

COLEÇÕES DE DADOS BRASILEIRAS PARA O ENSINO DE PORTUGUÊS

Marcia dos Santos Machado Vieira

Universidade Federal do Rio de Janeiro

Juliana Bertucci Barbosa

Universidade Federal do Triângulo Mineiro

RESUMO: No Brasil – um país plurilíngue com mais de duzentos milhões de falantes do Português –, a diversidade sociolinguística é patrimônio cultural que, embora esteja documentado em algumas coleções de dados linguísticos, carece de documentação via parâmetros de constituição de acervos de memória nacionalmente definidos. Coleções desse patrimônio heterogêneo podem ser exploradas em (inter)ações de ensino em contexto de língua materna e não materna, em outras ações discursivas relativas a tradução, interpretação, legendagem, dublagem, bem como em outras áreas do saber. Conscientes do potencial desse patrimônio, seis pesquisadores do GT de Sociolinguística da Associação Nacional de Pós-Graduação e Pesquisa em Letras e Linguística (ANPOLL) desenharam um projeto de repositório digital nacional para a catalogação e/ou a reunião das coleções brasileiras existentes e a serem constituídas: a Plataforma da Diversidade Linguística Brasileira. Este capítulo visa a oferecer ao leitor informações sobre coleções de dados de fala e escrita do território brasileiro e sobre o projeto de repositório, que está em desenvolvimento na Associação Brasileira de Linguística

(ABRALIN).¹⁶² Assim, busca-se contribuir para um trabalho com Português que promova atenção e respeito à diversidade linguística e sociocultural, aos modos de vida da população no Brasil.

INTRODUÇÃO

Há muitas línguas faladas no Brasil.¹⁶³ A estimativa é a de que mais de 250 línguas sejam usadas no Brasil, se considerarmos línguas indígenas, línguas de imigração, línguas de sinais (LIBRAS, por exemplo), afro-brasileiras. E, assim, coexistem, resistem ou convivem indivíduos e comunidades em situação de multilinguismo no país: o número de línguas usadas por um indivíduo pode ser bastante variado; há quem fale uma língua indígena ou outra língua (língua pomerana,¹⁶⁴ talian,¹⁶⁵ por exemplo), além do Português. O Decreto n. 7.387, de 09 de dezembro de 2010, que instituiu o Inventário Nacional da Diversidade Linguística (INDL), põe, oficialmente, em proeminência a urgência de documentação e salvaguarda da heterogeneidade linguística brasileira, bem como sua valorização.

Português, um dos dez idiomas mais falados no mundo (segundo *Ethnologue*, 2021 em gráfico a seguir), é a língua da maioria da população no Brasil, de mais de duzentos e treze milhões e setecentas mil pessoas.¹⁶⁶ E a estimativa é de contínuo crescimento dessa população.

¹⁶² Contato: data@abralin.org.

¹⁶³ Conferir, por exemplo: <https://pib.socioambiental.org/pt/L%c3%adnguas>, <http://portal.iphan.gov.br/indl>, <http://prodoclin.museudoindio.gov.br/index.php/documentando-linguas>, <https://www.museu-goeldi.br/assuntos/colecoes/linguistica>. (Acesso em: 09 out. 2021).

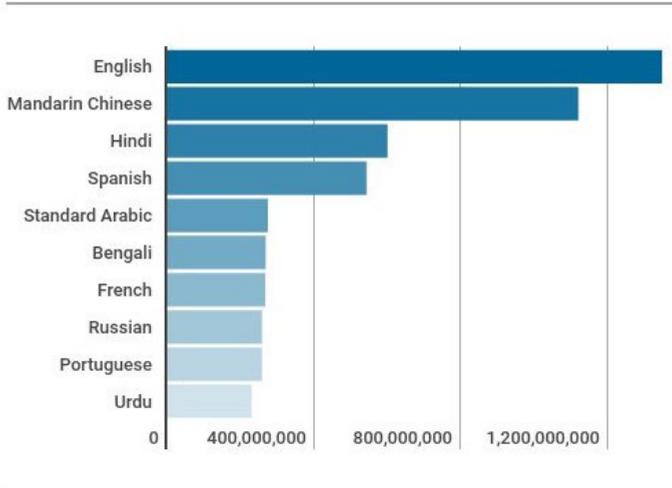
¹⁶⁴ “(...) a etnia descende de tribos eslavas e germânicas que viveram na região da Pomerânia (Pommerland), hoje conhecida como Pomerânia Oriental, devastada pela Segunda Guerra Mundial e atualmente situada entre a Alemanha e a Polônia. Os pomeranos do Brasil, situados nos estados do Rio Grande do Sul, Santa Catarina e Espírito Santo, falam um idioma extinto de origem baixo saxônica, o pomerano. Essa língua foi preservada graças a um projeto capixaba, o Programa de Educação Escolar Pomerana. Os pomeranos conservam a floresta de suas propriedades por acreditarem que sua destruição atrairia para suas casas tudo o que nela existe de malévolo, sob um ponto de vista material ou espiritual”. (Disponível em <http://portal.iphan.gov.br/noticias/detalhes/1107>. Acesso em: 09 out. 2021)

¹⁶⁵ “O Talian é uma das autodenominações para a língua de imigração falada no Brasil na região de ocupação italiana direta e seus desdobramentos desde 1875, em especial no nordeste do Rio Grande do Sul, Paraná, Santa Catarina, Mato Grosso e Espírito Santo. Sua origem linguística é o italiano e os dialetos falados, principalmente, na regiões do Vêneto, Trentino-Alto e Friuli-Venezia Giulia e Piemontes, Emilia-Romagna e Ligúria”. (Disponível em: <http://portal.iphan.gov.br/pagina/detalhes/183/>. Acesso em: 09 out. 2021).

¹⁶⁶ https://www.ibge.gov.br/apps/populacao/projecao/index.html?utm_source=portal&utm_medium=popclock&utm_campaign=novo_popclock (Acesso em: 09 out. 2021).

Gráfico 1 – Os 10 principais idiomas mais falados.

Top 10 most spoken languages, 2021



Ethnologue

Fonte: <http://www.ethnologue.com/guides/ethnologue200>. Acesso em: 09 out. 2021.

Logo, ações voltadas para o Português como língua de ciência, educação, cultura e relação em sociedade (nacional e internacional) requerem farta exemplificação para que se dê conta devidamente do mapeamento de sua realidade linguística plural e de sua contextualização histórica, sociocultural e geopolítica. Conscientes disso, seis pesquisadores do GT de Sociolinguística da ANPOLL¹⁶⁷ (Associação Nacional de Pós-Graduação e Pesquisa em Letras e Linguística) desenharam o Projeto Plataforma da Diversidade Linguística Brasileira,¹⁶⁸ apresentado ao Centro de Referência do Museu da Língua Portuguesa¹⁶⁹ em agosto de 2021 e, logo em seguida, à Associação Brasileira de Linguística (ABRALIN).¹⁷⁰ O desenho desse projeto foi exposto numa atividade¹⁷¹ do Congresso Internacional da Abralín de 2021 (InterAb 12) intitulada *Acervos de dados abertos à sociedade: memória*

¹⁶⁷ <https://anpoll.org.br/gt/sociolinguistica/> (Acesso em: 09 out. 2021).

¹⁶⁸ MACHADO VIEIRA, WIEDEMER, FREITAG, BARBOSA, PERES e MOLLICA, 2021.

¹⁶⁹ <https://www.museudalinguaportuguesa.org.br/mlp/centro-de-referencia/> (Acesso em: 09 out. 2021).

¹⁷⁰ <https://www.abralin.org/site/> (Acesso em: 09 out. 2021).

¹⁷¹ <https://www.youtube.com/watch?v=BsCvqcTo-qc&t=4374s> (Acesso em: 09 out. 2021).

linguística e sociocultural e potencialidade de (re)uso e noutra atividade¹⁷² do Seminário Viagens da Língua do Museu da Língua Portuguesa intitulada *Culturas multilíngues*. E, antes disso, os pilares que o sustentam, ciência e educação abertas à sociedade, foram abordados por ocasião do Festival do Conhecimento da UFRJ em 2021, em atividade¹⁷³ intitulada *Futuros possíveis para dados sociolinguísticos*.

BANCOS DE DADOS (SOCIO)LINGUÍSTICOS BRASILEIROS: O REPOSITÓRIO DIGITAL PLATAFORMA DA DIVERSIDADE LINGUÍSTICA BRASILEIRA (ABRALIN)

Em linhas gerais, o projeto *Plataforma da Diversidade Linguística Brasileira* consiste na organização de um repositório digital de amostras de fala e escrita, bem como de amostras de língua(s) sinalizada(s), presentes no território brasileiro. Essa plataforma deve ter uma arquitetura tecnológica e arquivística potente e plenamente capaz não só de catalogar e abrigar coleções de dados e metadados sociolinguísticos e histórico-culturais, mas também de potencializar o acesso a informações diversas por diferentes públicos e incentivar múltiplos sujeitos a cooperarem e interoperarem em prol da contínua atualização, reformulação, ampliação, curadoria e revitalização desse repositório.

Está no horizonte a possibilidade de tal repositório vir a figurar como referência nacional e internacional, inclusive em espaços museológicos, de documentação de amostras da memória linguística e cultural do país, empreendida por (socio) linguistas em interação com pesquisadores de outras áreas da ciência. Além disso, planeja-se que a plataforma possa ser acessada e aproveitada por futuros pesquisadores (de qualquer área do saber, como, por exemplo, inteligência artificial, tradução automática, diplomacia, historiadores entre outros), por educadores, estudantes, e, enfim, pela sociedade em geral.

Planejamos configurar como patrimônio cultural os seguintes objetos: i) o mapeamento e a catalogação nacional de coleções de dados linguísticos espalhados pelo Brasil e constituídas com base em parâmetros diversos via sistema de informações e metadados facilitador de buscas automatizadas; e ii) a construção de uma ferramenta digital para servir de repositório às coleções de (meta)dados (socio) linguísticos, com indexação a serviços de consultoria e assessoria especializados na área de Letras e Linguística, com produtos resultantes de pesquisas de dados nesse repositório destinados, entre outros fins, a ensino de línguas, educação patrimonial, ciência de pesquisa de campo, infraestrutura de armazenamento,

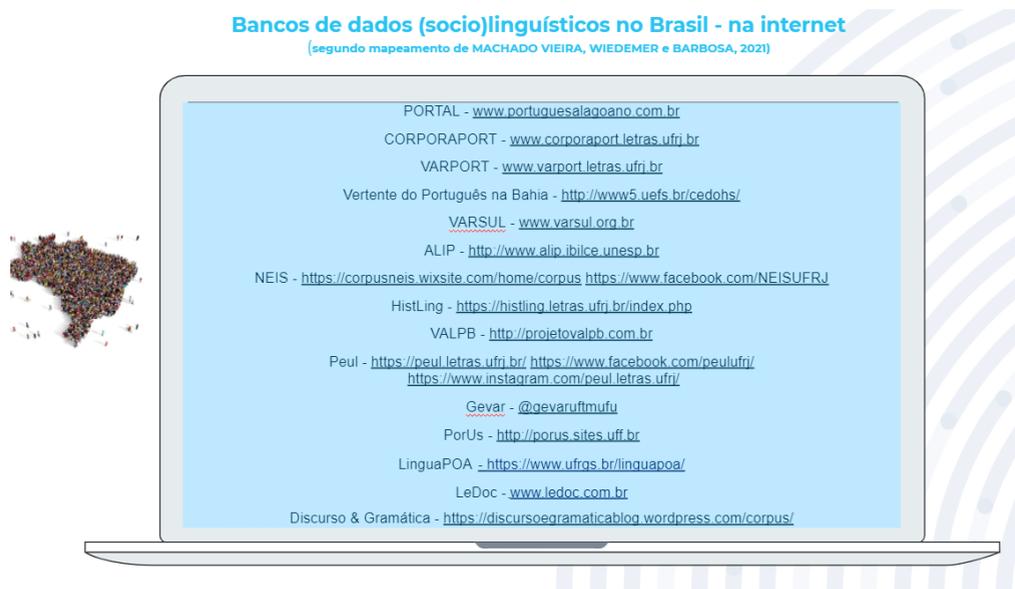
¹⁷² <https://www.youtube.com/watch?v=IaeiXl88yuo> (Acesso em: 09 dez. 2021).

¹⁷³ <https://www.youtube.com/watch?v=ZrZxsd5QQns> (Acesso em: 09 out. 2021).

patrimonialização, salvaguarda e valorização de (meta)dados sociolinguísticos. Essa ferramenta digital viabilizará a introdução de novas coleções de dados e o registro de demandas e *feedbacks* para o vital aprimoramento da plataforma digital e do trabalho com que ela se implementa e com acessibilidade a consulentes linguistas e não linguistas, brasileiros ou não.

Quanto aos bancos de dados já existentes e que poderão vir a ser os primeiros a integrar esse repositório ou o catálogo de bancos nele pretendido, a situação, em linhas gerais, é a seguinte: alguns bancos de dados já estão à disposição de consulentes, outros ainda não. Entre os que estão, segundo o que foi informado até agora por gestores de bancos de dados brasileiros que responderam ao Projeto “Mapeamento de bancos de dados (socio)linguísticos no Brasil” (MACHADO VIEIRA; WIEDEMER; BARBOSA, 2021), há, por exemplo, estes:

Figura 1 – Relação de sites de bancos de dados (socio)linguísticos brasileiros.



Fonte: Machado Vieira, Wiedemer e Barbosa (2021).¹⁷⁴

¹⁷⁴ PORTAL - www.portuguesalagoano.com.br, CORPORAPORT - www.corporaport.letras.ufjf.br, VARPORT - www.varport.letras.ufjf.br, Vertente do Português na Bahia - <http://www5.uefs.br/cedohs/>, VARSUL - www.varsul.org.br, ALIP - <http://www.alip.ibilce.unesp.br>, NEIS - <https://corpusneis.wixsite.com/home/corpus> <https://www.facebook.com/NEISUFRJ>, HistLing - <https://histling.letras.ufjf.br/index.php>, VALPB - <http://projetoalpb.com.br>, Peul - <https://peul.letras.ufjf.br/> <https://www.facebook.com/peulufjf/>, <https://www.instagram.com/peul.letras.ufjf/>, Gevar - @gevarufmufu, PorUs - <http://porus.sites.uff.br>, LinguaPOA - <https://www.ufrgs.br/linguapoa/>, www.ledoc.com.br, Discurso & Gramática <https://discursoegramaticablog.wordpress.com/corpus/>. (Acesso em: 09 out. 2021).

E entre os que ainda não são acessados via site (pelo menos até o momento), estão, por exemplo: PortVix (UFES), Falares sergipanos (UFS), Variação e Mudança no Português do Oeste de Santa Catarina (UFFS), Venda Nova do Imigrante – ES (UFES e PROFLETRAS-IFES), Português falado na região do norte de Minas (Unimontes), e Projeto Acomodação (Unicamp).

Além dos endereços já citados, vale considerar ainda outros que podem favorecer esse amplo conhecimento do Brasil, de sua comunidade linguística plurilíngue, tais como: Fonética-Fonologia de Thaís Cristófaros Silva (<https://fonologia.org/portugues/>); Dicionário Histórico do Português do Brasil (<https://dicionarios.fclar.unesp.br/dhpb/>), que possui “11.133.739 itens e 249.372 formas”; Projeto Atlas Linguístico do Brasil (<https://alib.ufba.br/>); Projeto Urbanização de Dialeto Rurais (dados de 1985) (<http://www.stellabortoni.com.br/>), um banco de dados com 271 páginas contendo transcrição das entrevistas (dados de fala).¹⁷⁵ Esses bancos de dados podem ser acessados sem ônus por pesquisadores ou professores que desejarem utilizar tais materiais para pesquisa, planejamento de aulas e produção de materiais didáticos. Infelizmente, esses dados ainda estão espalhados em diferentes sites. Seria importante termos uma plataforma como a que está sendo planejada (Plataforma Digital da Diversidade Linguística Brasileira) que reunisse as amostras em um único espaço virtual.

Cabe mencionar que os bancos de dados citados, vinculados a projetos de pesquisa desenvolvidos em universidades brasileiras, constituem uma amostra da vitalidade do patrimônio cultural da língua portuguesa e dos esforços científicos que lhes deram/dão existência e permanência (mesmo em situações adversas, decorrentes, por exemplo, da escassez de apoio financeiro). Servem também de ponto de referência e bússola para resgatar a memória das experiências alcançadas em pesquisa de campo, armazenamento e transcrição de materiais, para mapear o que pode ser aprimorado nos procedimentos implicados no trabalho de constituição, preservação, curadoria e difusão de bancos de (meta)dados e para formar e gerir novos quadros de pesquisadores para esse trabalho e novos (sub)sistemas de informações interconectados e interoperáveis em ambiente digital.

Em relação à preservação linguística, como apontou Freitag (2021), em sua fala “Por que precisamos de dados sociolinguísticos?”, para o já citado Festival do Conhecimento da UFRJ em 2021, a organização de um acervo de dados (socio)linguísticos de falares e escritas (sinalizadas ou não) do território brasileiro também

¹⁷⁵ E está em construção o Portal multimodal/multilíngue para o avanço da ciência aberta nas Humanidades, cuja versão preliminar está em <http://cienciaaberta.org>, conforme exposto em Berber Sardinha et al. (2021).

é uma ação de preservação linguística do Brasil, com os mesmos princípios de conservação e memória do que já ocorre em outras áreas, como a organização de Banco de Sementes.¹⁷⁶ Assim como um Banco de Sementes visa a armazenar sementes de modo a evitar que certas culturas desapareçam caso ocorra alguma praga ou desastre mundial, um repositório (digital) de dados linguísticos – como a *Plataforma da Diversidade Linguística Brasileira* – visa à salvaguarda de línguas e da memória de um povo de uma dada época, ou seja, no caso de alguma cultura/língua ser destruída, haverá sempre uma amostra de sua existência.

Tal acervo serve, ainda, como uma espécie de fiel da balança para educação linguística, educação patrimonial, ensino de língua, ainda mais num contexto em que se perspectivam variabilidade, dinamicidade, multimodalidade e multissemioticidade como propriedades basilares à concepção de língua, inclusive em orientações oficiais de ensino, como a Base Nacional Comum Curricular (BNCC, BRASIL, 2018), um documento de caráter normativo que define um conjunto de aprendizagens e competências essenciais que todos os alunos devem desenvolver ao longo das etapas e modalidades da Educação Básica no Brasil. A BNCC orienta em seu texto que as aulas de Língua Portuguesa, nas escolas brasileiras, devem ser ministradas evidenciando-se a variação linguística que existe em qualquer comunidade linguística, pois só assim o estudante poderá entender que uma língua é marcada por formas diferentes de se falar/escrever.

Para além do ensino de língua portuguesa como língua materna, um acervo de dados com amostras de diferentes falares brasileiros também pode ser utilizado no ensino-aprendizagem de português como língua não materna (ou português para estrangeiros, PLE), pois possibilitará o planejamento de aulas e a produção de materiais didáticos que evidenciem aos estudantes de PLE as características do português brasileiro e sua ampla variedade. O reconhecimento da diversidade linguística brasileira deve também estar presente na formação inicial e continuada dos professores de português (materna e não materna), pois, como afirmam Barbosa e Freire (2020, p. 654):

¹⁷⁶ Como exemplos de Banco de Sementes, há, no interior de uma montanha do arquipélago de Svalbard, na Noruega, o Silo Global de Sementes (Banco Mundial de Sementes), uma espécie de “cofre” de sementes, planejado para resistir a catástrofes climáticas e explosões nucleares, onde são depositadas amostras de diferentes espécimes vegetais. Outro exemplo são o Banco de Sementes A.L. Belo Correia, do Museu Nacional de História Natural e da Ciência de Lisboa, em Portugal e o Banco de Sementes da EMBRAPA (Empresa Brasileira de Pesquisa Agropecuária), no Brasil, ambos enviam amostras de sementes ao Banco Mundial (Fonte: <https://www.embrapa.br/busca-de-noticias/-/noticia/49411189/material-genetico-brasileiro-segue-para-deposito-no-banco-de-sementes-da-noruega>).

[...] o professor de língua portuguesa deve, entre outras ações: (i) reconhecer e trabalhar, em sala de aula, as características e as peculiaridades do Português Brasileiro; e (ii) tornar seus alunos mais sensíveis à diversidade cultural e linguística, incluindo a presente em sala de aula (FARACO, 2008). Assim, diante dessa problemática, parece-nos urgente refletir sobre a diversidade linguístico-cultural da língua portuguesa no Brasil e no mundo e relacioná-la ao ensino, sobretudo na formação do (futuro) professor.

Por fim, o acesso ao letramento digital também é uma das contribuições de repositório digital; pois é um aliado no enriquecimento cultural, apropriação de conhecimento científico e digital, além de aliado na construção de respeito linguístico e de autonomia dos sujeitos envolvidos num processo de aprendizagem ativa, pois requer/mobiliza a participação da sociedade (científica e não científica). E vale lembrar que: “Os letramentos digitais são resultantes das transformações e avanços tecnológicos, são marcados pelo multiculturalismo, pelo plurilinguismo, pela multisemiotividade” (SANTOS; GROSS; SPALDING, 2017, p. 128).

DISCUSSÃO

Ciência sociolinguística e Educação abertas e cidadãs são rumos com potencial de promover (i) sensibilização para a relação entre indivíduos e sociedade(s); (ii) partilha e coconstrução de conhecimentos e de vias para vencer desafios; (iii) ativismo (socio)linguístico; e (iv) estruturação de uma logística de trabalho em rede de cooperação e parceria. Investir em projetos e políticas nesse sentido é uma demanda na era da Ciência Aberta (<https://en.unesco.org/science-sustainable-future/open-science>):

In the context of pressing planetary and socio-economic challenges, sustainable and innovative solutions require an efficient, transparent and vibrant scientific effort - not only stemming from the scientific community, but from the whole society. The recent response of the scientific community to the COVID-19 pandemic has demonstrated very well, how open science can accelerate the achievement of scientific solutions for a global challenge.

The Open Science movement has emerged from the scientific community and has rapidly spread across nations, calling for the opening of the gates of knowledge. Investors, entrepreneurs, policy makers and citizens are joining this call.¹⁷⁷

¹⁷⁷ No contexto de desafios planetários e socioeconômicos urgentes, soluções sustentáveis e inovadoras exigem um esforço científico eficiente, transparente e vibrante – proveniente não apenas da comunidade científica, mas de toda a sociedade. A recente resposta da comunidade científica à pandemia COVID-19 demonstrou muito bem como a ciência aberta pode acelerar a obtenção de soluções científicas para um desafio global.

O movimento Open Science emergiu da comunidade científica e se espalhou rapidamente

O recurso a materiais que registrem a heterogeneidade sociocultural e linguística no Brasil via uma plataforma digital pode promover inclusão e justiça sociais. Ao serem explorados em diferentes ações e interações que ocorram nos mais diversos ambientes de ensino e aprendizagem, os materiais nessa plataforma poderão desfazer o mito do monolinguismo, preconceitos linguísticos e estereótipos quanto à comunidade linguística brasileira e, então, promover respeito linguístico. O acesso a dados autênticos de diferentes regiões brasileiras e suas comunidades de fala propiciará conhecer os modos de viver e de conceptualizar o mundo e as relações de pertencimento e de identidade de seus sujeitos. E isso só pode favorecer empreendimentos de pesquisa que visem à análise contrastiva de línguas e variedades, como o de comparação de línguas românicas (no Projeto VariaR), centrada em dados da realidade do uso. É, por fim, um caminho para se (re)descobrir o Brasil ou espaços de interação e de pertencimento que nele estão fertilmente configurados e às vezes, ou sequer são percebidos, ou são sub-representados. É preciso investir em reunir bons dados, autênticos e representativos da comunidade linguística brasileira, para ensinar Português perspectivando essa comunidade de milhões de falantes, naturalmente sem perder de vista que qualquer descrição de língua é sempre feita com base numa amostra, num recorte, que, queremos, atendam aos princípios FAIR: a padrões de encontrabilidade, acessibilidade, interoperabilidade e reusabilidade.

A já clássica afirmação de Labov (1994, p. 11) no labor da linguística histórica sobre “a arte de fazer bom uso de dados ruins” pode ser estendida a todos os conjuntos de dados de quem trabalha com descrição e análise linguística: nem sempre dispomos de dados de qualidade (dados linguísticos coletados em laboratório vs. dados de campo), e nem sempre os dados são facilmente acessíveis (organização em repositórios assistemáticos, sem ferramentas de busca consolidada vs. anotação e alinhamento dos dados e busca automática). Ademais, se no início das pesquisas na área de linguística até final do século passado, as limitações para constituir amostras de dados, especialmente de fala espontânea, eram decorrentes da tecnologia ainda incipiente, hoje, as limitações são decorrentes de restrições impostas por comitês de ética em pesquisa. E, ainda, a política de transparência com dados abertos requer condições de armazenamento e disponibilização, e novos desafios precisam ser superados, como os custos de manutenção, o sigilo e direito de uso dos dados, dentre outros. Dados linguísticos são base para desenvolvimento de tecnologias assistivas, o que coloca a constituição e a gestão de dados linguísticos em alinhamento com as áreas estratégicas para a ciência no Brasil (FREITAG et al., 2021, p. 3).

pelas nações, clamando pela abertura das portas do conhecimento. Investidores, empresários, formuladores de políticas e cidadãos estão aderindo a esta chamada.

PALAVRAS FINAIS

Esperamos, com as informações aqui reunidas e com o projeto Plataforma da Diversidade Linguística Brasileira, mudar a imagem do Português do Brasil normalmente ancorada na “pseudo-centralidade” linguística da região sudeste como referência ao trabalho com Português. E, com a reconfiguração dessa imagem, esperamos que exemplares até então à margem ou fora da sala de aula passem a ter lugar devido e, assim, a alargar as possibilidades de ser, agir e interagir. Compreendemos que um repositório de coleções de dados é vital para referenciar o desenvolvimento sustentável e ético de pesquisas e descrições (socio)linguísticas.

REFERÊNCIAS

BARBOSA, Juliana Bertucci; FREIRE, Deolinda de Jesus. A diversidade linguística no ensino de português como língua adicional e língua estrangeira. In: *Estudos Linguísticos* (São Paulo, 1978), 49(2), 2020, p. 651–673. Disponível em: <https://webcache.googleusercontent.com/search?q=cache:1PJo5FkeOT0J:https://revistas.gel.org.br/estudos-linguisticos/article/view/2714+&cd=3&hl=pt-BR&ct=clnk&gl=br>. Acesso em 10 out. 2021.

BERBER SARDINHA, Tony; BRAIT, Beth; SOUZA-E-SILVA, Cecília; CAMARGO, Zuleica; SAPGNUOLO, Renata Lamberti; BRAZ, Arianne Alfonso Brogini. Portal multimodal/multilíngue para o avanço da ciência aberta nas Humanidades. *Cadernos de Linguística*, 2(4), 1-16. DOI 10.25189/2675-4916.2021.V2.N4.ID406 <https://cadernos.abralin.org/index.php/cadernos/article/view/406/533>.

BRASIL. Secretaria de Educação Fundamental. *Base Nacional Comum Curricular*. Brasília: MEC/SEF, 2018.

CULTURAS multilíngues apresentado por Lina Gabriela Cortés, Juliana Bertucci Barbosa, Marcia dos Santos Machado Vieira, Kainan-Yuuki Veronese Anastacio, Vaz Pinto Có [s.l.,s.n.], 2021. 1 vídeo (1h07min). Publicado pelo Museu da Língua Portuguesa (2021). Disponível em: <https://www.youtube.com/watch?v=1aeiXl88yuo&t=1216s>. Acesso em: 09 dez. 2021.

FREITAG, Raquel Meister Ko *et al.* Desafios da gestão de dados linguísticos e a ciência aberta. In: *Cadernos de linguística*. Campinas, SP. v. 2, n. 1 (jan. 2021), p. 1-19.

FUTUROS possíveis para dados sociolinguísticos apresentado por Raquel Meister Ko Freitag, Juliana Bertucci Barbosa, Marcos Luiz Wiedemer, Marcia

dos Santos Machado Vieira [s.l.,s.n.], 2021. 1 vídeo (2h02min). Publicado pelo Festival de Conhecimento da UFRJ (2021) Disponível em: <https://www.youtube.com/watch?v=ZrZxsd5QQns>. Acesso em: 23 jul. 2021.

MACHADO VIEIRA, Marcia dos Santos; WIEDEMER, Marcos Luiz; FREITAG, Raquel Meister Ko; BARBOSA, Juliana Barbosa; PERES, Edenize Ponzo; MOLLICA, Maria Cecília de Magalhães Mollica. Plataforma da Diversidade Linguística Brasileira. Projeto apresentado à Pró-Reitoria de Pós-Graduação e Pesquisa da UFRJ e à Fundação Universitária José Bonifácio, em razão do Edital BNDES - Chamada Pública para seleção de propostas no âmbito da iniciativa Resgatando a História N. 01/2021, agosto de 2021.

MACHADO VIEIRA, Marcia dos Santos; WIEDEMER, Marcos Luiz; FREITAG, Raquel Meister Ko; BARBOSA, Juliana Barbosa. Mapeamento de Bancos de Dados (Socio)linguísticos no Brasil. Projeto desenvolvido pelo GT de Sociolinguística da ANPOLL e pela Comissão da Área de Sociolinguística da ABRALIN, junho de 2021.

SANTOS, Áurea Maria Brandão; GROSS, Letícia Granado; SPALDING, Marcelo. Conexões entre letramento digital e literatura digital. *Linguagem em foco*. Revista do Programa de Pós-Graduação em Linguística Aplicada da UECE, v. 9, n. 1, Fortaleza: EdUECE, 2017. Volume Temático: Novas Tecnologias e Ensino de Línguas.

