

PUBLICACIONES DE LA
ASOCIACIÓN ARGENTINA
DE HUMANIDADES DIGITALES



Análise com Voyant Tools do primeiro número da Revista de Humanidades Digitais

Bruna LESSA

bruna.lessa@ufba.br

Universidade Federal da Bahia (Brasil)

A Revista de Humanidades Digitais (RDH), configura-se como um marco na organização e disseminação da comunicação científica sobre o campo de estudos em Humanidades Digitais na Iberoamérica. A RHD, que nasceu da iniciativa de pesquisadores da Universidad Nacional de Educación a Distancia (UNED, España), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina) e Universidad Nacional Autónoma de México (UNAM, México), teve em seu primeiro número, a compilação de vinte e dois trabalhos apresentados no Segundo Congresso Internacional da Sociedade Humanidades Digitais Hispânicas: Inovação, Globalização e Impacto¹, realizado no Universidade Nacional de Educação a Distância (UNED) de Madrid, de 5 a 7 de outubro, 2015.

Este primeiro número, em especial, representa o ponto de partida para a formação de uma rede digital de informação que possibilitasse a potencialização das atividades científicas em Humanidades Digitais (HD) adquirir valor e autenticidade à medida em que é conhecida pela comunidade científica e a sociedade em geral. A possibilidade de publicação em língua hispânica, inglesa e portuguesa da Revista, contribui para a difusão do conhecimento, sobretudo por ser em acesso aberto, gerado em HD de modo global, promovendo com isso a horizontalidade no acesso e nas discussões em diferentes países, constituindo-se como espaço de mediação da informação científica que integra as ações de interação entre a sabedoria prática –*práxis*, o saber-fazer– *poiesis*, e a experimentação, *legein*, ao estabelecer conexões entre as diversas redes de pesquisa em HD.

Considerando esta iniciativa como um marco teórico para a divulgação dos estudos sobre Humanidades Digitais, é possível identificar em seu primeiro número os conceitos que incluem o corpus de estudo, análise e práticas em HD, evidenciando a natureza metodológica,



fundamentação teórica e possíveis linhas de pesquisa deste campo científico no ano 2015, período inicial das produções em massa sobre métodos, dispositivos e perspectivas heurísticas relacionadas com o digital no campo das Ciências Humanas e Sociais.

Para essa identificação, substanciada pela análise e visualização dos vinte e três textos (incluindo o editorial) do primeiro número da RHD, utilizou-se a ferramenta *Voyant Tools*², largamente utilizada para análises conjuntas de dados textuais em diferentes níveis. Esta ferramenta, disponível on-line, é um aplicativo de código aberto (*open source*) que permite ao usuário analisar textos individuais ou coleções de textos, disponíveis na internet ou não, utilizando técnicas de mineração de texto. A possibilidade desta *leitura distante*, termo cunhado por Franco Moretti (2013)³, permite associar diferentes aplicações para cada elemento textual e construir uma multiplicidade de conexões entre os conceitos presentes nas palavras presentes no texto e seu contexto.

Entre as aplicações que permitem a mineração de texto do *Voyant*, para a análise do corpus textual do primeiro número da RHD, destaca-se algumas. A princípio, as palavras irrelevantes foram excluídas do corpus, chegando-se aos termos mais frequentes em toda a coleção de artigos (com 169.308 palavras): texto/textos (730), digital (583), datos (402), que representam os termos em tendência nos 23 textos analisados. Por exemplo, o trabalho “Construcción de una Base de Datos y un Repositorio de Documentos de Investigación para el Proyecto TRACE”, é onde o termo *datos* é mais frequente. Para este artigo, inclusive, utilizou-se a aplicação *Link*, que a partir de um algoritmo de força produz uma rede por meio do uso de gráfico, quando o termo *datos* se destaca no tamanho em virtude da frequência geral no texto, e pela quantidade de conexões com outros termos. Já na aplicação *Trends*, junto ao termo *digital*, por exemplo, mostrou a frequência de ambos os termos ao longo do artigo exemplificado, evidenciando a interconectividade entre os conceitos –*digital* e *datos*– e sua relação direta com os estudos em HD (Figura 1).

¹ Disponível em: <http://hdh2015.linhd.es/>.

² O livro “Hermeneutica: Computer-Assisted Interpretation in the Humanities” (2016), de Geoffrey Rockwell e Stéfan Sinclair, é uma importante referência para o uso da ferramenta *Voyant Tools*, pois apresenta diferentes abordagens para análise de textos usando o aplicativo.

³ Estudioso em História Literária, o italiano Franco Moretti idealizou o conceito de “leitura distante” no livro “Distant Reading” (2013).

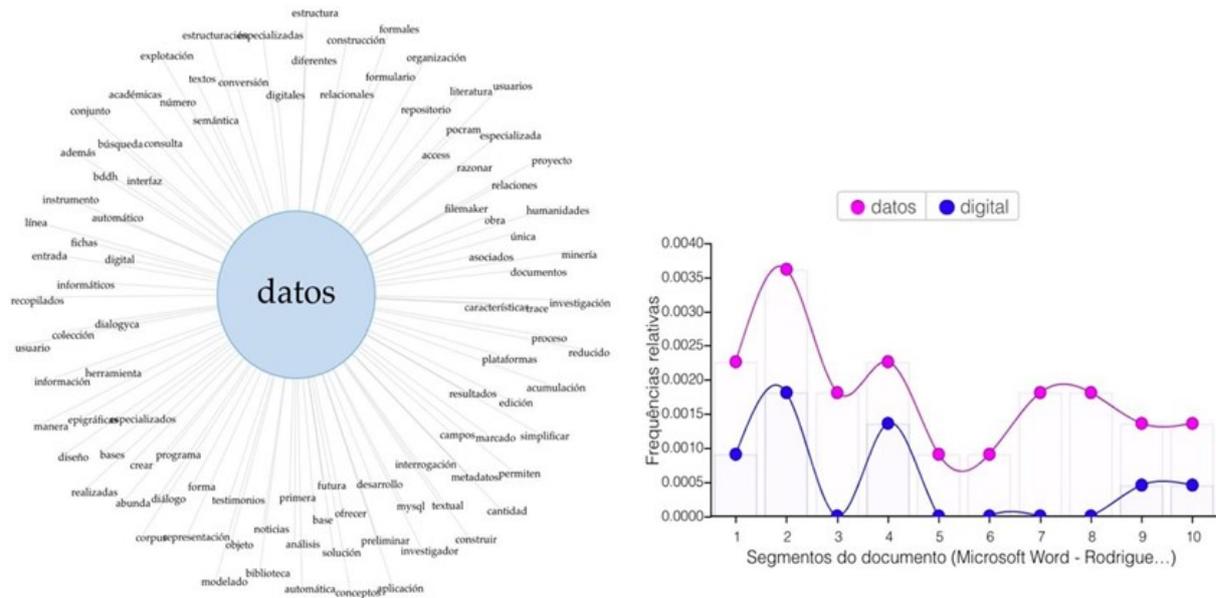


Figura 1. Aplicações Link e Trends do Voyant Tools para o termo *datos*. Fonte: elaborado pela autora a partir do corpus de análise.

Outra análise que se pôde verificar neste primeiro número da RHD, é em relação à representação geoespacial dos documentos que constituem o corpus. Foi possível visualizar, por meio da aplicação *DreamScape*, lugares do mapa geográfico citados nos textos. Grande parte das cidades, lugares e países estão correlacionados às fontes de informação utilizadas pelos autores para construção dos trabalhos, ou seja, onde foram publicadas as referências citadas desses documentos. Assim, identificou-se que a América do Norte e a Europa constituíram-se como os principais fornecedores de informações para os estudos em Humanidades Digitais neste número, seja pela referência aos autores e pesquisas citadas e/ou objetos de análise oriundos destas localizações geoespaciais.

Em se tratando de metodologias aplicadas no corpus, por exemplo, aplicou a *MicroSearch*, que permite escolher um termo para análise, o qual é identificado como pontos vermelhos indicados nos blocos que representam cada documento na coleção analisada. Seguindo a sintaxe de busca, orientada pela própria ferramenta, inseriu-se o termo *metodo!** para recuperar todas as variações semânticas relacionadas às metodologias delineadas nos trabalhos. Interessante destacar que o resultado foi um tanto tímido no que se refere ao uso deste termo nos trabalhos, embora, a partir da leitura individual de cada um, viu-se claramente o desenvolvimento de um caminho metodológico nas pesquisas. Por exemplo, em alguns trabalhos o termo é mencionado nas referências de uma obra citada, no entanto apresentam técnicas de análise e estudos originais no âmbito das HD ao longo do texto mesmo sem mencionar que se trata da metodologia utilizada.

A partir dessa leitura, pode-se inferir um caminho de amadurecimento da natureza metodológica em HD, apontando para um futuro de pesquisas onde seja necessário evidenciar quando começa e termina os procedimentos metodológicos aplicados em humanidades e suas convergências com outras disciplinas, organizando-se com isso um corpus metodológico em HD.

Para descobrir possíveis linhas de pesquisa em HD neste primeiro número da RHD, utilizou-se de forma conjunta quatro aplicações do *Voyant Tools*: 1) *Summary*, para identificar as palavras

mais específicas do corpus e elegê-las para uso nas outras aplicações utilizadas;

2) *Phrases*, para identificar o quanto é relevante uma frase no corpus; 3) *Contexts*, para a contextualização dos termos em análise; e 4) *Topics*, para compreender quais temas (grupos de termos) existem e como eles são distribuídos no contexto do corpus. Nesta análise conjunta, percebeu-se a ocorrência dos seguintes assuntos: Arquitetura da informação; Linguística; Paleografia; Educação; Periodismo Histórico; Organização e Representação do Conhecimento; Bibliotecas Digitais.

Nesta resenha crítica, procurou-se destacar o corpus teórico e metodológico, bem como a indicação de futuras linhas de pesquisa a partir do primeiro número da Revista de Humanidades Digitais, ressaltando sua contribuição enquanto canal de comunicação científica formal das pesquisas em HD, sobretudo por ser uma das revistas científicas pioneiras em debates exclusivos em Humanidades Digitais. Embora se tenha utilizado apenas algumas das 28 aplicações de análise disponíveis na ferramenta *Voyant Tools*, para esta leitura a distância dos artigos que compuseram o primeiro volume da Revista, ressalta-se que o objetivo da análise foi alcançado, criando possibilidades para estudos mais críticos sobre os fundamentos teóricos em HD e suas metodologias de análise, bem como suas influências e vieses de aplicações técnicas em diferentes contextos linguísticos e geográficos.

REFERÊNCIAS BIBLIOGRÁFICAS

Moretti, F. (2013). *Distant Reading*. Verso.

Rockwell, G., & Sinclair, S. (2016). *Hermeneutica: Computer-assisted interpretation in the humanities*. The MIT Press.