

DADIFICAÇÃO, VISUALIZAÇÃO E LEITURA DO MUNDO: QUEM FALA POR NÓS QUANDO OS NÚMEROS FALAM POR SI?

*Marcelo El Khouri Buzato**

RESUMO

Este trabalho está vinculado a uma investigação de dois anos sobre as implicações sociodiscursivas do fenômeno social da “dadificação”, i.e. coleta, manipulação e representação, para “extração de conhecimento”, de volumes astronômicos de dados digitais (Big Data) sobre o cotidiano do cidadão. Objetiva-se caracterizar os letramentos críticos (de dados) necessários, frente ao fenômeno, para (i) a conscientização do cidadão acerca das transgressões de seus valores éticos e culturais implicados nessas técnicas e epistemologias e (ii) fomentar o empoderamento do cidadão, sobretudo pela educação linguística e matemática, pela utilização dessas mesmas técnicas de geração, interpretação e comunicação de dados provenientes de repositórios de dados abertos disponibilizados pelo Estado, por força de lei, e por empresas de internet tais como Twitter e Facebook, via interfaces de aplicação (API). Nesta apresentação, abordarei especificamente os modos e estratégias pelas quais essas novas representações multimodais afetam os sujeitos em sua capacidade e disposição de ler – i.e. duvidar, criticar e interpelar – informações quantitativas/estatísticas representadas visualmente. Exemplificarei com um estudo exploratório sobre três mapas, um dos quais baseado em Big Data, que representam a distribuição racial no Brasil, no qual utilizei categorias de análise da teoria da multimodalidade em semiótica social. Concluo que, no caso do mapa do tipo Big Data, a visualização obtida gera uma poderosa sensação de envolvimento/empatia que induz o leitor a desconsiderar os componentes retórico e ideológico da representação.

Palavras-chave: Letramento de dados, Big Data; Multimodalidade.

ABSTRACT

This work is part of a two-year research project on the social-discursive implications of the social phenomena of “datafication”, i.e. the collection, manipulation and representation, for “extraction of knowledge”, of astronomical volumes of digital data (Big Data) on the daily life of the citizen. The objective is to characterize the critical (data) literacies necessary, in the face of the phenomenon, for (i) increasing citizen awareness of the transgressions of their ethical and cultural values implied in these techniques and epistemologies and (ii) fostering citizen empowerment, above all through the use of these same techniques of generation, interpretation and communication of data from open data repositories made available by the State, by law, and by Internet companies such as Twitter and Facebook, via application interfaces (API). In this presentation, I will specifically address the modes and strategies by which these new multimodal representations affect subjects in their capacity and willingness to read - i.e. doubt, criticize, and question - quantitatively/visually represented statistics. I will exemplify an exploratory study of three maps, one of which is generated on Big Data, representing the racial distribution in Brazil, in which I used categories of analysis from the theory of multimodality in social semiotics. I conclude that in the case of the Big Data type map, the visualization obtained generates a powerful sense of involvement/empathy that induces the reader to disregard the rhetorical and ideological components of representation.

Keywords: Data literacies; Big Data; Multimodality.

* Professor pesquisador do Programa de Pós-Graduação em Linguística Aplicada (Unicamp).

INTRODUÇÃO

Todos temos consciência hoje das diversas transformações sociais, políticas e culturais envolvidas no processo de digitalização do mundo em marcha acelerada desde meados do século XX, porém as consequências epistemológicas de tal processo, sobretudo para as ciências sociais e humanas, são ainda pouco discutidas, sobretudo nos estudos da linguagem.

Lankshear (2003) já destacava três tendências importantes nas condições de produção do saber advindas da digitalização, a saber:

- (1) Performatividade: substituição do conhecimento como crença justificada pela busca da máxima eficiência como critério de cientificidade
- (2) Dadificação: uso de grandes volumes de dados em linha (Big data), recursos de aprendizagem de máquina (*machine learning*) e inteligência artificial em crescente disputa com método científico tradicional e com os modelos de busca da verdade baseados em teorias
- (3) Ampliação da ideologia do “conhecimento neutro” e da crença no fim dos “enquadramentos ideológicos” da informação

Podemos tomar estas três tendências, respectivamente, como o “que”, o “como” e o “porque” da dadificação, destacando em particular, a terceira tendência, ilustrada na notória afirmação de Anderson (2008, parágrafo 7, minha tradução) a seguir:

Este mundo onde quantidades enormes de dados e matemática aplicada substituem qualquer outra ferramenta que seja levada em conta. Fora com todas as teorias de comportamento humano, da linguística à sociologia. Esqueçam taxonomia, ontologia e psicologia. Quem sabe por quê as pessoas fazem o que elas fazem? O ponto é que elas o fazem, e nós podemos rastrear e medir isso com uma fidelidade sem precedentes. Com dados suficientes, os números falam por si mesmos.

Este trabalho objetiva discutir que tipo de (multi)letramentos toram-se relevantes, talvez indispensáveis, aos sujeitos sociais hoje, menos como forma de “ouvir o que os números falam”, embora isso seja indispensável, mas como forma de lidar criticamente com essas três tendências subjacentes à definição do que conta como “realidade”, e, portanto, como ponto de partida para a luta política, para as afirmações identitárias, para os currículos educacionais a serem impostos ou imaginados daqui por diante, entre outras batalhas que nos esperam na era da dadificação.

LETRAMENTOS DE DADOS

Tendo em mente estudantes de graduação, Schield (2004, p.8, minha tradução) define letramento de dados como a habilidade de “acessar, avaliar, manipular, sumarizar e apresentar dados”. O autor acrescenta que o letramento de dados é intimamente relacionado com dois outros letramentos: o letramento de informação (capacidade de reconhecer quando certa informação é necessária e habilidade de localizá-la, avalia-la e usá-la eficazmente) e o letramento estatístico

(saber usar dados estatísticos como evidência argumentativa, que comparações entre estatísticas fazer e como comunica-las, saber reconhecer o caráter socioconstrutivo das estatísticas e levar em conta o contexto na sua interpretação).

Mais recentemente, Bhargava e D'Ignazio (2014, p. 2) definiram o letramento de dados como a habilidade de ler, trabalhar com, analisar e discutir com dados. Uma definição mais abrangente do conceito é fornecida pelo Aliança Data Pop, consórcio liderado pela *Rockefeller Foundation*, *Harvard Humanitarian Initiative* e *MIT Media Lab* entre outros atores, cuja missão é “juntar pesquisadores, especialistas, praticantes e ativistas para promover uma revolução Big Data centrada em pessoas”. Enfatizando o caráter social e político do letramento de dados, sem deixar de considerar as habilidades nele envolvidas, o consórcio propõe a seguinte definição: letramento de dados é “o desejo e a habilidade de engajar-se construtivamente na sociedade através de ou a respeito de dados”. (DATA POP ALLIANCE, 2015, p.8, minha tradução).

Essa definição é detalhada pelo consórcio, que destaca alguns de seus termos e seus significados. “Desejo e habilidade” visa destacar o papel da tecnologia como amplificadora da capacidade e intenção humanas. ”Habilidade” comporta vários níveis de letramento relacionados a diferentes posições e objetivos, a partir de um patamar mínimo, e não uma dicotomia entre letrados e não-letrados em dados. “Dados” é um conceito abrangente que se refere tanto a itens individuais quanto a conjuntos de fatos, tanto a textos, fotos e vídeos quanto a estatísticas e itens informativos, tanto a dados não estruturados (arquivos) quanto a dados estruturados (bancos de dados). ”Engajar-se construtivamente na sociedade” pressupõe que o sujeito letrado tem um senso de propósito ativo e que se engaja com os dados em relação a objetivos específicos. “Através de ou a respeito de” indica que um indivíduo pode ser letrado em dados mesmo sem ser capaz de se envolver com análise avançada.

À ideia de engajamento do cidadão com a sociedade “através de ou a respeito de” dados, subjaz a noção do Big Data como grande oportunidade de “mudar as coisas”, de empoderamento do cidadão pela apropriação de dados abertos, disponíveis para o público, e das técnicas da ciência de dados. Essa visão tende a obscurecer o fato de que já estamos todos de alguma forma engajados “através de” dados quando nos tornamos usuários de redes sociais, utilizamos nossos cartões de crédito ou baixamos aplicativos de celular que rastreiam nossos hábitos e nossa localização gerando dados contextuais involuntários.

Se a questão é empoderar o cidadão, trata-se, então, de pensarmos em diferentes formas e níveis de engajamento. Por exemplo, seria conveniente pensar-se em graus diferentes de letramento de dados para pessoas que procuram saber que dados seus são gerados e como e por quem são utilizados e pessoas que não o fazem. O mesmo se aplica a pessoas que usam os dados a seu próprio respeito que estão disponíveis com algum propósito e pessoas que não o fazem.

No caso de pessoas que se engajam ativamente com dados, em que medida seu engajamento se relaciona com habilidades e atitudes específicas? No caso das pessoas que não se engajam ativamente, que propósitos são capazes de imaginar quando se lhes apresenta um conjunto de dados sobre si? Seria o seu não engajamento uma questão de atitude ou de falta de habilidades necessárias?

Mais do que levantar as atitudes e seus porquês é importante, em face do avanço desse processo

na sociedade, havendo engajamento, de que maneira nós, dos estudos da linguagem, devemos preparar os cidadãos para combater, em vez de adquirir, a ideologia da neutralidade dos números, por meio de capacidades críticas de interpretação das representações de mundo produzidas no novo paradigma?

Esta pesquisa se ramifica em pelo menos duas direções. A primeira é o desenvolvimento desse conceito, “letramento de dados” pensando-se a dadificação como prática social. A segunda, que ilustro a seguir, é um levantamento das formas textuais centrais nessas práticas, prioritariamente aquelas a que tem acesso o cidadão comum. Nesse caso, o marco fundamental são as assim chamadas visualizações de dados.

VISUALIZAÇÕES DE DADOS: É VER OU LER?

O volume, velocidade e variedade dos dados que alicerçam o novo paradigma de “produção da realidade” é tal que mesmo dispendo de todos os números, ninguém pertencente à raça humana é capaz de interpretá-los diretamente. Por isso, vem tornando-se cada vez mais poderosa uma técnica ou prática advinda da confluência entre as ciências da computação, a estatística e a semiótica chamada visualização de dados, ou *DataVis* (*data visualization*).

A visualização de dados tem por missão tanto representar as correlações estatísticas obtidas nos processos de análise de dados feitos por técnicas de inteligência artificial e aprendizagem de máquina como comunicar esses achados para os diferentes públicos interessados, provendo-lhes um poderoso instrumento auxiliar na tomada de decisões. As representações produzidas nessas práticas teriam por finalidades precípua, portanto, facilitar *insights* sobre o significado das referidas correlações e, segundo, servirem como evidência de fatos até então “invisíveis” a olho nu.

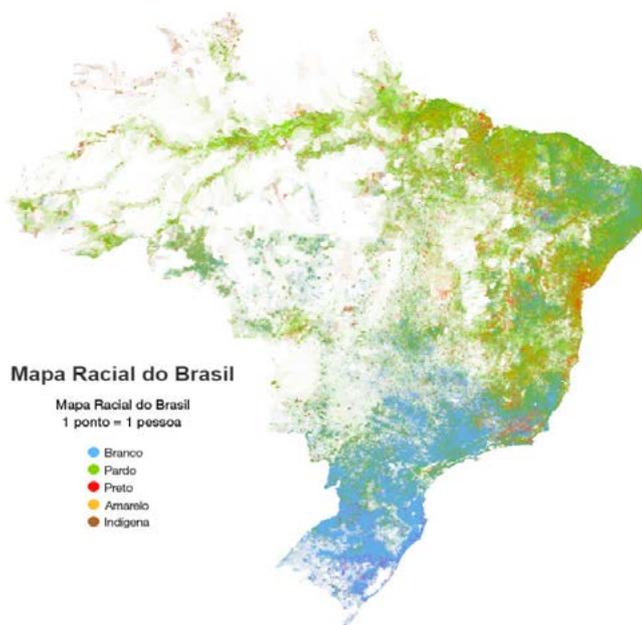
As formas ou gêneros (BUZATO, 2018) de visualização de dados mais comuns são os gráficos que aprendemos na escola, em especial os de barras, os de linha e os de pizza (diagramas de setores). Mas o Big Data traz consigo uma ampla gama de formatos voltados para finalidades específicas de natureza relacional, isto é, mostrar dispersão ou agrupamento, evolução temporal, densidade de fluxos, distribuição espacial e assim por diante. Não há espaço aqui para mostrar e explicar como funcionam diagramas aluviais, cloropletos, mapas de contorno, diagramas de arco etc., mas uma tipologia confiável pode ser consultada no site da biblioteca da Universidade Duke, EUA¹.

Neste trabalho, meu objetivo é simplesmente ilustrar de que modo o panorama ideológico naturalista-positivista de autores como Anderson (2008) se manifesta nessas formas textuais, e por que caminhos pode ser desmontado. Por essa razão, escolho uma representação mais simples, mas que combina o poder do Big Data com elementos que um brasileiro escolarizado médio já conhece: um “mapa de densidade de pontos um a um” (*one-to-one dot-map*) da distribuição racial no Brasil baseado no censo do IBGE mais recente (Figura 1). O que impressiona nessa visualização interativa é que cada ponto representa um brasileiro, ou um grupo de brasileiros, dependendo do

1 ZOSS, Angela. Introduction to Data Visualization: Visualization Types. Institutional. Disponível em: <http://guides.library.duke.edu/datavis/vis_types>. Acesso em: 28 jan. 2017.

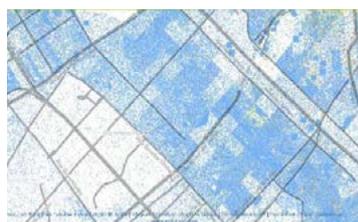
grau de zoom selecionado, em sua localização exata, ou seja, ele traz duas dimensões de dados (raça e localização) para milhões de pontos de dados (cada pessoa ou família) que podem ser agregados em diversas escalas espaciais, com amostras delimitadas segundo o interesse do leitor/visualizador, e não divisões geopolíticas abstratas.

Figural 1: Mapa de densidade de pontos da distribuição racial no Brasil



Em sua versão eletrônica interativa disponível na web², oferece a funcionalidade *zoom*, por meio da qual o leitor é capaz de perceber diferentes padrões de distribuição racial em diferentes escalas (rua, bairro, cidade e assim por diante). A Figura 2 compara, para exemplificar, a distribuição nas áreas centrais de São Paulo, Rio de Janeiro e Salvador)

Figural 2: Utilização da função zoom permite examinar padrões locais



São Paulo: predomínio da população branca, com negros e pardos em número pequeno e dispersos espacialmente



Rio de Janeiro: negros e pardos dispersos nas regiões menos valorizadas e concentrados em uma espécie de enclave.

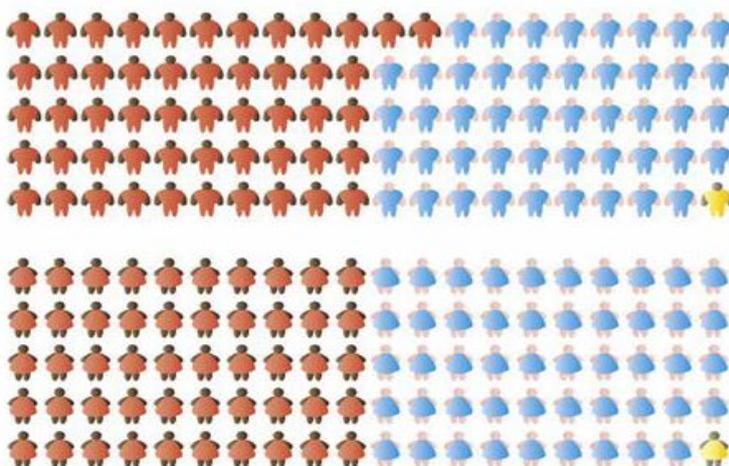
² Disponível em <<http://patadata.org/maparacial/#lat=-87.110736&lon=76.571172&z=3&o=t>>, acesso em 20/01/2017.



Salvador: pardos e negros predominam em grandes áreas relativamente contínuas entre si; brancos concentrados em diversos pequenos enclaves.

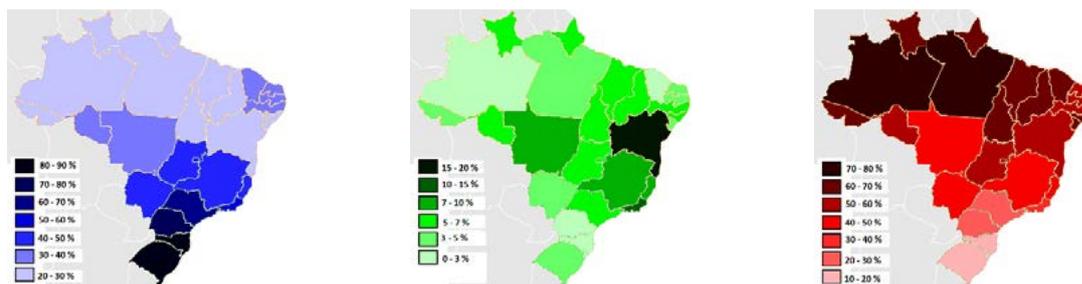
Como objeto de comparação, podemos considerar as Figuras 3 e 4, que utilizam basicamente os mesmos dados, com estratégias diversas de conversão de números em imagens:

Figura 3: Distribuição racial no Brasil de acordo com sexo e raça



Fonte: IPEA (2011, p.16). Roupas vermelhas = negros; azuis = brancos, amarelas = todos os demais

Figura 4: Distribuição percentual das raças (cores) por estado no Brasil



Fonte: Wikipedia (COMPOSIÇÃO ÉTNICA DO BRASIL, 2016)³.
Azul = brancos; verde = negros; vermelho = pardos

Mesmo utilizando exatamente as mesmas fontes de dados, as diferentes estratégias de representação têm efeitos de sentido absolutamente diversos. É relativamente fácil evidenciar alguns desses efeitos a partir de procedimentos usuais em análise do discurso visual, mas há outros de difícil apreensão por esse caminho.

³ Autoria de Henriquehorta (trabalho próprio) [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons.

Por exemplo, é fácil determinar como procedimento retórico o uso de cores primárias como um modo semiótico (KRESS, 2010) que (i) caracteriza o trabalho como neutro ou científico (já que as cores primárias são de fato as mais adequadas em termos de expressão de diferenças e distinção de categorias seja qual for o alvo da representação), mas também (ii) reforça (no caso das figuras 1 e 4) ou questiona (no caso da figura 3, que classifica pardos e negros como um mesmo grupo) a ontologia racial utilizada pelo IBGE (JAMES, 2017).

Do mesmo modo, percebe-se que a utilização do espaço como modo é muito semelhante nas Figuras 3 e 4 em contraste com a Figura 1. Enquanto naquelas utiliza-se padrões geométricos abstratos (formas geométricas na Fig.3, linhas imaginárias de divisão do território na Fig.4) como retentores de cores para “forçar” a noção de “homogeneidades opostas”, na Figura 1, espaço e cor atuam juntos criando aglomerados, dispersões, continuidades e descontinuidades em diferentes escalas, e, portanto, desmontando os mecanismos abstratos de retenção das outras duas representações.

Interessa-nos mostrar, entretanto, que a técnica de visualização de dados utilizada na Figura 1 não pode ter seu potencial esgotado por esse tipo usual de análise, já que ela se apresenta ao leitor menos como “representação” do que como “visualização” da realidade. Isso porque, ao utilizar o mapa em sua versão interativa, o leitor tem a possibilidade inusitada de “ver-se a si mesmo” no mapa, e, a partir daí, rever o mundo físico ao seu redor com “outros olhos” para a questão racial, perceber-se parte dela de um modo que nem o discurso verbal nem a representação visual de dados agregados e não georreferenciados com tal refinamento possibilitam.

Em contraste, a sensação de **ser** um dos pontos no mapa leva o leitor a percebê-lo como “mais realista”, ainda que os números utilizados sejam exatamente os mesmos que nos outros dois casos. Sentir-se envolvido no fenômeno, a despeito do enquadramento objetivo e da “gramática” estatística e geométrica da Figura 1, é um efeito de sentido que não pode ser atribuído exclusivamente nem à percepção pura, nem representação cultural. No entanto, tudo se passa como se o mapa de densidade de pontos fosse transparente, com muito menos evidências de “autoria” que os demais. Essa suposta transparência e/ou não autoria pode ser atribuída a dois fatores.

Primeiramente, a sensação de que se está vendo o fato, e não seu retrato ou tradução, é efeito de cada ponto de dados representar, com exatidão, um ponto físico no território urbano ou nacional, e esse, por sua vez, corresponder precisamente, quando se utiliza o máximo zoom, a um corpo ou um domicílio que pode ser visto, tocado, frequentado, ou seja, apreendido (previamente, se na minha cidade) pela experiência direta. Não fossem os limites técnicos e éticos, poder-se-ia efetivamente trocar cada ponto por um retrato da pessoa ou família representada, por exemplo. Novamente, a ninguém, se não ao próprio ponto, se pode atribuir o local em que está ou quem representa.

Em segundo lugar, o mapa de pontos utiliza uma combinação muito especial de iconicidade e abstração que facilita o esquecimento da autoria. Nele, as relações espaciais são icônicas, ou seja, as ruas se parecem com ruas, a vegetação se parece com vegetação, a densidade de aglomeração dos pontos acompanha a taxa de aumento ou diminuição do zoom e assim por diante. Já os pontos de dados, em si, não são icônicos, são pontos coloridos de cores primárias que não correspondem aos fenótipos dos grupos raciais representados, mas à melhor combinação técnica para que as visualizações em diferentes níveis de zoom funcionem. Assim, a questão racial, em si, torna-se abstrata e, portanto, neutra, enquanto as circunstâncias e relações espaciais dos grupos raciais são “reais”.

Na figura 4, por contraste, o autor representou os pontos de dados como pessoas, com fenótipos correspondentes a seus grupos raciais, e roupas que trazem as mesmas cores primárias usadas na figura 1. Podemos ver nisso um grau maior de iconicidade, porém os pontos em si estão agregados (cada ícone representa uma porção de população, mas não a pessoa tal que mora em tal casa). Assim, houve aí a iconização de uma abstração, um primeiro indício de autoria.

Ademais, mesmo utilizando-se de cores primárias, como na figura 1, vê-se facilmente, na figura 4, que a criação dos ícones envolveu escolhas estéticas do tipo que cor de roupa vai com que fenótipo ou que tipo de silhueta terão os ícones, enquanto no mapa de pontos, a questão do fenótipo desaparece totalmente. Fica, na figura 1, somente a sensação de estar-se muito mais perto do real, embora os números sejam os mesmos.

CONSIDERAÇÕES FINAIS

Se é obvio que essas nova práticas de representação/visualização da realidade estão profundamente enraizadas em um processo mais amplo de mudança epistemológica decorrente da dadificação do mundo, não está bem claro de que forma devemos preparar os cidadãos para tirarem proveito de discursos multimodais como o da Figural 1 para repensarem suas subjetividades e seus contextos de vida, tampouco para interpelarem esses discursos criticamente.

Para estabelecer uma noção de letramentos de dados (críticos) que efetivamente permita o desenvolvimento de estratégias educacionais, sobretudo no campo das ciências humanas, não basta que se melhore a formação dos alunos, professores e pesquisadores dessa área em matemática, estatística, técnicas de visualização de dados e outros recursos que já se mostram indispensáveis para a sociedade como um todo. É preciso que encontremos formas de ampliar nossa caixa de ferramentas analíticas nas ciências da linguagem para abarcar dimensões do sentido que não se deixam capturar pelas formas e modos semióticos empregados, senão pela fenomenologia que estabelece a ponte entre leitura, visualização e experiência e/ou entre quantificação, expressão e realidade.

Um dos conceitos que pode ser trabalhado nesse sentido de ampliação e reinvenção é o de gênero textual/discursivo. Argumento em Buzato (2018) que as visualizações podem ser vistas como gêneros híbridos de narrativa e banco de dados, conforme o insight de Manovich (2000), mas não está clara a concepção de gênero que melhor atende a essas questões.

Outro caminho de desenvolvimento, seria o que procurei ilustrar aqui, qual seja, formas de análise do discurso multimodal que incorporem os efeitos de sentido derivados do poder de que o Big Data tem de aproximar o leitor/visualizador de fenômenos que fazem parte de sua experiência corporificada, mas não podiam ser vistos a olho nu; e, por outro lado, a forma como o Big Data traz ao mundo, visualmente, fenômenos totalmente estranhos à nossa experiência, mas que são estatisticamente inquestionáveis e, portanto, podem ser usados para estabelecer novas ontologias, o que implica a abertura de novos fronts políticos, como tem acontecido, por exemplo, com o aquecimento global, um fenômeno tipicamente constituído a partir do uso de Big Data nas pesquisas sobre ciências atmosféricas.

Finalmente, essas tarefas são claramente interdisciplinares, e é preocupante que a grande maioria dos pesquisadores e educadores da área de línguas/linguagens não tenha a formação mínima necessária para um engajamento de alto nível com o tema, e com os pesquisadores das ciências exatas nele envolvidos. Estou certo de que um caminho produtivo para a aquisição desse conhecimento entre nós, no Brasil, é o estabelecimento das assim chamadas humanidades digitais como um campo efetivo, de vocação aplicada, aberto à colaboração e apropriação mútua entre as duas grandes culturas epistêmicas (as humanas/sociais versus as exatas/biológicas) que hoje se encontram divorciadas, quando não em guerra (BUZATO, 2017).

REFERÊNCIAS

ANDERSON, Chris. The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. **WIRED**, 23 jun. 2008. Disponível em: <<http://www.wired.com/2008/06/pb-theory/>>. Acesso em: 26 jul. 2016.

BHARGAVA, Rahul D'IGNAZIO, Catherine. Designing Tools and Activities for Data Literacy Learners. In: **Data Literacy Workshop at Web Science**. Oxford, UK: University of Oxford, 2015. Disponível em: <<https://www.media.mit.edu/publications/designing-tools-and-activities-for-data-literacy-learners/>>. Acesso em: 18 ago. 2016.

BUZATO, Marcelo. Critical Data Literacies: going beyond words to challenge the illusion of a literal world. In: TAKAKI, Nara Hiroko; MONTE MOR, Walkyria (Org.). **Construções de sentido e letramento digital crítico na área de línguas e linguagem**. Campinas, SP: Pontes Editores, 2017. p. 119–142.

BUZATO, Marcelo El Khouri. *Data Storytelling* e a dadificação de tudo: um gênero bastardo de mãe narrativa e pai banco de dados. In: LIMA-LOPES, Rodrigo Esteves de; BUZATO, Marcelo El Khouri (Org.). **Gênero Reloading**. Campinas, SP: Potes Editores, 2018. p. 95–123.

COMPOSIÇÃO ÉTNICA DO BRASIL. In: Wikipédia, a enciclopédia livre. [S.l.: s.n.], 2016. Disponível em: <https://pt.wikipedia.org/w/index.php?title=Composi%C3%A7%C3%A3o_%C3%A9tnica_do_Brasil&oldid=47362644>. Acesso: 4 Abr. 2017.

DATA POP ALLIANCE. Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. 2015. Disponível em: <<https://datatherapy.files.wordpress.com/2015/10/beyond-data-literacy-2015.pdf>>.

IPEA-INSTITUTO DE PESQUISA ECONÔMICA APLICADA (Org.). **Retrato das desigualdades de gênero e raça**. 4. ed. Brasília: Ipea, 2011.

JAMES, Michael. Race. In: ZALTA, EDWARD N. (Org.). **The Stanford Encyclopedia of Philosophy**. Spring 2017 ed. [S.l.]: Metaphysics Research Lab, Stanford University, 2017. Available at <<https://plato.stanford.edu/archives/spr2017/entries/race/>>. Access: 1 Apr. 2017.

KRESS, Gunther R. **Multimodality: a social semiotic approach to contemporary communication.** London; New York: Routledge, 2010.

LANKSHEAR, Colin. The Challenge of Digital Epistemologies. **Education, Communication & Information**, v. 3, n. 2, p. 167–186, 2003.

MANOVICH, Lev. Database as a genre of new media. **AI & Society**, v. 14, n. 2, p. 176–183, jun. 2000.

SHIELD, Milo. Information Literacy, Statistical Literacy, Data Literacy. **IASSIST Quarterly**, v. 28, p. 6–11, 2004.