

Leitura Distante em Português: resumo do primeiro encontro

Diana Santos, Daniel Alves, Raquel Amaro, Isabel Araújo Branco, Olivia Fialho, Cláudia Freitas, Suemi Higuchi, Marcia Langfeldt, João Marques Lopes, Alckmar Luiz dos Santos, Emanuel Pires, Barbara Ramos, Danielle Sanches, Rebeca Schumacher Fuão, Paulo Silva Pereira, Paula Terra

Este artigo pretende refletir sobre o *Primeiro Encontro sobre Leitura Distante em Português* (doravante ELD) que aconteceu na Universidade de Oslo a 27 e 28 de outubro de 2019.¹ Organizado com o objetivo de reunir pesquisadores do Brasil e Portugal interessados na área, o encontro congregou cerca de vinte pesquisadores em onze palestras e duas sessões de discussão para refletirem acerca de questões relevantes ao tema, como indicado na descrição do evento:

Neste encontro será discutida a recente subdisciplina dos estudos literários e da linguística computacional intitulada leitura a distância, leitura distanciada ou leitura distante, em relação à literatura lusófona. O objetivo é encontrar caminhos, juntar esforços, e propor metas culturalmente apropriadas.

Elaboramos este texto com o intuito de disseminar os assuntos e ideias discutidos nesses dois dias de encontro. Após uma descrição abreviada de todas as intervenções, tentaremos destacar os principais pontos de debate, assim como algumas sugestões avançadas pelos participantes, na esperança de contribuir para a ampliação do debate sobre leitura distante dentro da comunidade lusófona contemporânea.

¹ <https://www.hf.uio.no/ilos/forskning/aktuelt/arrangementer/gjesteforelesninger-seminarer/2019/encontro-sobre-leitura-distante-em-portugues.html>

Leitura distante

A leitura distante (em inglês, *distant reading*) é uma área interdisciplinar específica e em crescente evolução que combina os domínios dos Estudos Literários, da Linguística Computacional e da Informática Aplicada na análise de grandes coleções de textos, que, pela sua natureza, compreende dados de volume significativo. Os primeiros trabalhos desenvolvidos nesta área preocuparam-se com textos literários (Moretti, 1999, 2005), mas os seus usos não se restringem a este tipo de fontes.

O investimento nesta área tem contribuído ativamente para o crescimento da disponibilização e do tratamento digital de textos literários e para o desenvolvimento de ferramentas, métodos e algoritmos de extração de informação de várias ordens. Atualmente é uma das áreas mais ativas no âmbito das Humanidades Digitais (Schöch et al., 2017).

Por oposição à leitura próxima, ou seja, a leitura de aspectos e elementos internos de uma dada obra literária, em que a perspectiva qualitativa prima sobre a quantitativa, a leitura distante pretende dar primazia inicialmente à perspectiva quantitativa, nomeadamente ao eleger como objeto de leitura — como faz Franco Moretti no início desse campo de estudos — «todas» as obras de um país, período, continente, testando as caracterizações existentes da História da Literatura, do género literário, da periodização literária, da estilística, entre outras (Rybicki & Maciej, 2011).

O potencial de inovação desta metodologia no campo dos Estudos Literários torna-a extremamente relevante como base de convergência entre a história literária, de natureza interpretativa, e a análise de

dados quantitativos, abrindo portas a novas perspectivas de teoria e história literárias e a uma nova dimensão de estudos comparados entre períodos, línguas, culturas (Wilkins, 2015).

Desde a sua proposta por Moretti, vários pesquisadores têm contribuído para o desenvolvimento da área. Mathew Jockers (2013), Geoffrey Rockwell e Stéfan Sinclair (2016), Andrew Piper (2018) e Katherine Bode (2012, 2018) estão entre alguns dos que têm se detido de maneira mais aprofundada sobre o assunto, ainda que nem sempre utilizando das mesmas maneiras de abordar o objeto ou de demonstrar os resultados, e sendo, muitas vezes, extremamente críticos em relação ao próprio Moretti (Piper, 2017, Bode, 2017).

Contudo, mesmo que as potencialidades desse tipo de estudo tenham se evidenciado mais com o surgimento e a popularização da internet e dos processos de digitalização, é preciso deixar claro que uma abordagem interpretativa de textos, calcada em distribuições e frequências, por exemplo, existe há séculos, ainda que de forma embrionária.

Nesse sentido, o que se convencionou chamar de leitura distante, muito disso graças ao trabalho de Moretti, que cunhou a expressão, possui raízes mais profundas na historiografia dos estudos literários – e sociais também –, e não deve ser entendido como algo absolutamente novo. Para um panorama mais substancial sobre o conceito e suas relações com a crítica literária, conferir o trabalho de Ted Underwood, “A Genealogy of Distant Reading” (2017).

Para o português, além dos nossos próprios esforços, apenas conhecemos ainda o volume publicado em França em 2014, por Cabral (2014).

Intervenções no ELD

Alckmar Luiz dos Santos, professor da Universidade Federal de Santa Catarina, abriu o encontro apresentando as diversas atividades do NuPILL (Núcleo de Pesquisas em Informática, Literatura e Linguística), projeto que constitui a maior base de dados sobre histografia literária em português², coordenado por ele. Após descrever brevemente a história da instituição, o trabalho realizado e comentar sobre as áreas de ênfase recente, ele trouxe duas ferramentas publicamente acessíveis desenvolvidas pelo grupo: o *Aoidos*, ferramenta para análise de versos, e o *DLNotes2*, ferramenta para anotações livres e semânticas (Mittmann et al. 2016 e Mittmann et al. 2013, respectivamente).

<Inserir Fig1_Alckmar>

Figura 1: Interface do programa Aoidos.

<Inserir Fig2_Alckmar>

Figura 2: Usando a ferramenta DLNotes2 sobre *Dom Casmurro*, de Machado de Assis.

Emanoel Pires de Assis, professor da Universidade Estadual do Maranhão, descreveu seu trabalho no desenvolvimento do Portal Maranhão³, um exemplo de projeto utilizando o NuPILL aplicado para a literatura maranhense, seu principal interesse de pesquisa. Para seu pós-doutorado em 2020, ele pretende aprofundar-se na forma em que personagens negras, masculinas e femininas são retratadas nos romances maranhenses.

² <http://www.literaturabrasileira.ufsc.br>

³ <https://www.literaturamaranhense.ufsc.br/>

<Inserir Fig3_Emanoel>

Figura 3: O portal da Biblioteca Digital da Literatura Maranhense, do NuPILL.

Suemi Higuchi apresentou sua tese em andamento na área da história da política brasileira, e seu uso da leitura distante através do *Dicionário Histórico-Biográfico Brasileiro* (DHBB)⁴, cf. Higuchi et al. (2019). A pesquisadora comparou a potencialidade da leitura distante, que permite uma visão panorâmica dos dados, com a da leitura próxima, onde verbetes são analisados caso a caso. Por fim, enfatizou as especificidades de uma investigação histórica e a necessidade da revisão humana após a coleta inicial dos dados, essencial na extração de informação confiável.

<inserir Fig4_Suemi>

Figura 4: Lendo o Dicionário Histórico-Biográfico Brasileiro de forma distante.

Esta foi a única intervenção no ELD que não tratou especificamente do campo literário, embora o estudo das paisagens (geografia humana e física) também tenha aparecido na apresentação de Danielle Sanches, Daniel Alves e Diana Santos, explicando como seria possível obter informação relevante sobre outros domínios usando a literatura como matéria-prima. Em sua intervenção, além de apresentarem o projeto BILLIG⁵ através do qual se pretende cruzar metodologias de análise espacial e linguística computacional, os pesquisadores compararam as principais localidades geográficas na literatura brasileira e portuguesa em romances e novelas de 1840 a 1919 assim como discutiram brevemente os distintos perfis geográficos de duas obras canônicas

4 O material encontra-se aberto e disponível para consulta em <https://cpdoc.fgv.br/acervo/dhbb> na forma de verbetes integrais de uma base de dados; e o seu texto foi anotado e tornado pesquisável através do projeto AC/DC, de <https://www.linguateca.pt/acesso/corpus.php?corpus=DHBB>

5 <http://billig.fcsh.unl.pt/>

do Portugal oitocentista, *O Primo Basílio* e *As Pupilas do Senhor Reitor*. Para tanto, o Projeto BILLIG utiliza o *Atlas das Paisagens Literárias de Portugal Continental*⁶, procurando construir uma abordagem complementar entre leitura distante e leitura “tradicional” que concilie uma análise micro, sustentada pela leitura atenta e individual de cada obra literária, com uma análise macro (Moretti 2005 e 2013; Jockers 2013), cujo interesse está na identificação de padrões gerais através de métodos quantitativos e de análise, com recurso a sistemas de informação geográfica (Alves e Queiroz, 2013 e 2015). Esta apresentação utilizou dois mapas para ilustrar a aplicação dessa metodologia à Literateca, visando discutir com o público participante do ELD o aprimoramento futuro da tarefa de extração de dados geográficos do conjunto das obras, assim como sua visualização no âmbito do BILLIG.

<inserir Fig5_billig>

Figura 5: Um percurso literário no Atlas das paisagens literárias de Portugal Continental.

<inserir Fig6_billig>

Figura 6: Mapa literário das localidades mencionadas numa amostra de cerca de 200 obras das literaturas portuguesa e brasileira do século XIX, da Literateca.

Cláudia Freitas, professora da PUC-Rio, destacou em sua palestra alguns exemplos aplicados de corpora linguisticamente anotados em diferentes contextos nas Humanidades Digitais. Em primeiro lugar, apresentou um estudo sobre a caracterização de personagens femininas e masculinas feito através do corpus OBRas, utilizando a anotação morfossintática disponível em todo o material do AC/DC. O trabalho resultou numa quadripartição categórica, dividida entre características físicas, emocionais, sociais e de personalidade. Assim demonstrou-se que, na literatura brasileira contida no OBRas, há uma predominância de aspectos físicos nas descrições

⁶ <https://ielt.fcsh.unl.pt/paisagensliterarias/>

das mulheres e, em contrapartida, predominam traços de caráter na descrição dos homens. Em seguida, como outra dimensão potencial de caracterização de personagens, foram examinadas ocorrências verbais no discurso relatado, desta vez viabilizada pela anotação semântica dos verbos do dizer, também presente no AC/DC (Freitas et al., 2016). Novamente, foram utilizados o mesmo corpo, assim como as marcações de gênero feminino e masculino, como operadores analíticos. Por fim, Cláudia Freitas trouxe para a discussão a necessidade de atenção para algumas singularidades da língua portuguesa, em particular o fenômeno do sujeito oculto, relevante tanto para a caracterização de personagens quanto para a atribuição de quem fala. Constatou, por exemplo, que cerca de 30% das frases nas obras de Machado de Assis não contêm sujeito explícito (a esse respeito, refira-se Freitas et al., 2019).

<inserir Fig7_Claudia>

Figura 7: A caracterização das personagens femininas e masculinas no OBRas.

A pesquisadora Marcia Langfeldt apresentou alguns estudos e ferramentas utilizados no seu doutorado sobre a Amazônia na literatura brasileira (Langfeldt, 2018), levantando a questão dos estudos de estilometria e recepção como associados ou associáveis ao campo de leitura distante, mais especificamente no caso de Euclides da Cunha e sua obra *À margem da história*. Tendo escrito sua tese num contexto francófono, destacou alguns exemplos de iniciativas de *humanités numériques* que colocam em evidência a relação produtiva entre informática e os estudos literários: O laboratório *Plateforme Géomatique*⁷ (EHESS, Paris), o laboratório *Observatório da Vida Literária OBVIL*⁸ (Sorbonne, Paris) e a estrutura de pesquisa *HUMA-NUM*⁹.

⁷ <https://psigehess.hypotheses.org/>

⁸ <http://obvil.sorbonne-universite.site/>

⁹ <https://www.huma-num.fr/>

<inserir Fig8_Marcia>

Figura 8: Grafo conceitual relacionado com a palavra *vida* na obra de Euclides da Cunha.

Paulo Silva Pereira, diretor do curso de doutoramento em Literatura de Língua Portuguesa e professor também na área de Humanidades Digitais na Universidade de Coimbra, apresentou suas reflexões sobre a área, sua natureza e alcance, pistas sobre teorias e práticas que deveríamos levar em consideração enquanto comunidade (embrionária), numa abordagem comparativa e transnacional. Partiram dele as propostas mais concretas de organização coletiva para o futuro próximo, que serão descritas na secção seguinte.

Isabel Araújo Branco, Diana Santos, Raquel Amaro e Paulo Silva Pereira apresentaram a história da iniciativa COST, “Distant reading for European Literary History”,¹⁰ incluindo um panorama do trabalho realizado até à data no âmbito dessa iniciativa e um inquérito sobre leitura distante feito à comunidade portuguesa dos estudos literários.¹¹ Discutiram também algumas das respostas obtidas, tanto de aceitação como de rejeição desse paradigma, oferecendo uma imagem realista da divulgação do tema nos meios académicos portugueses atualmente.

Ainda relacionado com a ação COST, os pesquisadores Rebeca Schumacher Fuão e João Marques Lopes discursaram sobre seu projeto, desencadeado na Literateca por conta da compilação de um conjunto significativo de obras publicadas entre 1840 e 1919, assim como as dificuldades que encontraram no processo de escolha e classificação das quase duzentas obras do período (cf.

10 Mais informações em <https://www.distant-reading.net/> e <https://www.cost.eu/actions/CA16204/#tabs|Name:overview>

11 Passível de ser consultado em https://docs.google.com/forms/d/1-75ASiir91Pv2u-7NfbEI40gDLgYo-wB384wufZyO3Q/viewform?edit_requested=true

Santos et al., 2020). A discussão do trabalho também terminou por levantar a questão da insuficiência na representatividade de mulheres escritoras dentro das definições históricas de “escolas literárias” luso-brasileiras. Paulo Silva Pereira sugeriu, assim, que um “novo cânone” poderia dar origem a novas taxonomias (veja-se também Silva, 2014).

A comunicação de Diana Santos (adaptada de Santos e Simões, 2019) dedicou-se à apresentação e ilustração da Literateca (Santos, 2019) como um ambiente ou infraestrutura para estudar a literatura lusófona, destacando vários exemplos de estudos embrionários no campo: diferenças entre autoras e autores, referência a emoções nas literaturas brasileira e portuguesa, estudos no campo da saúde e do papel do médico na literatura, entre outros. A ênfase não foi nos resultados, mas nas potencialidades.

<inserir Fig9_Literateca>

Figura 9: Redes de personagens relativas a duas obras da literatura portuguesa do século XIX: *As Pupilas do Senhor Reitor*, de Júlio Dinis, e *O Primo Basílio*, de Eça de Queirós.

A palestra de Barbara Ramos concentrou-se no estudo e anotação das emoções em textos de português no âmbito da Literateca, agregando minúcias importantes e ampliando a discussão sobre a complexa tarefa de pesquisa do perfil emocional, tanto na língua quanto na literatura. Relatando o andamento de sua tese na PUC-Rio, Barbara abordou três questões centrais: a identificação de novos termos de sentimento com base no uso da língua (Ramos & Freitas, 2019), a relação da emoção com o corpo (Ramos et al., 2020) e o agrupamento das palavras de emoção em categorias mais amplas.

Temas de discussão

Um tema frequentemente abordado no evento foi a definição e delimitação da leitura distante. Por um lado, há exemplos que apontam para a existência de várias tecnologias e práticas, tanto no estudo da literatura quanto da linguística, que poderiam ser consideradas precursoras desta metodologia, como a estilometria citada na apresentação de Marcia Langfeldt, assim como a menção de outra proposta recente holandesa de estudar os hábitos de leitura de uma população nacional de leitores (Riddel, Allen & Karina van Dalen-Oskam, 2018; van Craneburg et al., 2019). Por outro lado, houve argumentos a favor de uma delimitação mais restrita do campo, sugerindo que uma priorização acertada talvez permita um avanço mais produtivo nos trabalhos da área. De toda forma, houve consenso geral a respeito da necessidade de um esforço coletivo de integração para ampliar o conhecimento sobre tudo que está sendo pesquisado e desenvolvido nessa área dentro da língua portuguesa. De toda maneira, é importante ressaltar alguns pressupostos e algumas possíveis perspectivas de avanço nessa área. Cabe, assim, dizer que o que Franco Moretti propôs inicialmente como leitura distante não se restringe ao exame de grandes quantidades de obras, publicadas em largos intervalos de tempo. É possível trabalhar com grandes quantidades de dados a partir de uma única obra literária (como o fez também o próprio Moretti), de maneira que a leitura distante diz respeito, principalmente, à exploração de uma metodologia de leitura e de análise que vem se associar à leitura tradicional ou, se quiserem, à leitura próxima. O levantamento de dados permite analisar a obra sob perspectivas que esta leitura nunca poderia vislumbrar e provocá-la a outras possibilidades de análise e de interpretação. De outro lado, a leitura próxima pode muito bem ser utilizada para referendar ou invalidar alguns levantamentos quantitativos que poderão nunca deixar de ser apenas isso, por não conseguirem se incorporar de modo fecundo ao exame da obra.

Foi também discutida a questão da diferença em termos de quantidade ou volume daquilo que geralmente se caracteriza como leitura distante e os trabalhos inspirados em linguística com corpos, que correspondem a uma dimensão bastante menor. Será preciso muito cuidado na preservação de um diálogo aberto entre as duas metodologias que presumidamente cumprem os mesmos objetivos ou têm os mesmos requisitos. As apresentações de Cláudia Freitas, Diana Santos, Suemi Higuchi e Barbara Ramos usaram a anotação humana total ou parcial, estratégia incomum até ao momento dentro do campo da leitura distante “tradicional”. Uma preocupação seria equacionar até que ponto a revisão da anotação pode ser incluída neste processo, ainda que este tipo de revisão não seja tradicionalmente considerado na leitura distante. Como argumentou Claudia Freitas, seus resultados (a anotação) podem adicionar mais uma camada a este tipo de leitura, pois ao trabalhar com padrões observáveis a partir das quantidades e em material textual não anotado, acabamos inevitavelmente prestando mais atenção naquilo que parece “mais diretamente” observável. No entanto, quando invocamos propriedades de distribuição de fenômenos – sejam estes linguísticos ou não – somos forçados a constatar a presença constante de uma cauda longa destas ocorrências e outras singularidades que, tomados como um todo, representam uma imensa parcela do material em questão (Santos, 2014; Freitas, 2017 e Manning & Schutze, 1999). Para Cláudia Freitas, é justamente nesse contexto que a anotação – atividade classificatória por excelência – pode atuar como um auxiliar poderoso, na medida em que procura atribuir sentidos mais gerais às ocorrências únicas, viabilizando a incorporação destas no fluxo geral dos dados. Da mesma forma, tal perspectiva revela a anotação como tarefa capaz de amenizar críticas que veem a leitura distante como uma abordagem meramente quantitativa. Tomando-se as classes e o processo de anotação como atividades interpretativas, trata-se de uma

maneira de concatenar a dimensão qualitativa à quantitativa, fazendo uso do melhor dos dois mundos.

Outra questão relevante, sobretudo levantada por Paulo Silva Pereira, seria a do estabelecimento de uma base comum em torno de objetivos particulares para potencializar uma colaboração mais extensa e profícua, designadamente no âmbito da criação de recursos comuns. Dadas as dificuldades no trabalho de digitalização de grandes quantidades de texto, seria útil delimitar campos mais específicos para a utilização de ferramentas computacionais de análise.

Duas hipóteses de trabalho que poderiam ser exploradas dizem respeito, por um lado, a novas abordagens do romance histórico produzido em Portugal e no Brasil, no século XIX e primeiras décadas do seguinte, tanto numa perspectiva individualizada (o caso português e o caso brasileiro), como de índole comparativa e, por outro, à escrita de autoria feminina. É indiscutível a vivacidade de obras literárias de pendor histórico no âmbito da literatura portuguesa, como se pode comprovar pelo volume de poemas, romances e dramas históricos publicados nos dois últimos séculos (Chaves, 1980; Marinho, 1999; Boechat et al., 2000). Inicialmente, sua função consistia em consolidar uma certa ideia de “Nação” através do fundo do medievalismo, no caso de Portugal, e do indianismo, no caso do Brasil. A partir de meados do século XIX, os autores começaram a incorporar novos cenários narrativos, incluindo matéria contemporânea à data. Os eventos mais significativos do século XIX (desde o período das Invasões napoleónicas ao estabelecimento da República em 1910, em Portugal; da Independência à República e à Revolução de 30, no Brasil) podem ser reconstruídos a partir desta cadeia diacrónica de produção literária. Seria interessante descobrir o que as novas ferramentas de leitura distante permitiriam identificar nesse largo *corpus*. Nesse sentido, talvez se possa apontar um dos resultados do já

referido projeto *Atlas das Paisagens Literárias*, no qual se procurou uma análise dos olhares literários sobre um evento histórico em concreto, a Revolução do 25 de Abril (Queiroz e Alves, 2015). Apesar de não se enquadrar na cronologia acima proposta, este exemplo tem a vantagem de se referir ao mesmo enquadramento temático e de apresentar uma via de conciliação ou complementaridade entre os métodos da leitura próxima e a leitura distante.

Quanto à escrita de autoria feminina, o desafio ainda é mais aliciante. Contrariando a ideia de que a produção literária por mulheres no contexto português seria escassa ou de sofrível qualidade, vários trabalhos e projetos de pesquisa têm procurado resgatar esse acervo pouco conhecido, nas últimas décadas (Barros, 1924; Osório & Esteves, 2005; Schumacher & Vital Brazil, 2001; Macedo & Amaral, 2005).¹² Afirmar que a história da literatura se encontra fortemente marcada pelo conceito masculino de autoria é um mero truísmo, mas não se trata aqui de explorar o alcance dos mecanismos institucionais que estão por detrás desse fenómeno de exclusão ou de marginalização das mulheres escritoras, e conseqüente edificação de um cânone esmagadoramente de autoria masculina (Erkkila, 1992; Rector, 1999; Edfeldt, 2006; Silva, 2014; Lousada & Cantarin, 2016). A constituição de um *corpus* formado simultaneamente por obras de autoria feminina e masculina em regime de paridade (ou, pelo menos, de representação mais equilibrada entre homens e mulheres), como acontece no âmbito da ação COST, constitui uma oportunidade única de se estudar o fenómeno literário sem o apagamento histórico da produção de mulheres. Tradicionalmente, tem sido difícil relacionar o trabalho de mulheres escritoras com determinadas categorias de classificação que servem de suporte a boa parte das histórias literárias, como as de movimentos e gerações literárias (Spender, 1997; Owen & Pazos Alonso, 2011), pelo que é interessante ver as conseqüências que daí resultam. Por outro lado, torna-se

¹² Veja-se, a título de exemplo, o projeto *Escritoras. Women Writers in Portuguese before 1900* (<http://www.escritoras-em-portugues.eu/>).

possível analisar sob outro ponto de vista – agora com base em evidências empíricas – certos lugares comuns associados à prosa feminina que sublinham a sua exacerbada sentimentalidade, a sua condição menor face aos autores considerados hiper-canónicos e a incapacidade de expressão artística relevante (Santos & Amaral, 1997; Klobucka, 2008). Além disso, as ferramentas estilométricas podem abrir novas perspetivas em termos de estudo do discurso a partir de uma perspetiva de género. Por exemplo, tentar perceber se as convenções de género (literário) podem ter alguma prevalência sobre os usos de género (sexual) ou se, do ponto de vista diacrónico, as diferenças aumentam ou diminuem. Nesse sentido, podemos dizer que traz um contributo válido para a discussão em torno da canonicidade, da memória cultural e da tradição literária em articulação com as questões de género. Em suma, como sempre acontece em projetos de leitura distante, a escala de análise interfere não só a nível da quantidade de textos que são objeto de tratamento numa dada pesquisa, mas também na formulação de novas questões, abrindo caminho para uma compreensão mais sistemática das convenções que regem as práticas literárias num determinado período histórico-cultural e no modo como certas obras canónicas se destacam face a um cenário de estereotipia semântica e formal. Interessa aqui, pois, considerar a história literária não como um conjunto exemplar de estudos de caso, mas enquanto sistema abrangente.

Outro tema de discussão no ELD foi se deveria haver um foco maior na divulgação do(s) método(s) de leitura distante na comunidade lusófona de estudos literários, ou se seria mais prudente no momento simplesmente reunir os esforços daqueles já interessados no assunto, havendo divergência de posições neste sentido. Outra questão trata de possíveis aproximações com outros grupos da grande área das Humanidades que também se utilizam dos mesmos métodos, como História e Ciências Sociais.

Uma das constatações do encontro foi a presença quase exclusiva de participantes das áreas de linguística e dos estudos literários, faltando quase absolutamente a terceira vertente, a informática. Esperamos poder mitigar esta discrepância em edições futuras, visto que a implementação e o teste dos métodos computacionais são essenciais. Julgamos ser também benéfico para o desenvolvimento desta abordagem procurar integrar em futuros eventos especialistas de outras áreas disciplinares, como a História, a Geografia e a Ciência Política, entre outras.

Outra questão abordada foi a da separação ou junção de esforços das comunidades brasileira e portuguesa. Embora em alguns casos esta união esteja acontecendo, (tanto o NuPILL como a Literateca reúnem obras das duas literaturas), existem também vozes e iniciativas que consideram mais fundamental o estudo independente de cada literatura e comunidade em si. Esta dicotomia é claramente visível na própria ação COST, cujo objetivo de comparar as histórias da literatura europeia põe alguns entraves ao estudo conjunto da literatura lusófona. A abordagem por ora seguida pelos participantes portugueses no COST é a de criar uma coleção portuguesa e outra lusófona em separado (ou seja, além da coleção com obras portuguesas, ter uma coleção alargada que inclui também obras brasileiras e outras obras que não cumpram os requisitos da coleção ELTEc-por, por exemplo mais obras de um mesmo autor português, ou obras mais curtas do que o limite do COST). Além disso, e embora uma ação COST seja restrita a parceiros europeus, conseguiu-se o estatuto de observador para o Brasil (PUC-Rio), o que significa que, embora não financiado, pode participar nas discussões e ter acesso a todas as iniciativas e dados.

Para os que se interessam pelo estudo da literatura lusófona, há projetos em curso ou em estágio embrionário cujo objetivo é exatamente investigar a influência e proximidade de alguns escritores brasileiros em relação a escritores portugueses (nomeadamente os poetas Alberto de Oliveira e Guerra Junqueiro, e o romancista Coelho Neto), assim como estudos de recepção de autores do outro lado do Atlântico e vice-versa. A esse respeito, foram apresentadas algumas curiosidades em relação à historiografia literária lusófona. O pesquisador João Marques Lopes observou que, no dia do nascimento de José Saramago, 16 de novembro de 1922, o *Diário de Notícias* (principal jornal português à época) apresentava em grandes parangonas um louvor aos maiores escritores vivos da língua portuguesa: Guerra Junqueiro, português, e Coelho Neto, brasileiro, ambos pouco lidos na atualidade.

Como reflexo da análise de leitura distante, Marcia Langfeldt também chamou atenção para a criação de topônimos com o nome do escritor Euclides da Cunha (como de uma cidade no Brasil e centenas de logradouros), além de outros escritores membros da Academia Brasileira de Letras. Esta iniciativa acabou por constituir a criação de um campo literário nacional implementado pela instituição, configurando um autêntico *lieu de mémoire* no qual se constrói uma relação dialética entre História e memória. Um exemplo interessante é o caso do poeta Manuel Bandeira, cujo nome está atribuído não apenas a mais de duzentos logradouros em todo o Brasil, mas também a uma rua da cidade do Porto, em Portugal, reverberando para além-fronteiras esta relação entre o espaço e a cultura de um país.

Outras questões abordadas no encontro foram as possibilidades de intervenções educativas entre faculdades e instituições distintas, de forma a partilhar experiências e conhecimentos. Para tal, bastariam, inicialmente acordos informais de palestras remotas, como o estabelecido entre a

UFSC, a UEMA, a PUC-Rio e a Universidade de Oslo (já informalmente iniciado) e a troca de formações e visitas cruzadas financiadas pelo projeto BILLIG entre a NOVA-FCSH e a Universidade de Oslo.

Discutiu-se também até que ponto outras abordagens nos estudos literários poderiam ter um diálogo frutífero com a leitura distante em português, sobretudo aquelas que usam metodologias empíricas e estatísticas, como as que analisam aspectos cognitivos e emoção na recepção literária (Miall & Kuiken, 1994; Miall, 2006; Fialho, 2012; Fialho et al., 2012) ou a recepção literária de forma mais abrangente (Cf. publicações no journal *Scientific Study of Literature*; Burke, Fialho & Zyngier, 2016), já com um percurso sólido em estudos em língua inglesa, apresentadas por Olívia Fialho. A questão da delimitação dos estudos do que é a leitura distante em contraponto com outras abordagens quantitativas nos estudos literários é pertinente em geral, e não só em relação ao português, mas, se a comunidade lusófona conseguir integrar os dois tipos de estudos em vez de os considerar em oposição, isso poderá ser um excelente objetivo, sobretudo visto que não conhecemos ainda quaisquer trabalhos nessas áreas sobre o português.

Paula Terra chamou a atenção para a importância da constituição de um espaço de pesquisa com ampla representatividade discursiva, valorizando a inclusão de vozes minoritárias, para que tais pesquisadores possam exercer maior protagonismo acerca dos temas e questões de seu interesse a partir de seus próprios lugares de fala (sejam esses marcadores de gênero, classe, raça, orientação sexual etc).

A necessidade de considerar e enfatizar a leitura distante como complementar, e não substituta, da leitura próxima e de quaisquer outros métodos pertinentes no estudo da literatura também foi

um assunto recorrente. Houve um consentimento coletivo de que a disseminação desta área deve sempre enfatizar a complementaridade entre os dois campos teóricos. No projecto “Atlas das Paisagens Literárias”, por exemplo, tem sido reforçada essa mensagem, uma vez que não parece possível abdicar da riqueza e subjectividade de uma leitura próxima, tendo a leitura distante tudo a beneficiar dessa iteração (Alves e Queiroz 2015).

Por fim, discutiu-se bastante sobre a preocupação com o trabalho envolvido na preparação de acervos para a leitura distante. A passagem de documentos em formato pdf, muitas vezes como imagens, para o formato de texto, está longe de ser uma tarefa simples. Os métodos automáticos capazes de produzir material de qualidade ainda não são uma realidade, e o trabalho humano necessário para a manutenção e melhoria de acervos textuais, apesar de especializado, não é, tradicionalmente, reconhecido como atividade de pesquisa.

Conclusão

O encontro não teve outras conclusões que não fossem a necessidade de prosseguir o trabalho e de conjugar esforços para que mais estudiosos ou interessados possam contribuir para resultados que beneficiem todos, e que só poderão ser obtidos através de cooperação. A posteriori, podemos considerar que o termos conseguido escrever este documento a trinta e duas mãos foi a primeira prova de que existe uma comunidade empenhada em continuar ou abraçar este campo de pesquisa e colocar a leitura distante na ordem de trabalhos dos estudos literários lusófonos, não como única ou principal metodologia, mas como mais uma ferramenta de análise no campo das Humanidades Digitais.

Agradecimentos

Emanoel Pires agradece à FAPEMA; Danielle Sanches, Daniel Alves e Diana Santos aos EEA grants, que financiaram o projeto BILLIG; Isabel Araújo Branco, Diana Santos, Paulo Silva Pereira e Raquel Amaro agradecem à ação COST *Distant Reading for European Literary History*, financiada pelo Horizon 2020 e pela União Europeia, e Cláudia Freitas agradece à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) pelo financiamento no âmbito do projeto PRINT. Para os trabalhos usando o material da Literateca, agradecemos à FCCN o alojamento dos servidores, e declaramos que as computações efetuadas utilizaram também os recursos da *UNINETT Sigma2 - the National Infrastructure for High Performance Computing and Data Storage in Norway*.

Referências

[Todos os URL foram verificados a 4 de março de 2020]

ALVES, Daniel & Ana Isabel Queiroz (2013). “Studying Urban Space and Literary Representations Using GIS: Lisbon, Portugal, 1852-2009”. *Social Science History* 37.4: 457–81. DOI: <https://doi.org/10.1215/01455532-2346861>.

— (2015). “Exploring Literary Landscapes: From Texts to Spatiotemporal Analysis through Collaborative Work and GIS”. *International Journal of Humanities and Arts Computing* 9.1: 57–73. DOI: <https://doi.org/10.3366/ijhac.2015.0138>.

BARROS, Thereza Leitão de (1924). *Escritoras de Portugal: Génio Feminino Revelado na Literatura Portuguesa*. 2 vols. Lisboa: ed. de autora.

BODE, Katherine (2012). *Reading by Numbers: recalibrating the Literary Field*. London & New York: Anthem Press.

— (2017). “A response to some responses”. blogue, 28 de junho de 2017. <https://katherinebode.wordpress.com/a-response-to-some-responses/>

— (2018). *A World of Fiction: Digital Collections and the Future of Literary History*. Ann Arbor: University of Michigan Press.

BOECHAT, Maria Cecília Bruzzi, Silvana Maria Pessoa Oliveira & Paulo Motta Oliveira (2000). *Romance histórico: recorrências e transformações*. Belo Horizonte: FALE/UFMG.

CABRAL, Maria de Jesus, Maria Hermínia Amado Laurel & Franc Schuerewegen (2014). *Lire de près, de loin: close vs distant reading*. Paris: Garnier.

CHAVES, Castelo Branco (1980). *O Romance Histórico no Romantismo Português*. Amadora: Bertrand.

van CRANENBURGH, Andreas, Karina van Dalen-Oskam & Joris van Zundert (2019). "Vector space explorations of literary language". *Language Resources and Evaluation* 53: 625-650.

EDFELDT, Chararina (2006). *Uma história na história: representações da autoria feminina na História da literatura portuguesa do século XX*. Montijo: Câmara Municipal do Montijo.

ERKKILA, Betsy (1992). *The Wicked Sisters: Women Poets, Literary History & Discord*. Oxford: Oxford University Press.

FIALHO, Olivia, Sonia Zyngier & David Miall (2011). "Interpretation and Experience: Two Pedagogical Interventions Observed". *English in Education* 45.3: 236-253.

FREITAS, Bianca, Cláudia Freitas & Diana Santos (2016). "QUEMDISSE?: Reported speech in Portuguese". *Proceedings of the Tenth International Conference on Language Resources and*

Evaluation (LREC 2016). Ed. Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk & Stelios Piperidis . 4410-4416.

FREITAS, Cláudia (2017). “Estudos linguísticos e Humanidades Digitais: corpus e descorporificação”. *Gragoatá* 22.44: 1207-1227.

FREITAS, Cláudia, Elvis de Souza & Luísa Rocha (2019). "Quantificando e qualificando o sujeito oculto em Português". *STIL 2019 – XII Symposium in Information and Human Language Technology and Collocates Events , October 15-18, 2019, Salvador, BA, Proceedings of conference*. 288-293.

HIGUCHI, Suemi, Diana Santos, Cláudia Freitas & Alexandre Rademaker (2019). "Distant reading Brazilian history". *Proceedings of the Digital Humanities in the Nordic Countries 4th Conference (Copenhagen, Denmark, March 5-8, 2019)*. Ed. Constanza Navarreta, Manex Agirrezabal & Bente Maegard. 190-200.

JOCKERS, Matthew L (2013). *Macroanalysis: Digital Methods and Literary History*. Illinois: University of Illinois Press.

KLOBUCKA, Anna (2008). “Sobre a hipótese de uma herstory da literatura portuguesa”. *Veredas: Revista Internacional da Associação de Lusitanistas* 10: 13-25.

LANGFELDT, Marcia Caetano (2018). *A Amazônia e os impasses da civilização em relatos dos séculos XX e XXI*. Université Sorbonne Paris Cité.
<https://tel.archives-ouvertes.fr/tel-02178877/document>

LOUSADA, Isabel Cruz & Márcio Matiassi Cantarin (coord.) (2016). *Do feminino plural ou a singularidade pela voz e a escrita de mulheres*. Lisboa & São Paulo: CLEPUL/UNESP.

MACEDO, Ana Gabriela Macedo & Ana Luísa Amaral (orgs.) (2005). *Dicionário da Crítica Feminista*. Porto: Edições Afrontamento.

MANNING Christopher C. & Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing*. Cambridge, Massachusetts: The MIT Press.

MARINHO, Maria de Fátima (1999). *O Romance Histórico em Portugal*. Porto: Campo das Letras.

MITTMANN, Adiel, Roberto Willrich & Renato Fileto (2013). "DLNotes2: Ferramenta de anotações estruturadas e semânticas voltada ao ensino da literatura". *Escritorios electrónicos para las literaturas: nuevas herramientas digitales para la anotación colaborativa*. Ed. L. P. Núñez. Madrid: Universidad Complutense de Madrid. 137-152.

MITTMANN, Adiel, Aldo von Wangenheim & Alckmar Luiz dos Santos (2016). "A System for the Automatic Scansion of Poetry Written in Portuguese". *Computational Linguistics and Intelligent Text Processing, 17th International Conference, CICLing 2016, Konya, Turkey, April 3-9, 2016, Revised Selected Papers, Part II*. Springer. 611-628.

MOREIRA, Nazilda M. de B. (2002). "Da margem para o centro: a autoria feminina e o discurso feminista do século XIX". *Gênero e representação: teoria, história e crítica*. Ed. Constância Lima Duarte, Eduardo de Assis Duarte & Kátia da Costa Bezerra. Belo Horizonte: Faculdade de Letras da UFMG. 143-147.

MORETTI, Franco (2005). *Graphs, Maps, Trees: Abstract Models for Literary History*. London: Verso Books.

— (2013). *Distant Reading*. London: Verso Books.

ODAG, Özen (2013). "Emotional engagement during literary reception: Do men and women differ?". *Cognition & Emotion*, 27.5: 856-874.

OSÓRIO, Zília & João Esteves (Dir.) (2005). *Dicionário no Feminino (séculos XIX-XX)*. Lisboa: Livros Horizonte.

OWEN, Hilary & Cláudia Pazos Alonso (2011). *Antigone's Daughters? Gender, genealogy, and the politics of authorship in 20th-Century Portuguese Women's Writing*. Lewisburg, PA: Bucknell University Press.

PIPER Andrew (2017). "Data, data, data. Why Katherine Bode's new piece is so important and why it gets so much wrong about the field". 23 de junho de 2017. [29 de fev. 2020].

<https://txtlab.org/2017/06/data-data-data-why-catherine-bodes-new-piece-is-so-important-and-why-it-gets-so-much-wrong-about-the-field>

— (2018). *Enumerations: Data and literary study*. Chicago: University of Chicago Press.

QUEIROZ, Ana Isabel & Daniel Alves (2015). “Walking Through the Revolution: A Spatial Reading of Literary Echoes”. *JSSE-Journal of Social Science Education* 14.2: 4-16.
<https://doi.org/10.2390/jsse-v14-i2-1351>

RAMOS, Barbara & Cláudia Freitas (2019). ““Sentimento de quê?”: uma lista de sentimentos para a Análise de Sentimentos”. *STIL 2019 – XII Symposium in Information and Human Language Technology and Collocates Events , October 15-18, 2019, Salvador, BA, Proceedings of conference*. 38-47.

RAMOS, Barbara, Diana Santos & Cláudia Freitas (2020). “Looking at body expressions to enrich emotion clusters”. *Proceedings of the Digital Humanities and Natural Language Processing Workshop at the 14th International Conference on the Computational Processing of Portuguese Language, 2020*. Ed. Maria José Bocorny Finatto, Saturnino Luz, Senja Pollak & Renata Vieira.

RECTOR, Monica (1999). *Mulher, objecto e sujeito da literatura portuguesa*. Porto: Edições Universidade Fernando Pessoa.

REIS, Carlos (2016). *Pessoas de Livro: Estudos sobre a Personagem*. 2. ed. Coimbra: Imprensa da Univ. de Coimbra.

RIDDEL, Allen & Karina van Dalen-Oskam (2018). "Readers and their roles: Evidence from readers of contemporary fiction in the Netherlands". *PLoS ONE* 13.7: e0201157. DOI: <https://doi.org/10.1371/journal.pone.0201157>

ROCKWELL, Geoffrey & Stéfan Sinclair (2016). *Hermeneutica: computer-assisted interpretation in the humanities*. Cambridge, MA: The MIT Press.

RYBICKI, Jan & Eder Maciej (2011). "Deeper Delta across Genres and Languages: Do We Really Need the Most Frequent Words?". *Literary and Linguistic Computing* 26.3: 315-321.

SANTOS, Diana (2014). "Podemos contar com as contas?". *New Language Technologies and Linguistic Research: A Two-way Road*. Ed. Sandra Aluísio & Stella Tagnin. Newcastle: Cambridge Scholars Publishing. 194-213.

— (2019). "Literature studies in *Literateca*: between digital humanities and corpus linguistics". *Humanists and the digital toolbox: In honour of Christian-Emil Smith Ore*. Ed. Martin Doerr, Øyvind Eide, Oddrun Grønvik & Bjørghild Kjelsvik. Oslo: Novus forlag. 89-109.

SANTOS, Diana & Alberto Simões (2019). "Towards a computational environment for studying literature in Portuguese". *DH Budapest 2019, Digital Humanities Conference* (Budapest, 25-27 September 2019).
<https://www.linguateca.pt/Diana/download/PresentationBudapestSantosSimoese.pdf>

SANTOS, Diana, Emanuel Pires, João Marques Lopes, Rebeca Schumacher Fuão & Cláudia Freitas (2020). "Periodização automática: Estudos linguístico-estatísticos de literatura lusófona". Em apreciação.

SANTOS, Maria Irene Ramalho dos & Ana Luísa Amaral (1997). *Sobre a 'Escrita Feminina'*. Coimbra: Centro de Estudos Sociais.

SCHÖCH, Christof et al. (2017). Memorandum of Understanding for the implementation of the COST Action Distant Reading for European Literary History CA16204, Technical Annex. COST – European Cooperation in Science and Technology.
https://e-services.cost.eu/files/domain_files/CA/Action_CA16204/mou/CA16204-e.pdf

SCHUMACHER, Shuma & Érico Vital Brazil (orgs.) (2001). *Dicionário Mulheres do Brasil de 1500 até Actualidade*. Rio de Janeiro: Jorge Zahar Editor. 2ª Edição.

SILVA, Fabio Mario da (2014). *A autoria feminina na literatura portuguesa. Reflexões sobre as teorias do Cânone*. Lisboa: Edições Colibri.

SPENDER, Dale (1997). "Women and Literary History". *The feminist reader*. Ed. Catherine Belsey & Jane Moore. Blackwell: Palgrave Macmillan. 16-25.

UNDERWOOD, Ted (2017). "A Genealogy of Distant Reading". *DHQ: Digital Humanities Quarterly* 11.2. <http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html>

WILKENS, Mathew (2015). "Digital Humanities and its Applications in the Study of Literature and Culture". *Comparative Literature* 67.1: 11-20.