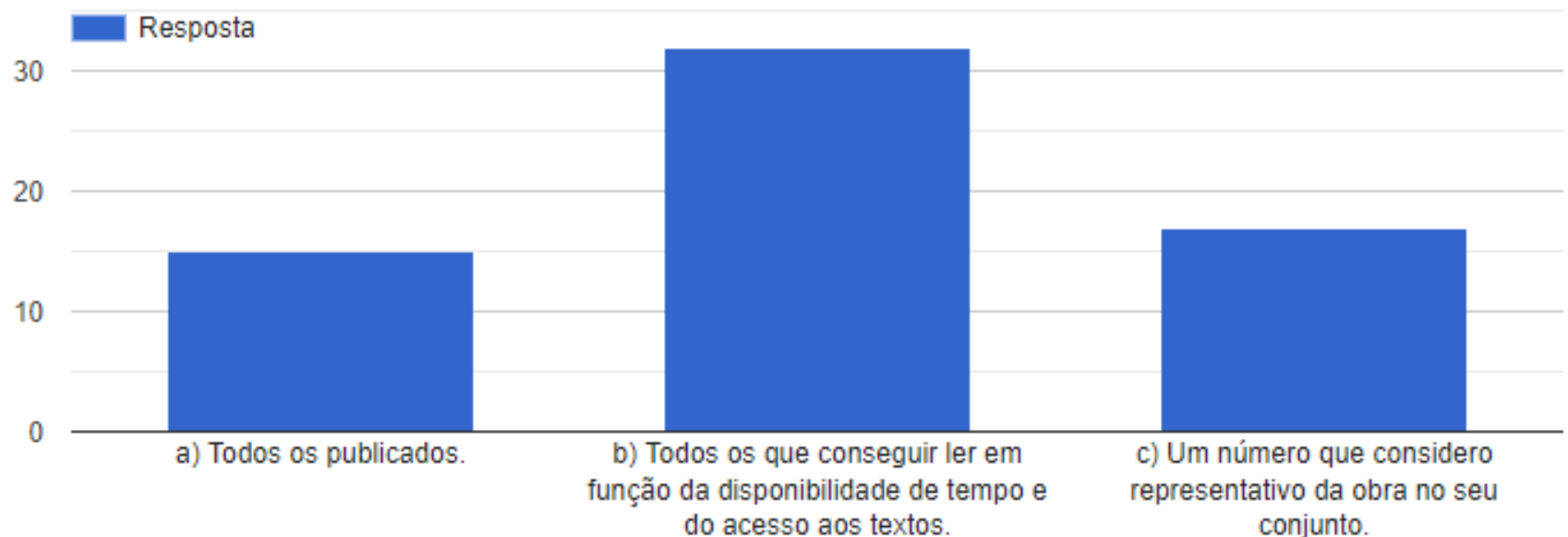
Roteiro

- Quem somos
- O que é esta ação COST
- Um inquérito sobre os hábitos de pesquisa nos estudos literários em Portugal
- Relatórios do que já se fez no COST
 - Construção da coleção ELTeC-por
 - Estudos sobre ferramentas existentes e futuras
 - Estudos teóricos

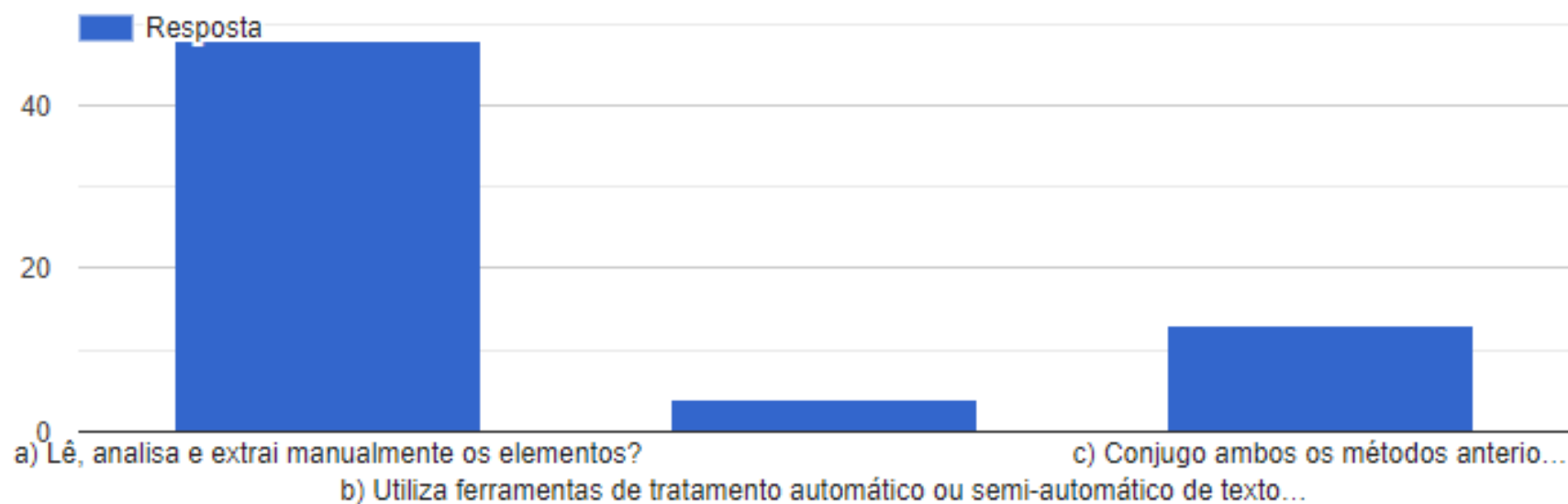
Inquérito sobre leitura distante em Portugal

- Efetuado por Isabel Araújo Branco e Raquel Amaro em setembro de 2019.
- Contactados oito centros de investigação e oito departamentos de sete universidades portuguesas (UL, UNL, UC, UP, UAI, UÉ e UBI).
- 62 respostas de investigadores em estudos literários.

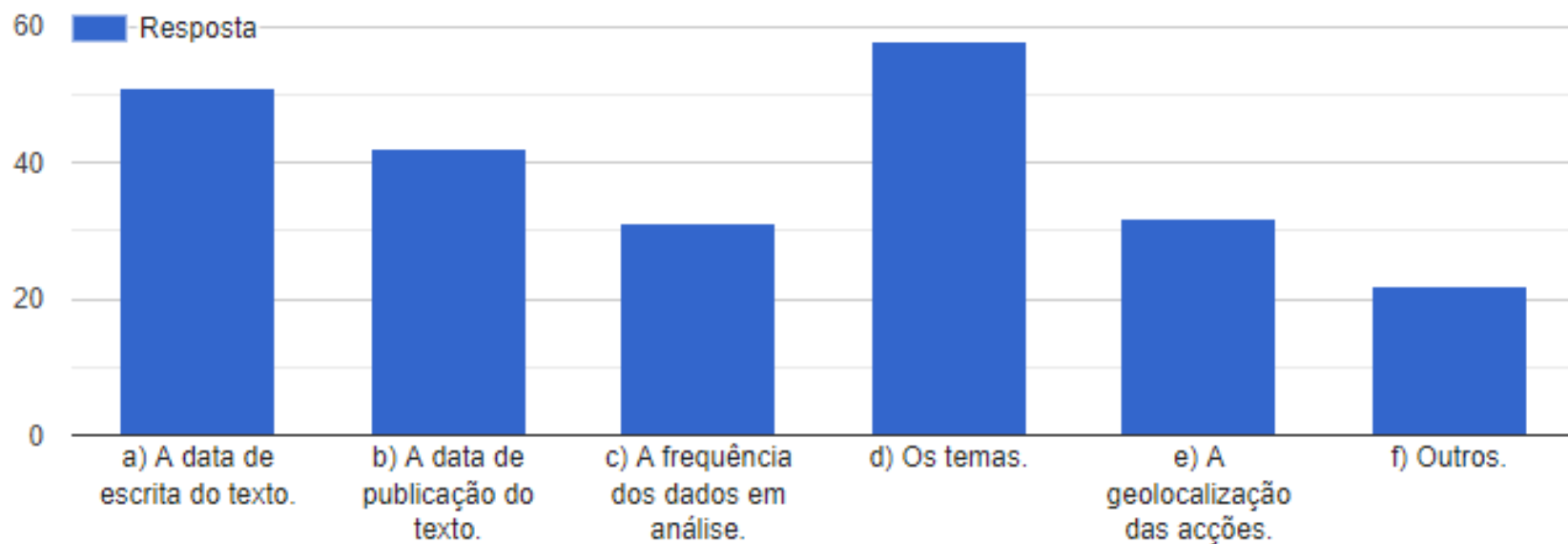
1. Em média, quantas obras lê do mesmo autor para classificar o período ou movimento em que se integra, o estilo e outras características dos textos narrativos?



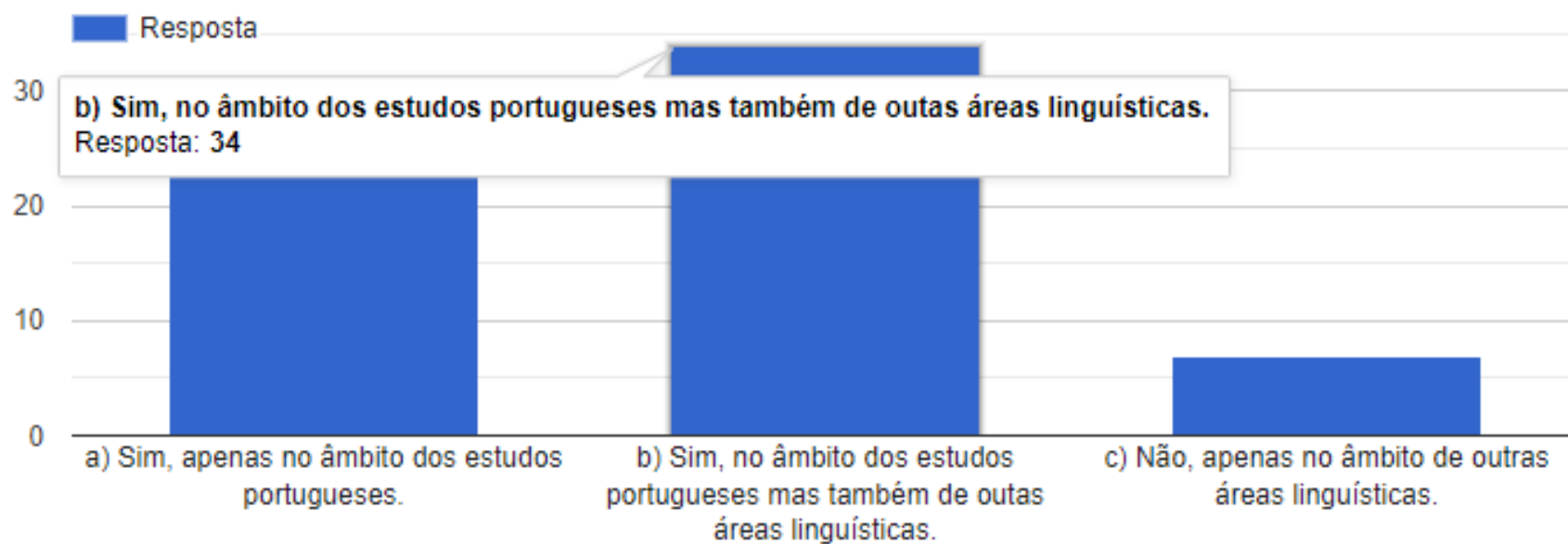
2. Como faz a caracterização de um texto?



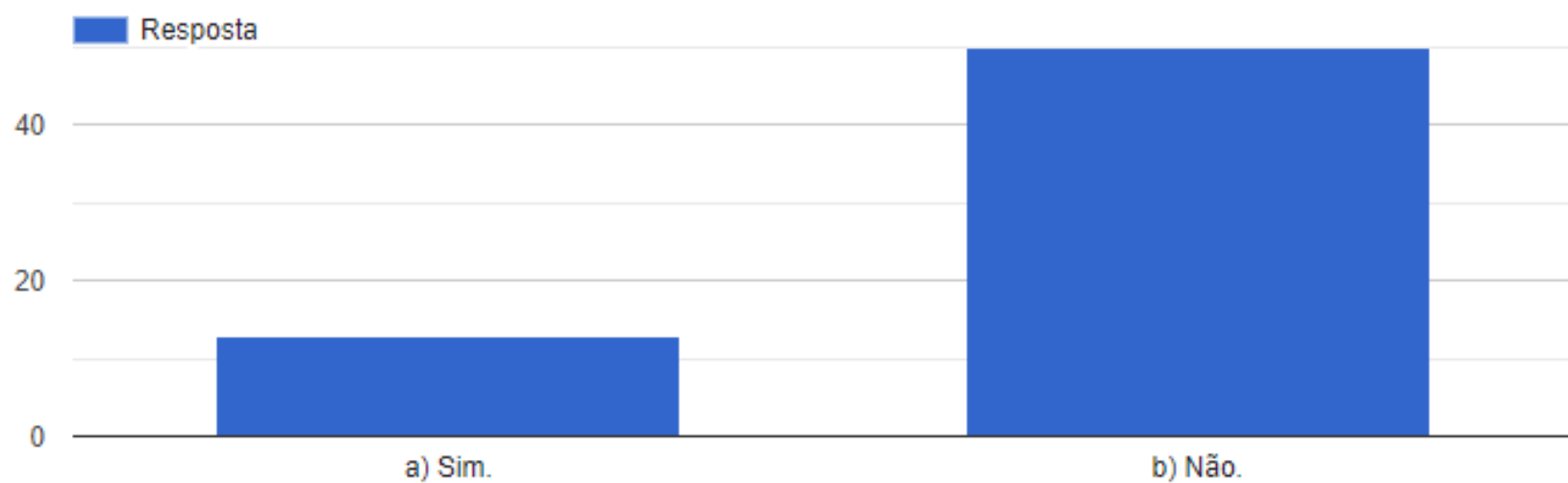
3. O que considera na análise dos dados?



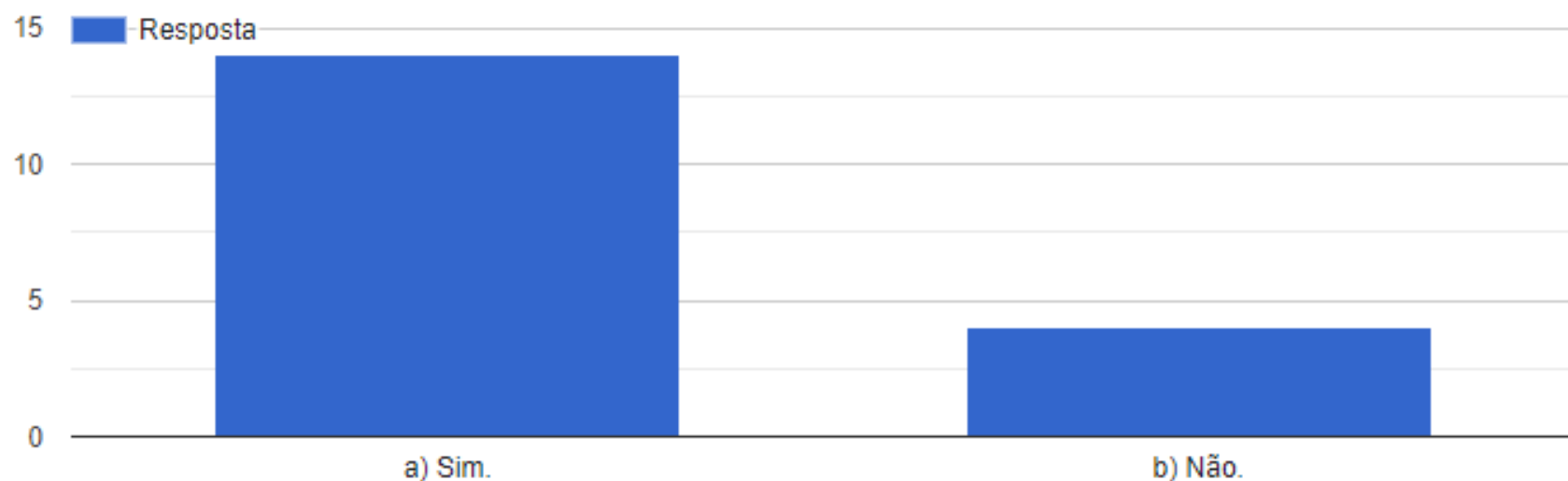
5. Integra o seu trabalho no âmbito dos estudos portugueses?



6. Conhece a «Distant Reading» ou «Leitura Distante»?



7. Se conhece a «Distant Reading» ou «Leitura Distante», considera-a útil para o seu trabalho?



«A Distant Reading é uma revolução metodológica em comparação com os métodos com que fomos formados, se bem feita pode, de facto, abrir imensos novos horizontes para a investigação na literatura, questionando a fundo objectivos do nosso trabalho.»

«Usada com moderação mas sem substituir a leitura profunda.»

«A leitura distante falseia frequentemente a recolha e consequente análise dos dados relativos aos textos. Várias estatísticas de Franco Moretti estão erradas devido a confusões entre o que as obras anunciam no título e o que elas realmente são.»

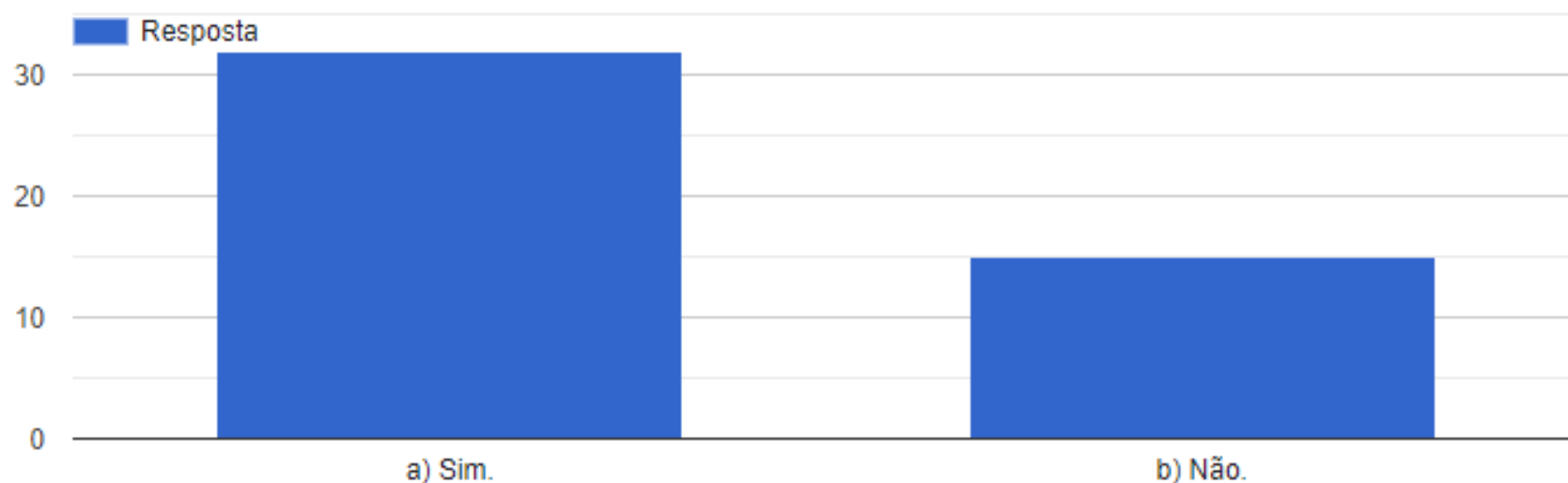
«É mais uma metodologia de análise a que vale a pena estar atento, embora não a considere a mais determinante.»

«Utilizo-a pontualmente porque me permite apreciação rápida de um grande número de textos. O facto de a) trabalhar com textos anteriores ao século XX; b) me interessarem muitas vezes textos não canónicos; c) os textos analisados não possuírem versão digital; d) ou me interessar pela leitura lenta ou pela *close reading*, são 4 razões para não a utilizar mais vezes.»

«Facilita a pesquisa de dados.»

«Pelo volume de obras a ler, pela facilidade de extrair dessas obras informação geolocalizada e pelas capacidades de análise do seu conteúdo, os métodos de leitura distante são essenciais aos meus trabalhos no âmbito do Atlas das Paisagens Literárias.»

8. Se está a tomar contacto com a «Distant Reading» ou «Leitura Distante» pela primeira vez, parece-lhe útil para o seu trabalho?



«É mais uma forma de trabalhar textos literários.»

«Tendo em conta a informação oferecida no preâmbulo a este inquérito. Parece-me, porém constituir muito mais um instrumento de análise entre muitos outros possíveis. Se tomado como único meio de validação de um trabalho de análise textual, será obviamente extremamente redutor no seu alcance.»

Terei de ter mais elementos convincentes da sua pertinência.»

«Dependendo das áreas cronológicas recobertas e, ainda, da sua possibilidade de não reflectir apenas o cânone.»

«Pode ser interessante integrar outras perspetivas de análise.»

«O armazenamento de dados e informações é relevante não só para a leitura, mas para a análise e aprofundamento do texto e dos seus vários sentidos.»

«Somos investigadores em literatura e o nosso dever é ler de verdade.»

«Sim, na medida em que - pelo que percebi - poderá facilitar a análise de textual uma vez que "lê" o texto à luz de dados previamente recolhidos acerca das estruturas estilísticas mas também genológicas e periodológicas que nele detetou; poderá iluminar aspetos até então desconhecidos ou pouco claros em alguns textos, favorecendo não só o trabalho de análise, como também agilizando o tempo da mesma, proporcionando, do mesmo modo, espaço para a análise de outros textos, sejam estes mais recentes ou não. Sim, ainda, porque facilita o trabalho de identificação dos hipotextos e de outras marcas transtextuais essenciais para a compreensão do texto em análise.

Não, porque devem ser previstos/acautelados erros de interpretação e/ou leitura errónea de dados pelo processo de generalização computacional, uma vez que poderá não considerar dados que se encontrem nas subcamadas do texto e que sejam importantes para que se traga nova luz sobre esse mesmo texto, sendo a plurissignificação uma das características essenciais de literariedade/do processo comunicacional de um texto, mas também é prova da sua intemporalidade. Não, sobretudo, porque padroniza/uniformiza/impõe/prescreve a/uma leitura e análise do texto literário, eliminando a subjetividade necessária de que vive o trabalho de crítica e análise textual no âmbito dos estudos literários.»

«Pelo que li no início deste inquérito parece-me um meio ilusório e absurdo de estudar literatura. A literatura não é "cientificável" e muito menos analisável desta forma simplista.»

«Parece-me uma forma de simplificação de um processo (a leitura) que só tem valor pela sua complexidade, como exercício intelectual que se pretende enriquecedor.»

«Pode dar-me mais domínios de análise de texto.»

«O uso combinado de métodos comparatistas e informáticos potencia os resultados pretendidos e a inovação.»

«É um regresso à Estilística, com dados mais fidedignos.»

«Aumento da informação classificada.»

«Trata-se de uma ferramenta de trabalho tão útil como qualquer outra cuja função seja a de facilitar o trabalho sobre obras de larga projeção.»

«A possibilidade de trabalhar grandes corpus por vários investigadores e em diferentes lugares, com perspetivas diversas, permite uma caracterização mais fundamentada e uma imagem do todo.»

«Trabalho de grupo e pesquisas deste género valorizam algo que está a ser desbastado nos currículos escolares e desconsiderado por políticas escolares e académicas, com consequências desastrosas para o homem enquanto ser humano vivendo em sociedade em harmonia e não em competição, com capacidade de reflexão, enquanto cidadão, e que sabe ter pausas neste mundo em enorme velocidade, e depois diz, ainda adolescente, que não tem tempo ou paciência para ler um romance como "Os Maias", "O Primo Basílio", "Memorial do Convento", "O Ano da Morte de Ricardo Reis" ou "A Casa do Pó": e esse algo é a literatura.»

«Creio que não. Sigo um outro método de análise textual.»

«Parece ser mais um instrumento de trabalho.»

«É sempre bom aceder a mais informação.»

«Facilidade de acesso a textos pouco conhecidos.»

«Penso que facilite o acesso aos livros que não existem em Portugal e que estão digitalizados noutras bibliotecas pelo mundo fora, sendo estas mais disponíveis para os investigadores que as obras nos fundos nacionais.»

«Os métodos computacionais permitem uma validação mais objetiva de dados, desde que não sejam interpretados de forma meramente quantitativa.»

«Sendo uma perspectiva ainda fora do meu trabalho, conhecê-la poderia agregar algo interessante, complementar.»

Relatório sobre o grupo WG1

Boas notícias!

- A primeira coleção a ser tornada pública foi a portuguesa, v.0.7, 69 obras
- O português está em sexto lugar em número de obras (o primeiro é o húngaro, o segundo o francês):

<https://distantreading.github.io/ELTeC/>

Critérios para a coleção ELTeC(-por)

- Romances e novelas (*novels*)
- Publicados (primeira edição) entre 1840 e 1920 (1919)
- Dez autores com 3 obras, o resto só com uma obra, o que implica 80 autores diferentes!
- Equilíbrio possível entre
 - vintenas (1840-1859, 1860-1879, 1880-1899, 1900-1919),
 - tamanho em palavras (10000-49999, 50000-99999, >100000)
 - canonicidade (alta, baixa, não especificada)
 - género do autor

<https://distantreading.github.io/ELTeC/por/index.html>

Problemas na escolha das obras

- Muito poucas obras já existentes em formato de texto e revistas (grande interseção)
 - Projeto Vercial
 - Projeto Gutenberg
 - Luso-livros, livrolivre, projeto adamastor
- Difícil de encontrar obras mesmo em pdf
 - Archive.org
 - Projetos de digitalização: NuPiLL, Câmara Municipal de Setúbal, etc.
- Pedidos de digitalização à BN de Portugal

Revisão do texto

- As obras têm grandes problemas de ROC, devido a ortografia antiga e pouca qualidade da impressão
- Rever uma obra presente no archive.org leva entre dez a vinte horas de trabalho
- Apenas tivemos uma bolsreira, a Adeliana Silva (contratada pela Raquel Amaro), que tratou de dez obras
- Tentativas de “apoio das massas” não deram resultado (apelo a alunos do politécnico de Bragança)

Grupo 2

- Criação de amostras para estudar o tipo de problemas dos textos literários do século XIX e inícios do século XX
- Comparação de anotação gramatical (POS-tagging) para as várias línguas (primeiros resultados em Cinkova et al. 2019)
- Comparação de reconhecimento de entidades mencionadas (REM) para as várias línguas (primeiros resultados em Stankovic et al. 2019)
- Comparação de atomização e segmentação e deteção de discurso direto
- Comparação de “sentiment analysis”
- Escolha do formato XML para o nível 2 da coleção ELTeC (a determinar em janeiro de 2020)

Grupo 3

- O WG3 ("Teoria Literária e História") tem por missão:
 - i) explorar preocupações teóricas que decorrem da aplicação de métodos de Leitura Distante na história literária;
 - ii) refletir sobre questões que surgem da pesquisa complementada por dados sobre a história literária em diferentes tradições literárias na Europa;
 - iii) promover o debate sobre as vantagens e limitações das metodologias de Leitura Distante e abordagens computacionais para o estudo da história literária europeia;
 - iv) reavaliar e redefinir conceitos-chave da história literária como género, estilo ou autoria;
 - v) examinar os resultados obtidos e explorar as suas implicações em termos de periodização e canonicidade na história literária europeia.

Na prática

A existência desta ação COST foi o acaso que permitiu que quase todos aqui estejam presentes:

- 1) porque a Diana se começou a interessar pela leitura distante, e criou um grupo informal aqui em Oslo (também graças ao Ernest Bowes)
- 2) o que levou o Emanuel a contactá-la e ao João do Brasil, vir cá e depois pôr-nos em contacto com o Alckmar
- 3) a já muito antiga colaboração da Diana com a Cláudia fez com que ela também se interessasse pelo COST (e escrevemos o artigo Santos, Freitas & Lopes (2018) para o DHRio)

Referências

- Cinkova, Silvie, Tomaž Erjavec, Cláudia Freitas, Ioana Galleron, Péter Horváth, Christian-Emil Ore, Pavel Smrž and Balázs Indig. "Evaluation of taggers for 19th-century fiction". *DH Budapest 2019, Digital Humanities Conference* (Budapest, 25-27 September 2019).
- Santos, Diana. "Presentation of the COST action Distant Reading for European Literary History", ILOS, 30 august 2018.
<https://www.linguateca.pt/Diana/download/COSTagosto2018.pdf>
- Santos, Diana, Cláudia Freitas & João Marques Lopes. "Comparando a literatura lusófona com outras literaturas: recursos para leitura a distância em português". In Suemi Higuchi & Cláudio José Silva Ribeiro (eds.), *I Congresso Internacional em Humanidades Digitais no Rio de Janeiro (HdRio2018)*, CPDOC/FGV, 2018, pp. 375-383.
- Stanković, Ranka, Diana Santos, Francesca Frontini, Tomaž Erjavec & Carmen Brando. "Named entity recognition for Distant Reading in several European literatures". *DH Budapest 2019, Digital Humanities Conference* (Budapest, 25-27 September 2019).