



## Humanidades digitais e estudos históricos do léxico Digital Humanities and Historical Lexicology

Maria José Bocorny FINATTO\*

### Um cenário para novos encontros: as Humanidades Digitais

No contexto de uma nova *transdisciplina* (Golub, Liu, 2021) que passamos a conhecer por *Humanidades Digitais (HD)*, nós, como linguistas, (re)encontramos o cenário-desafio de extrair dados e conhecimentos dispersos a partir de arquivos digitais que representam registros de línguas escritas ou faladas. Nesse cenário, está o estudo dos textos e dos discursos, mas em novos formatos, geralmente bastante informatizados.

Nas *HD*, pesquisa-se sobre as heranças e os patrimônios, culturais, científicos e artísticos, incluindo os patrimônios em forma de imagens e em outros suportes, sendo conduzidos trabalhos, via de regra, em miradas sócio-históricas e diacrônicas. Nessas miradas, normalmente, aplicam-se recursos computacionais ao tratamento de dados e de acervos digitalizados, o que tende a ajudar a sua transformação em informação qualificada e devidamente segmentada. Assim, os todos de massas de dados digitalizados transformam-se em partes que podem ser armazenadas e buscadas conforme diferentes demandas e interesses de recuperação de informação.

Quanto mencionam-se as *HD*, geralmente encontramos colegas da História, da Documentação, da Arquivologia, dos Estudos do Patrimônio e da Computação à frente de uma série de iniciativas e de eventos. Felizmente, cada vez mais estudiosos da linguagem têm participado desses encontros e debates em torno da informação digital. Igualmente, no cenário da Educação, uma série de discussões têm sido feitas sobre os desafios do novo quadro das *HD* conforme ele se apresenta (Rocha; Costa,

---

\* Doutora em Letras/Estudos da Linguagem pela UFRGS, Docente do PPG-Letras-UFRGS, pesquisadora PQ do CNPq. [maria.finatto@gmail.com](mailto:maria.finatto@gmail.com)

2021), como está sendo percebido e como pode ser explorado para o bem comum. Direitos autorais, aspectos éticos e propriedade dessas fontes digitalizadas também têm sido muito discutidos. Desse modo, à medida que os registros, acervos, *corpora* e todos os rastros das Humanidades, salvaguardados, recortados, ampliados, reproduzidos ou representados em diferentes formatos de arquivo interessam às HD, instaura-se algo a ser compreendido, especialmente no tocante à inserção digital e à cidadania.

No Brasil, tivemos a nossa *I Jornada Nacional de Humanidades Digitais* em outubro de 2020, em um evento *on-line*<sup>1</sup>. Na ocasião, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) tomou a iniciativa de apresentar e dinamizar o novo campo e trouxe a oportunidade do debate. A interdisciplinaridade na produção, salvaguarda e pesquisa de acervos digitais foi uma das tônicas desse evento. Mais tarde, em 2022, tivemos um outro evento multidisciplinar associado ao tema das HD, junto ao Arquivo Público do Estado do RS, o *Linguística, História e Computação: dos acervos aos conhecimentos e vice-versa*, com mais de 400 participantes<sup>2</sup>, em atividades remotas e presenciais. No Brasil, desse modo, avançamos no debate dos temas das conservações e digitalizações dos patrimônios, de diferentes formatos e tipologias.

No estrangeiro, o assunto das HD, igualmente, tem mobilizado diferentes grupos institucionais e acadêmicos, com uma acolhida “natural” na pesquisa sobre Recuperação da Informação (*Information Retrieval*), reengenharia de *software* e afins. Afinal, há toda uma base computacional em torno de uma história de dados produzidos, compartilhados e conservados em um ambiente virtual. Para um exemplo disso, temos já, há três edições, junto ao PROPOR, o maior evento sobre o Processamento do Português na área da Informática, *workshops* específicos sobre HD e

---

<sup>1</sup> Disponível em: <https://plataforma9.com/congressos/i-jornada-nacional-de-humanidades-digitais> . Acesso em: 23 dez. 2023.

<sup>2</sup> Disponível em: <https://sites.google.com/view/linguistic-histor-e-computacao/p%C3%A1gina-inicial> Acesso em: 23 dez. 2023.

Processamento da Linguagem Natural (PLN), o *DHandNLP*<sup>3</sup>. Nessa via, assistimos também ao estabelecimento, em nível acadêmico, de diferentes *Laboratórios de Humanidades Digitais*, como já temos na Universidade Nova de Lisboa<sup>4</sup> e na Universidade de Évora<sup>5</sup>, em Portugal, para citar apenas alguns. Nesse âmbito, também a conservação e salvaguarda de informações e de dados de estudos e de pesquisas, produzidos e lançados apenas na internet, é um tema recorrente, visto que tem sido comum a obsolescência de plataformas e de sistemas, que precisam ser revitalizados para seguirem ativos. Enfim, especialmente no período da Pandemia, quando justamente mais buscamos históricos e dados *on-line*, assistimos à “morte” e ao desaparecimento (repentino?) de diferentes *sites* de pesquisa acadêmica. Isso serviu de alerta para toda uma comunidade que se ocupa desses modernos patrimônios digitais, que também podem ficar “velhos” e inoperantes, sobretudo se utilizam tecnologias comerciais não atualizáveis. Enfim, descobrimos os dramas das perdas de milhões de dados, produzidos a duras penas, em décadas de trabalho, geralmente com verbas de pesquisa pública (para o breve relato de uma experiência, veja Finatto, 2022).

Nesse cenário “da informação maciçamente digitalizada” ou um tanto distante dele, na área de Letras, já temos, de longa data, toda uma tradição de trabalhos linguísticos, lexicológicos, lexicográficos e especialmente os estudos filológicos, dedicados a estudar e situar documentos escritos antigos, impressos, manuscritos e registros de fala transcritos. Sem medo de errar, posso dizer que já éramos digitais

---

<sup>3</sup> Para conhecer os trabalhos apresentados nas diferentes edições do *DHandNLP*, veja: <https://propor2024.citius.gal/index.php/iii-workshop-on-digital-humanities-and-natural-language-processing/>

<sup>4</sup> Este laboratório de *HD* estabeleceu-se, formalmente, em 2019. Seu diferencial, conforme entendo, é o foco em diferentes saberes e metodologias da área das Artes e das Humanidades (história, arte, literatura, arqueologia, música, geografia etc.) e das Ciências Computacionais. Disponível em: <https://dhlab.fcsh.unl.pt>

<sup>5</sup> Fundado em 2022, o Laboratório de Humanidades Digitais do CIDEHUS, reúne pesquisadores de diferentes especialidades, como linguistas, historiadores da ciência e especialistas em Computação. Está conectado ao CIDEHUS Digital, é um repositório de dados em acesso aberto, que oferece um crescente volume de materiais, incentivando a sua reutilização. Disponível em: <http://www.cidehusdigital.uevora.pt>

desde sempre, especialmente desde quando passamos a utilizar apoio computacional para a produção e guarda e compartilhamento de *corpora* textuais.

A antiguidade das fontes quando se lida com as *HD*, naturalmente, é sempre uma condição relativa, pois os pontos de referência podem ser variados: um passado em exame pode ser tanto a década de 1980 quanto o século XVIII, por exemplo. Desse modo, com a digitalização de acervos (que também é um processo que demanda cuidados e profissionais com formação especializada), cada vez mais divulgam-se trabalhos de Paleografia, de edição e crítica textual, cujas fontes são ou tornaram-se disponíveis em forma de arquivos, felizmente a maioria com acesso público.

Assim, dinamizou-se, especialmente no Brasil, já há mais de uma década, uma “Filologia Digital” (Paixão de Sousa, 2013), âmbito que nos permitiu o acesso a todo um patrimônio de manuscritos raros (Rocha; Lose; Barreiros, 2020) e a fontes impressas pouco acessíveis (Murakawa, Nadin, 2020), em diferentes diacronias, registros e contextos culturais (Castro, 2013), que nos mostraram toda uma outra História Social. Assim, os livros, as bibliotecas e os arquivos, digitalizados, tornaram-se novos objetos de consumo (Guerreiro; Borbinha, 2016), instaurando-se também novas relações das pessoas com os acervos físicos tradicionais. Ao passo que acessamos obras antigas apenas pela via *on-line*, nos perguntamos: quem ainda frequentará essas bibliotecas físicas? Quem cuidará da segurança desses materiais originais tão preciosos?

Nesse processo, parafraseando um dos artigos da nossa seção temática, vale dizer que se acentua, pela via das *HD*, todo o fluxo de uma “libertação de conteúdos”, em vários aspectos. Afinal, saímos das fontes físicas dos arquivos em concreto, espacial e geograficamente limitados, e a partir delas encontramos arquivos transformados em *corpus/corpora*, em suportes que nos permitem uma leitura a distância de diferentes tipos de fontes e de acervos. Todavia, vale lembrar, para chegar nesse ponto, houve todo um trabalho de coleta e curadoria de dados em museus, arquivos públicos e bibliotecas. Além disso, muitas vezes, esquecemos que o que hoje acessamos no

conforto de alguns cliques são fontes exaustivamente garimpadas, tratadas e digitalizados pelos próprios investigadores, com seus próprios recursos e equipes. E, nesse processo, ainda carecemos de políticas de acesso ou de disponibilização do material arquivístico em formatos *on-line*, sem contar a carência de políticas para gestão, promoção e salvaguarda dos próprios arquivos físicos – e da valorização do pessoal que neles trabalha. Vale lembrar: sem os arquivos, acervos e bibliotecas físicas e sem pessoas que deles cuidam nada haveria.

Nesse resgate do concreto ao suporte virtual, via digitalização, há ainda muitas perguntas importantes que precisamos repetir, buscando alguma ponderação: quem gere e cuida desses acervos, sobretudo dos acervos em formato físico? O que merece ou (mais) precisaria ser digitalizado? O que são as prioridades? A digitalização estará ao lado da fonte original, que se poderá encontrar sempre que for preciso? Qual o papel institucional e político dos arquivos e acervos públicos nesse processo?

Conforme mencionado, as *DH* têm se situado na interseção da Informação, da Computação e das Humanidades. Com dados digitais em cronologias, temos pesquisa trans- e multidisciplinar e colaborativa, novas formas de ensino e pesquisa, nas suas realidades presenciais e *on-line*. Novos formatos e modalidades, em tese mais acessíveis, impactam o patrimônio cultural e desenharam uma “cultura digital”. Portanto, há um importante campo de pesquisa em crescimento. Um campo para o qual os estudos da linguagem e as Ciências do Léxico, especialmente se conjugados com as tecnologias do Processamento de Linguagem Natural (PLN), da Linguística de *Corpus*, da Filologia, da Paleografia e da Linguística Histórica, têm muito a oferecer. Afinal, existem muitas técnicas de pesquisa linguística, feitas com algum, pouco ou muito apoio computacional, que podem fornecer ajuda útil para o estudo de textos e registros de valor histórico, de literaturas e de qualquer outro discurso humano que esteja sob investigação nessas condições.

## Sobre os textos da seção Humanidades Digitais e estudos históricos do léxico

Situado o âmbito das *HD*, vale agora destacar os artigos que trazemos para os nossos leitores e leitoras. Nossa seção inicia com o artigo *Desambiguação do vocábulo jogo a partir de uma análise semântica histórica e o atual contexto de ubiquidade do entretenimento*, de Leonardo Dias Avanço. A ideia do trabalho, a partir de estudos da linguagem, é situar os valores da palavra jogo em um *continuum* histórico enquanto se analisam os processos que tornam a palavra ambígua.

O segundo trabalho é *A conquista do oeste brasileiro: estudos exploratórios sobre toponímia histórica*, de Marcia Meurer. Este trabalho demonstra como, com o acesso facilitado a diferentes bancos de dados, destacando materiais raros digitalizados, como fontes de cartografia antiga de difícil acesso e leitura, é possível realizar estudos contextualizados e interdisciplinares centralizados no léxico toponímico.

O terceiro artigo da seção é *Análises estatístico-computacionais de atribuição de autoria: Augusto dos Anjos e a obra psicografada Parnaso de Além-Túmulo*, produzido por Ana Paula Nunes de Sousa e Emanuel Pires. Nesse trabalho, vemos textos transformados em *corpora* e técnicas computacionais a eles aplicados para tentar identificar padrões de escrita e traços de uma autoria. Em um estudo de lexicometria, mostra-se o potencial da ferramenta digital *Hyperbase*.

O quarto artigo do grupo é *A terminologia têxtil n'O Jornal das Senhoras (1852 – 1855): os termos têxteis na imprensa feminina do século XIX*, cujas autoras são Ana Vitória Gomes Moreira e Vanessa Regina Duarte Xavier. Com essa pesquisa, ambas buscam dar a conhecer uma parcela do léxico relacionado a tecidos nesse gênero jornalístico e no espaço temporal do século XIX enquanto nos trazem uma série de contrapontos com registros lexicográficos e com glossários especializados.

O quinto trabalho da nossa seção intitula-se *Sociolinguística Histórica, Filologia e Humanidades Digitais: um olhar interdisciplinar*, da autoria de Evelyne Patricia Figueiredo de Sousa Costa e de Leici Lanherr Moreira. Este artigo nos traz, além de importantes reflexões sobre as *HD* e ao acesso à pesquisa sócio-histórica e sobre o

trabalho com acervos, a proposta de (re)construir o perfil sociocultural do primeiro secretário da capitania do Rio Grande de São Pedro. Isso é feito analisando-se a documentação indireta e as variações gráficas em manuscritos oitocentistas à luz da Sociolinguística Histórica, em suas relações com a Filologia e as *HD*.

Para enfeixar a seção, o sexto artigo é *Conceptualizações de atleta paraolímpico no contexto brasileiro: um estudo diacrônico baseado em frames*, de Aline Nardes dos Santos, Diego Spader de Souza e Rove Chishman. O objetivo geral do trabalho é verificar a presença de *frames* semânticos que caracterizam o atleta paraolímpico em uma perspectiva diacrônica. O trabalho parte de um *corpus* de notícias, disponível em uma *hemeroteca digital*, e investiga registros linguísticos que abrangem as décadas de 1960 a 2000, no contexto brasileiro.

Na condição de uma entusiasmada pelo tema das *HD* em perspectivas sócio-histórica e diacrônicas, ao mesmo tempo em que testemunho um conjunto de esforços sempre renovados no âmbito dos Estudos e das Ciências do Léxico, desejo que os textos da nossa seção temática nos apontem novos rumos e que possam nos inspirar para novos estudos e novos resgates. Apesar de todas as dificuldades, seguimos mostrando o vigor da ciência linguística brasileira e marcando o quanto o investimento para os acervos e arquivos públicos, físicos e digitais, é vital. Sem cuidar do nosso patrimônio, em suas diferentes modalidades e suportes, não teremos herança, nem memória, tampouco um futuro para estimar. Por fim, fica o agradecimento à *Domínios de Lingu@gem*, pela acolhida desta seção temática, e a todos os pareceristas e autores que aceitaram o desafio.

## Referências

CASTRO, I. **O retorno à Filologia**. Miscelânea de Estudos Linguísticos, Filológicos e Literários *in memoriam* Celso Cunha. Rio de Janeiro: Nova Fronteira, 1995.

FINATTO, M. J. B. Para uma história de *sites* de projetos de pesquisa da UFRGS. Porto Alegre: **Jornal da Universidade**. Disponível em: <https://www.ufrgs.br/jornal/para-uma-historia-de-sites-de-projetos-de-pesquisa-da-ufrgs/> Acesso em: 28 de dez. 2023.

GOLUB, K.; LIU, Y.-H. (ed.) **Information and Knowledge Organisation in Digital Humanities: Global Perspectives** (1st ed.). London: Routledge, 2021. Disponível em: <https://doi.org/10.4324/9781003131816>. Acesso em: 28 de dez. 2023. DOI <https://doi.org/10.4324/9781003131816>

GUERREIRO, D.; BORBINHA, J. L. Humanidades Digitais: Novos desafios e oportunidades. **Revista Internacional del Libro**, Digitalización y Bibliotecas, v. 2, n. 2, 2016. DOI <https://doi.org/10.37467/gkarevdig.v2.779>

MURAKAWA, C. de A. A.; NADIN, O. L. De Males e Dores: a variação terminológica na denominação de doenças no Português do Brasil colonial. **Papéis**, Campo Grande, v. 24, Nº Especial, p. 289-320, 2020. Disponível em: <https://periodicos.ufms.br/index.php/papeis/article/view/12405>. Acesso em: 28 de dez. 2023.

PAIXÃO DE SOUZA, M. C. A Filologia Digital em Língua Portuguesa: alguns caminhos. In: GONÇALVES, M. F.; BANZA, A. P. (coord.). **Património Textual e Humanidades Digitais: da antiga à nova Filologia**. Évora: CIDEHUS, 2013. Disponível em: <https://books.openedition.org/cidehus/1089>. Acesso em: 23 dez. 2023. DOI <https://doi.org/10.4000/books.cidehus.1089>

ROCHA, L. M. B. M.; COSTA, C. J. de S. A. Compreensão das humanidades digitais nos domínios da educação para a formação docente. **Revista Intersaberes**, [S. l.], v. 16, n. 38, p. 766-796, 2021. Disponível em: <https://www.revistasuninter.com/intersaberes/index.php/revista/article/view/2107>. Acesso em: 23 dez. 2023. DOI <https://doi.org/10.22169/revint.v16i38.2107>

ROCHA, J. P.; LOSE, A. D.; BARREIROS, P. N. Filologia, Edição Digital e Estudos do Processo Criativo: Análise do Poema em Trovas “Quadras á minha Dôr”. **Manuscritica: Revista de Crítica Genética**, [S. l.], n. 41, 2020, p. 5-19. Disponível em: <https://www.revistas.usp.br/manuscritica/article/view/180128>. Acesso em: 23 dez. 2023. DOI <https://doi.org/10.11606/issn.2596-2477.i41p5-19>