

REGISTRO DE PROJETO

ACERVOS DE DADOS ABERTOS À SOCIEDADE: MEMÓRIA LINGUÍSTICA E SOCIOCULTURAL E POTENCIALIDADE DE (RE)USO



EDITORES

- Miguel Oliveira, Jr. (UFAL)
- René Almeida (UFS)

AVALIADORES

- Tommaso Raso (UFMG)
- Tony Sardinha (PUC-SP)

SOBRE OS AUTORES

- Marcia dos Santos Machado Vieira
Conceptualização, Análise Formal,
Aquisição de Financiamento,
Metodologia, Administração do
Projeto, Supervisão, Validação,
Visualização e Escrita.
- Juliana Bertucci Barbosa
Conceptualização, Análise Formal,
Aquisição de Financiamento,
Metodologia, Administração do
Projeto, Supervisão, Validação,
Visualização e Escrita.
- Raquel Meister Ko Freitag
Conceptualização, Análise Formal,
Aquisição de Financiamento,
Metodologia, Administração do
Projeto, Validação, Visualização
e Escrita.
- Maria Manuel Borges
Conceptualização, Escrita.
- Ana Lígia Silva Medeiros
Conceptualização, Escrita.

DATAS

- Recebido: 06/10/2021
- Aceito: 08/01/2022
- Publicado: 24/01/2022

COMO CITAR

MACHADO VIEIRA, M.S.; BARBOSA, J.B.; FREITAG, R.M.K.; BORGES, M.M.; MEDEIROS, A.L.S. (2021). Acervos de dados abertos à sociedade: memória linguística e sociocultural e potencialidade de (re)uso. *Cadernos de Linguística*, v. 2, n. 1, e607.

Marcia dos Santos MACHADO VIEIRA

Universidade Federal do Rio de Janeiro (UFRJ)

Juliana Bertucci BARBOSA

Universidade Federal do Triângulo Mineiro (UFTM)

Raquel Meister Ko FREITAG

Universidade Federal de Sergipe (UFS)

Maria Manuel BORGES

Universidade de Coimbra (UC)

Ana Lígia Silva MEDEIROS

Fundação Casa de Rui Barbosa (FCRB)

RESUMO

Este artigo apresenta a proposta de uma plataforma digital nacional de arquivos de dados linguísticos e metadados interconectados para suporte às investigações linguísticas, com amostras de textos falados, escritos e/ou sinalizados das diferentes regiões brasileiras. Tal plataforma brasileira digital de bancos de dados linguísticos atende as demandas atuais da Ciência Aberta, que visam ao compartilhamento de informações, seu uso e reuso e (inter)ações coletivas não apenas dentro da comunidade científica, mas também para além dela, atingindo toda a sociedade. Além disso, evidenciamos como um repositório on-line

possibilita a criação de redes de cooperação entre os linguistas atuantes em território nacional, contribui para a formação inicial e continuada de novos pesquisadores, promove a Educação Patrimonial e a difusão e popularização da área nos mais diversos espaços.

ABSTRACT

This article presents the proposal of a national digital platform of linguistic data files and interconnected metadata to support linguistic investigations, with samples of spoken, written and/or signed texts from different Brazilian regions. Such a digital Brazilian platform of linguistic databases meets the current demands of Open Science, which aim at sharing information, its use and reuse and collective (inter)actions not only within the scientific community, but also beyond it, reaching the entire society. In addition, we show how an on-line repository enables the creation of cooperation networks between linguists working in the national territory, contributes to the initial and continued training of new researchers, promotes Heritage Education and the dissemination and popularization of the area in the most diverse spaces.

PALAVRAS-CHAVE

Gerenciamento de Dados Linguísticos Brasileiros; Sustentabilidade; Inovação Científica; Ciência Aberta.

KEYWORDS

Brazilian Linguistic Data Management; Sustainability; Scientific Innovation; Open Science.

INTRODUÇÃO

O desenvolvimento sustentável de pesquisas sociolinguísticas exige cada vez mais recursos, especialmente na era das humanidades digitais, com sistemas de comunicação hiperconectados e com tecnologias que rapidamente se modificam e transformam comportamentos (inclusive, linguísticos). Pensando na sustentabilidade das pesquisas sociolinguísticas em médio prazo, apresentamos a proposta de uma plataforma digital brasileira de arquivos de dados linguísticos e metadados interconectados, para suporte às investigações sociolinguísticas, com amostras de textos falados, escritos e/ou sinalizados das diferentes regiões brasileiras, com curadoria científica de uma comissão composta por linguistas de todas as regiões brasileiras. Embora pensada para atender a uma demanda específica de uma área, a disponibilização de amostras de língua em uma plataforma pode servir a outros propósitos e a outros públicos.

O desenho da plataforma foi apresentado, em agosto de 2021, no projeto intitulado “Plataforma da Diversidade Linguística Brasileira”, no escopo do GT de Sociolinguística da ANPOLL (Associação Nacional de Pós-Graduação em Letras e Linguística), mas contando com a parceria de pesquisadores da área de ciências da informação e da informática. A proposta da plataforma decorre do desenvolvimento de ações do GT de Sociolinguística da ANPOLL, como o projeto coletivo “SocInTec Brasil – Sociolinguística em (inter)ação: tecnologias sociais para o desenvolvimento da sustentabilidade no Brasil”, e o projeto “Mapeamento de Banco de Dados Sociolinguísticos”, desenvolvido sob as coordenações do GT de Sociolinguística (triênio de 2018-2021) e da Comissão Científica da Área de Sociolinguística da ABRALIN (Associação Brasileira de Linguística).

Uma plataforma/repositório nacional digital de bancos de dados linguísticos¹ atende as demandas atuais da Ciência Aberta, que visam ao compartilhamento de informações e (inter)ações coletivas não apenas dentro da comunidade científica, mas também para além dela, isto é, envolvendo a sociedade. A Ciência consiste tanto no corpo de conhecimento acumulado quanto no processo pelo qual o conhecimento é desenvolvido (GREEN, 2017; VICENTE-SAEZ; MARTINEZ-FUENTES, 2018). O processo nuclear da Ciência é tornar o novo conhecimento disponível à comunidade científica para que possa ser discutido, validado, objeto de revisão ou fornecer novos modos de interpretar os dados. Dependemos, assim, do passado para projetar e desenhar o futuro e dispomos de ferramentas que nos permitem fazê-lo a uma escala e com um grau de eficiência impossível até agora, pelo fato de dispormos de redes mundiais de informação, como a Internet e a *World Wide Web*.

1 Banco de dados, neste artigo, é o conjunto das várias coleções brasileiras (*corpora*) de textos falados, escritos e/ou sinalizados a serem reunidas e constituídas. O repositório em si abrigará outros materiais e informações.

É precisamente essa disponibilidade de resultados, dados e ideias que a Ciência Aberta procura garantir, o que implica a abertura do ciclo de informação científica ao longo de todo o processo. Facilitar o acesso ao trabalho de investigação não apenas difunde e estimula a criação e partilha de novo conhecimento, como potencializa o seu impacto junto da comunidade em que se insere, além da sociedade em geral, e estimula o seu uso promovendo novos olhares e nova geração de conhecimento.

É este fator inclusivo dos novos modos de fazer ciência uma marca distintiva deste novo paradigma da Ciência Aberta, que elege a abertura e a ética da investigação como valores principais. Esse movimento (de Ciência Aberta²) vem modificando sensivelmente o desenvolvimento da área de pesquisa científica. Uma das áreas que mais se beneficiou foi a de património, pois se ampliaram as possibilidades de acesso aos acervos sob a guarda das instituições de memória. Durante muito tempo estas instituições se defrontaram com a dicotomia entre a necessidade de preservar as coleções e, ao mesmo tempo, possibilitar o acesso aos pesquisadores dos documentos relevantes para seus estudos (VICENTE-SAEZ; MARTINEZ-FUENTES, 2018).

Com o avanço das políticas de digitalização de acervos e da organização em repositórios, ampliaram-se significativamente as possibilidades de pesquisa, bem como o uso e reuso de seus resultados. Porém, a intensificação da utilização das tecnologias de comunicação e informação exigiu uma série de novos procedimentos para que todas as etapas do ciclo de uma política de digitalização pudessem ser contempladas, desde o planejamento, a captura, o armazenamento, a organização e o (re)uso dos documentos, entre outros desafios. Esses procedimentos incluem ciclos complexos que exigem atenção para que a tecnologia seja utilizada em benefício do pesquisador. Um exemplo disso, segundo Medeiros (2012) e Medeiros et al (2017), é o trabalho realizado junto aos acervos arquivísticos, bibliográficos e museológicos da Fundação Casa de Rui Barbosa (FCRB), no Rio de Janeiro, que culminou na organização do Repositório Rui Barbosa de Informações Culturais (RUBI), na digitalização da coleção de cordel e do Portal de Crônicas Brasileiras, esse último em parceria com o Instituto Moreira Salles.

Tal ação realizada junto à FCRB é um exemplo de que é possível, mesmo nas áreas de Ciências Humanas e Linguística, participar desse movimento da Ciência Aberta, que contribui não só para o compartilhamento de informações com a sociedade, mas também a preservação da memória. É seguindo essa demanda urgente e exemplos como o da FCRB, que apresentamos projetos relacionados aos acervos de dados abertos linguísticos.

2 Por exemplo, "Portal multimodal/multilíngue para o avanço da ciência aberta nas Humanidades" (BERBER-SARDINHA et al, 2021).

Na área de Linguística, a questão dos repositórios para arquivos de dados linguísticos tem entrado na pauta de discussão em diferentes ações do evento ABRALIN AO VIVO³, e, na área da Sociolinguística em específico, o simpósio “Gestão de dados linguísticos” (FREITAG et al, 2021), que, a partir de práticas e experiências em nível individual ou de grupos de pesquisa que, reunidas, evidencia a urgência de uma ação de natureza coletiva, a criação e manutenção de repositórios de dados. Esta discussão está presente no GT de Sociolinguística da ANPOLL há algum tempo, como em simpósio específico sobre esse tema no 1º Fórum Internacional em Sociolinguística, em 2019, o painel “Futuros possíveis para dados sociolinguísticos” no Festival de Conhecimento da UFRJ, em 2021, e, desde o início de 2021, no projeto coletivo “SocInTec Brasil – Sociolinguística em (inter)ação: tecnologias sociais para o desenvolvimento da sustentabilidade no Brasil”.

Tendo em vista o cenário plurilíngue e pluricultural do Brasil, este projeto visa a criar redes de cooperação entre os sociolinguistas atuantes em território nacional e contribuir para a formação inicial e continuada de novos pesquisadores, para fomento e suporte à pesquisa de orientação sociolinguística, para o desenvolvimento de tecnologias sociais, de arranjos produtivos e de espaços de feedback e de escuta de demandas de diferentes comunidades (que se valem de linguagem, línguas e variedades de língua, nas suas mais diversas rotinas de interação ou de trabalho), bem como para a promoção, difusão e popularização da área nos mais diversos espaços.

As tecnologias sociais a serem desenvolvidas pelo projeto visam a otimizar a formação (treinamento de equipes, capacitação e atualização teórico-metodológica), o compartilhamento de amostras linguísticas e de informações sociolinguísticas, bem como sua aplicação e disseminação às mais diversas áreas do conhecimento na sociedade (como, por exemplo, a interface Sociolinguística e Educação), entre outras ações. A nucleação desta rede permitirá a difusão do conhecimento sociolinguístico produzido, com a inserção em políticas públicas e a promoção do desenvolvimento sustentável e do respeito aos direitos humanos, especialmente no campo dos direitos linguísticos.

O desenvolvimento sustentável está ligado a parte de um Protocolo Internacional da Assembleia Geral da Organização das Nações Unidas (ONU) no qual o Brasil assumiu o compromisso de implementar a Agenda 2030 para o Desenvolvimento Sustentável. O intuito do GT de Sociolinguística é desenvolver tecnologias sociais orientadoras para planejamento de ações e políticas públicas inclusivas, capazes de levar o Brasil ao efetivo alcance do desenvolvimento sustentável na área de sociolinguística. Temos como visão estratégica o compromisso com a disseminação de formação de recursos humanos

3 Por exemplo: “Grandes projetos em rede: NURC, Gramática do Português Falado e ALiB” (2020), “Grandes Projetos da Linguística em rede: Tycho Brahe, PROHPOR e PHPB” (2020), “Archiving and Language Documentation” (2020), “Sociolinguística Histórica no Brasil: caminhos e desafios” (2020).

altamente qualificados para o respeito aos direitos humanos e desenvolvimento sustentável, mobilizando a língua enquanto patrimônio.

Apesar dessas movimentações, ainda não há iniciativas com resultados efetivos. Freitag et al. (2021) destacam que “as associações de pesquisa teriam um papel seminal na execução dessas ações, por permitirem visualizar a abrangência e qualidade na distribuição dos dados.” Isso requer o mapeamento dos bancos brasileiros de dados linguísticos, para traçar um panorama de quantos e quais são os bancos de dados existentes, e qual a natureza dos dados que os compõem.

O projeto Mapeamento de Bancos de Dados Sociolinguísticos no Brasil⁴ recebeu notícia da existência de pelo menos 24 bancos de dados (socio)linguísticos no Brasil, organizados por parâmetros diferentes. Os bancos brasileiros até então reportados no mapeamento dizem respeito a dados de variedades da língua portuguesa (apenas uma das línguas usadas no Brasil a lhe servir de referência cultural, embora seja a oficial) e reúnem principalmente dados de fala. Não há informação (ou não foi informado até o momento), dentro da rede de contatos da área de Sociolinguística, de banco de dados de língua de sinais, língua indígena, de imigração, de herança, afro-brasileira. Tal resultado evidencia a carência de informações sobre coleções de dados representativas da diversidade linguística brasileira, além da falta de um espaço que reúna informações sobre as que já existem e estão espalhadas pelo Brasil.

Outro aspecto que chama a atenção é a concentração de bancos de dados sociolinguísticos na faixa litorânea brasileira e em torno de instituições universitárias, por conta de geralmente sua constituição estar atrelada a pesquisas de programas de pós-graduação (Figura 1).

4 O projeto Mapeamento de Bancos de Dados Sociolinguísticos no Brasil foi executado conjuntamente pela coordenação do GT de Sociolinguística da ANPOLL (triênio 2018-2021) e pela coordenação da Comissão da Área de Sociolinguística da ABRALIN com o objetivo de i) dar visibilidade às informações sobre os bancos de dados; ii) desenvolver estratégias coletivas de transcrição de dados, configuração, reposição e continuidade de mapeamentos, planejamento, gestão e preservação de registros/coleções de dados; iii) criar uma rede integrada de ações de pesquisas e de pesquisadores; iv) conceber um catálogo nacional de informações e, possivelmente, um repositório-referência nacional de bancos de dados. Especificamente quanto ao mapeamento, os pesquisadores da área de Sociolinguística foram consultados para noticiar a existência de bancos de dados.



Figura 1. Coordenadas geográficas de Bancos de Dados (socio)linguísticos no Brasil⁵. Fonte: MACHADO VIEIRA, WIEDEMER, BARBOSA, 2021.

O mapeamento realizado até o momento é revelador do quadro de lacunas em termos de representatividade das normas faladas, escritas e sinalizadas no Brasil para além das que se apresentam na faixa litorânea do território brasileiro. Esse quadro evidencia, por um lado, a necessidade de constituição, no que diz respeito à língua oficial usada no país, de acervos de dados do Português Brasileiro por parâmetros comparáveis em todo o território brasileiro (modalidade expressiva é apenas um deles). Evidencia, por outro lado, a necessidade de obter informações, a partir de interações com outras áreas da pesquisa linguística e com outras instituições, como o IPHAN (Inventário Nacional da Diversidade), por exemplo, sobre acervos brasileiros de outras línguas usadas no país⁶ e investir em uma plataforma que sirva de referência de catalogação, registro e/ou abrigo amplos de materiais que capturem línguas usadas no Brasil (estimadas em mais de 250) e que conte com curadoria técnica transdisciplinar.

Em continuidade a essa agenda, a mesa “Acervos de dados abertos à sociedade: memória linguística e sociocultural e potencialidade de (re)uso”, integrante da programação do InterAb12 (2021), abriu o diálogo interdisciplinar entre áreas, em busca de respostas e estratégias sustentáveis para promover a aproximação das concepções da Ciência Aberta em relação ao armazenamento, reutilização e autoria da organização de

5 Informadas aos coordenadores do mapeamento até setembro de 2021.

6 Cabe informar que o formulário que propicia o mapeamento (<https://forms.gle/Be4MtJgrWvf5zE617>) tem sido amplamente divulgado desde o final de junho de 2021: entre os mais de cem membros do GT e a partir deles; no Congresso Internacional da ABrALIN (InterAb 12).

bancos de dados linguísticos a partir da urgência em reunir acervos de dados linguísticos espalhados pelo Brasil num espaço que sirva de referência a consulentes e a necessidade de viabilizar sua (re)utilização para a ciência, educação, tecnologia social de interação e comunicação, inovação científica, inteligência artificial, patrimônio histórico-cultural e artes. É, com esse horizonte em perspectiva, que se contextualiza o projeto “Plataforma da Diversidade Linguística Brasileira”.

1. PLATAFORMA DA DIVERSIDADE LINGÜÍSTICA BRASILEIRA

*“Para levantar una carga muy pesada,
es preciso conocer su centro.
Así, para que los hombres puedan embellecer
sus almas,
es necesario que conozcan su naturaleza”.*
Ergonáutica⁷.

Como mostramos, a salvaguarda e a promoção de documentação das línguas existentes no país são tarefas urgentes, uma vez que, via suporte digital ou não-digital, qualquer divulgação, comunicação ou negociação de interesses perpassa a linguagem, requerendo, para ser persuasiva, diagnóstico e conhecimento das identidades do público-alvo, que é afetado e também se (re)configura via linguagem. Ademais, qualquer descrição de linguagem na ciência depende de documentação. A documentação da diversidade linguística contribui para vários campos do saber, bem como para a otimização da atuação e do diálogo desses campos em comunidades brasileiras.

O projeto “Plataforma da Diversidade Linguística Brasileira” toma por objeto o patrimônio linguístico brasileiro, patrimônio imaterial, a ser capturado contemporaneamente em cada região do país e a ser catalogado por meio de parâmetros nacionais comuns: mapeamento e documentação digital de coleções de dados linguísticos brasileiros que permitam conservar e valorizar a memória sociocultural, afetiva e linguística registrada em amostras faladas, escritas e sinalizadas, bem como servir de referência, nacional e internacional, a diferentes públicos (amplo, escolar e científico).

Inexiste um repositório de abrangência nacional de coleções de dados e metadados sociolinguísticos parametrizados (correspondendo a atributos sociolinguísticos comuns e, inclusive, atendendo às normativas de pesquisas com seres humanos, de CEP-CONEP), digitalizados e disponíveis, em interface amigável, que sirva de referência a consulta on-

⁷ “Fragmento de un poema que escribí siendo estudiante de ingeniería, cuyo tema era el navegar por dentro el alma de la ingeniería del futuro, siendo yo un “ergonauta” de mi profesión transformada hacia el conocimiento de la naturaleza humana.” (no prefácio de Rolf Behncke C. na obra *El árbol del conocimiento: las bases biológicas del entendimiento humano*, de Humberto Maturana R. e Francisco Varela G. (2009: p. IX); <http://repositorio.uchile.cl/handle/2250/119932>)

line e buscas por conteúdos anotados. Na era das humanidades digitais, é fundamental contar com *Big Data* para a sustentabilidade e o desenvolvimento de análises estatísticas comparáveis em várias áreas, tanto geográficas quanto científicas.

A organização digital ajuda a resolver essa carência e a promover redes de colaboração interinstitucionais para conceber uma arquitetura de acervos de memória sociolinguística que conte com pesquisa de campo a ter impacto em diferentes comunidades (via engajamento de profissionais e serviços de cada região), áreas do saber/fazer (por implicar a interação de expertises diversas). Em especial, destacamos que o uso e o reuso de dados decorrentes de documentação sociolinguística pode ter muito mais impacto na sociedade considerando que materiais podem circular em espaços de perfil museográfico, subsidiar a produção de materiais didático-pedagógicos (em contextos de língua materna e não-materna) que contemplem a ampla gama de normas linguísticas brasileiras, e promover a cultura brasileira. O acesso a esse material pode auxiliar no desenvolvimento de estratégias de promoção ao respeito linguístico às diferenças entre comunidades e indivíduos, conduzindo à relações entre linguagem e poder socialmente justas e cientificamente sustentáveis e que ou barrem apropriação dessas diferenças para perpetuação de estereótipos ou ativem o bom senso crítico de quem os (re)produz sobre as consequências inclusive no imaginário coletivo nacional e internacional sobre o Brasil.

É também uma ação que tem repercussão internacional, haja vista a existência de programa no Brasil (Programa de Leitorado da CAPES) direcionado à promoção de línguas e culturas brasileiras e programas no exterior voltados para repositórios de línguas no mundo (<https://glottobank.org/>, por exemplo) e para movimentos políticos de difusão de línguas (<https://www.francophonie.org/>, por exemplo). A constituição de um repositório brasileiro tem o potencial de servir de referência ao nosso patrimônio linguístico dentro e fora do Brasil.

1.1. OBJETIVOS DA PLATAFORMA

Entendemos que uma plataforma da diversidade linguística brasileira precisa valorizar como patrimônio: a) a catalogação nacional de iniciativas individuais e coletivas de pesquisa de campo e de administração de acervos de dados linguísticos espalhados pelo Brasil e constituídos segundo parâmetros diversos; b) a ferramenta digital para pesquisa e acesso ao repositório de coleções de dados sociolinguísticos.

Como uma das premissas da sociolinguística é a de que as línguas estão sempre em mudança, o processo de alimentação de uma “Plataforma da Diversidade Linguística Brasileira” é cíclico e contínuo (Figura 2).



Figura 2. Etapas de *input* e *output* da Plataforma da Diversidade Linguística Brasileira. Fonte: MACHADO VIEIRA, WIEDEMER, FREITAG, BARBOSA, PERES e MOLLICA, 2021.

A constituição de um catálogo nacional da diversidade linguística brasileira e uma ferramenta digital para acesso amigável a esse acervo catalogado, demanda uma gestão consorciada responsável por cinco áreas de ações: i) infraestrutura para armazenamento/preservação e suporte técnico-científico e tecnológico; ii) implementação, salvaguarda e restauração de amostras já constituídas; iii) implementação para a série histórica, com a organização de acervos para estudo diacrônico; iv) implementação de novas coleções constituídas com base em parâmetros nacionais comuns; e v) pesquisa, divulgação e popularização para acesso a público diverso e para multidirecionamento em termos de uso/reuso.

Para a ação de infraestrutura de armazenamento e suporte técnico-científico e tecnológico, é necessário investimento na formação e capacitação de recursos humanos especializados para atuarem na organização, recuperação, transcrição, etiquetagem e curadoria de coleções de dados. Nesta mesma direção, é preciso investir na formação de recursos humanos para a alimentação dos dados linguísticos e metadados a serem disponibilizados ao público.

A implementação quanto à série histórica demanda o desenvolvimento de arquitetura de software para permitir a participação ativa dos usuários da plataforma, que podem contribuir com amostras linguísticas, a serem coletadas on-line, juntamente com metadados de identificação que permitam indexar a amostra linguística não só no tempo e no espaço, como também socialmente. Paralelamente, é necessário também o desenvolvimento de arquitetura de software para disponibilização aos usuários da plataforma em tempo real dos resultados da coleta de amostras linguísticas on-line em projeção cartográfica e temporal, a partir dos metadados coletados. Este processo requer acompanhamento e supervisão de recursos humanos qualificados.

1.2. AÇÕES DE DIFUSÃO

A “Plataforma da Diversidade Linguística Brasileira” tem o potencial de subsidiar ações de difusão e popularização da diversidade linguística brasileira como elemento integrado a um ecossistema mundial de manifestações político-linguísticas e como patrimônio imaterial dos brasileiros.

O acesso ao conhecimento da diversidade linguística promove empoderamento social e efetiva participação cidadã, e colabora para a melhoria do ensino e respeito à diversidade. A possibilidade de participação dos usuários na constituição das amostras que alimentarão a plataforma, seja como amostras autênticas, seja por meio da indicação de materiais reveladores da diversidade linguística brasileira, enseja ao cidadão comum experimentar a ciência, apoderar-se de lugar (de analista) e/ou processo que é tipicamente assumido por sociolinguistas: focalizar uma ou mais comunidade(s) de fala/discurso, observar, formular pergunta e hipótese, cuidar de como experimentar/registrar o que percebe/observa, examinar o resultado, tecer generalizações e compartilhar o que foi apreendido nesse processo. Esta é uma oportunidade de formação para a ciência e de educação linguística, campo em que o Brasil ainda é muito carente de iniciativas.

As profundas desigualdades na distribuição das oportunidades educacionais e do conhecimento são ainda desafios grandes para a inclusão social em nosso país. Esta proposta contribui para que os dados linguísticos organizados e coletados ultrapassem os muros universitários e cheguem à comunidade em geral. E essas ações mobilizam outras relativas à (re)qualificação de recursos humanos e engajamento da população regional/local no propósito de preservação do patrimônio histórico, haja vista o interesse em abranger acervos de usos reais e singulares de línguas por todo o território nacional, bem como educação patrimonial.

Além disso, prevemos ainda outras ações de divulgação e popularização das ações do projeto, como por exemplo:

- Produção de cursos e materiais para treinamento de professores da educação básica e líderes comunitários para ações de valorização da diversidade linguística brasileira e mesmo de coparticipação no processo de constituição de novas amostras;
- Elaboração de mostras/feiras de ciências nas comunidades envolvidas;
- Produção de materiais com acessibilidade em libras e descrição em áudio;
- Treinamentos para líderes comunitários para administrarem recursos educacionais, bem como para serem nossos aliados na preservação de nosso patrimônio linguístico;
- Realização de treinamento que vise a contemplar as características identitárias do grupo ao qual o público-alvo pertence: (i) reconhecendo a realidade sociolinguística da

sala de aula e da comunidade onde está atuando; (ii) observando marcas de diversidade linguística regionais (urbano/rural, centro/periferia, nordestino/sulista, por exemplo), e sociais (maior ou menor domínio da norma culta em decorrência de fatores sociais como o nível socioeconômico da família, por exemplo) e de contato linguístico (comunidades bilíngues).

Tais ações podem permitir que qualquer indivíduo da sociedade participe como agente ativo de experiências que envolvem fenômenos científicos de forma responsável e motivadora, para que haja estímulo à compreensão de conceitos e à popularização da ciência. Ao mesmo tempo, este tipo de ação promove o engajamento de linguistas com a comunidade, para além da coleta de dados.

Grande parte das amostras linguísticas que alimentarão a “Plataforma da Diversidade Linguística Brasileira” são desenvolvidas por instituições de ensino superior da rede federal decorrentes dos programas de expansão e interiorização da universidade. Casqueiro, Iffi e Silva (2020) destacam que os municípios que receberam campi ou universidades novos tiveram redução no percentual de pobres, bem como o aumento da renda per capita e do percentual de ocupados com nível superior e com nível médio. Estes resultados, aliados às potencialidades de mudanças nos indicadores educacionais, reforçam a importância social e econômica desta proposta, que tem potencial de transformação da realidade sociopolítica e econômica com ciência, tecnologia e informação.

Uma logística de ações incluídas e capilarizadas por todas as regiões do país é importante base para a construção de sociabilidade, de civilização e de cidadania nos espaços de negociação de interesses individuais e coletivos. É também um meio de engajar pessoas de diferentes regiões brasileiras em prol da preservação do patrimônio linguístico.

1.3. PÚBLICO-ALVO

Destacamos três perfis de público para o alcance das ações da Plataforma da Diversidade Linguística Brasileira:

- a) **Público amplo:** exposição virtual e interativa da diversidade linguística brasileira, e, principalmente, por meio da plataforma/repositório on-line, a possibilitar a todos ver, ouvir e repetir diferentes falares do território nacional, em formato de exposição permanente, aberta e gratuita;
- b) **Público escolar:** repositório de pesquisa de amostras linguísticas representativas da diversidade regional e social brasileira, a fim de subsidiar a implementação da BNCC e do direito de aprendizagem à diversidade linguística, bem como servir de referência sobre a linguagem nacional a políticas educativas espalhadas no mundo e possibilitar

produção de materiais/recursos educacionais a partir dos falares representativos da diversidade brasileira presentes na plataforma a ser desenvolvida;

- c) **Público acadêmico:** repositório de amostras linguísticas preparadas conforme padrões internacionais de representação e armazenamento, para subsidiar pesquisas científicas e desenvolvimento tecnológico (processamento de linguagem natural). Os dados presentes na plataforma possibilitarão a continuidade de levantamento de características das línguas/variedades presentes em nosso território, incluindo regiões ainda não estudadas pelos linguistas e pesquisadores de outras áreas.

2. O PAPEL PEDAGÓGICO DA EDUCAÇÃO LINGUÍSTICA

“El origen de lo humano, tiene que haber ocurrido en un momento histórico en el que surge el lenguaje en la convivencia. **Si ustedes atienden a lo que hace el lenguaje, se darán cuenta que coordina conductas, relaciones, haceres, sentires, ideas.** ¿Qué tiene que pasar para que se den las coordinaciones? Tenemos que encontrarnos, si no nos encontramos conversando, ese fenómeno de la información no va a pasar, tenemos que transformarnos juntos y para eso los seres humanos tenemos que convivir y tenemos que querer convivir, tiene que gustarnos el convivir.”

“Somos generadores de los mundos que habitamos, y habitamos los mundos que hacen posibles que nosotros habitamos en ellos. No nos encontramos con algo independiente de nuestra historia, somos responsables de las consecuencias porque generamos los mundos que vivimos, de modo que podemos generar los mundos que queremos vivir.”
(Humberto Maturana, disponível em <https://www.uchile.cl/noticias/145130/profesor-humberto-maturana-tenemos-que-transformarnos-juntos>, acesso em 01/11/2021)

A educação linguística da sociedade brasileira ainda carece de investimentos. Começamos pela percepção expressa por vozes que estão fora do território teórico de Letras e Linguística e que, supomos, plenamente representadas neste excerto de crônica de Luana Génot:

“Quando vi o vídeo “Sudestino” no canal Porta dos Fundos, não me contive. Vi e revi em *looping*. Já passou da hora de nós, sudestinos, entrarmos na nossa caixinha e deixarmos de achar que o centro do Brasil é o Sudeste.

Precisamos parar de acreditar que o que se faz no Sudeste é a norma e que os demais são os “outros”, os diferentes, os que têm sotaque fofo, os que comem comidas estranhas, os que são do outro lado do Brasil.

“Meu” (gíria de sudestinos de Sampa), é óbvio que como indivíduos, somos múltiplos etnicamente e temos várias influências culturais ao longo da vida, mas precisamos reconhecer que, como em toda e qualquer região do Brasil e do mundo, compartilhamos hábitos que não podem ser negados e tampouco colocados como padrão absoluto.”

(...)

“Morando em um país de dimensões continentais, devemos ter mais dúvidas do que certezas quando o assunto é o que se faz em todas as regiões do Brasil fora do eixo Sul-Sudeste”. (Revista Ela, O Globo, 29 de agosto de 2021, p. 26; disponível em <https://oglobo.globo.com/ela/queridos-sudestinos-25175013>, acesso em 01/11/2021)

Soma-se à percepção de que é necessário “descobrir os Brasis que cabem dentro do nosso Brasil” e “navegar nesse Brasil”, a constatação de que também no espaço educacional reinam lacunas, padrões estimados com base na “pseudocentralidade do

Sudeste” (referências (entre aspas) feitas por Luana Génot) que acabam por, equivocadamente, servir de único referencial do Português do Brasil. Enfim, são alguns problemas a resolver.

Apesar dos esforços empreendidos por políticas públicas, a educação brasileira ainda enfrenta desafios a superar para a melhoria de seus indicadores de sucesso. As avaliações oficiais que envolvem o pleno domínio da Língua Portuguesa, como Avaliação Nacional de Alfabetização, Prova Brasil e PISA, vêm sistematicamente evidenciando o profundo fosso de desigualdade existente em nosso sistema educacional, marcado por assimetrias regionais e sociais. Essas assimetrias educacionais têm forte impacto no desenvolvimento econômico, interferindo na oferta de mão de obra qualificada, o que reduz a geração de emprego e de renda.

E, pelo viés das assimetrias regionais e sociais, é que consideramos a importância da diversidade linguística como patrimônio imaterial brasileiro. Políticas públicas promotoras da Educação enfatizam o papel central da diversidade linguística no ensino de Língua Portuguesa: o tratamento e o reconhecimento da diversidade linguística está presente nos documentos oficiais da educação brasileira, tais como os Parâmetros Curriculares Nacionais (PCN) (BRASIL, 1997), o Programa Nacional do Livro Didático (PNLD) (BRASIL, 2018) e, mais recentemente, a Base Nacional Comum Curricular (BNCC) (BRASIL, 2018).

A BNCC (BRASIL, 2018) assume uma concepção de língua dinâmica e multimodal:

A variedade de composição dos textos que articulam o verbal, o visual, o gestual, o sonoro, o tátil, que constituem o que se denomina multimodalidade de linguagens, deve também ser considerada nas práticas de letramento. A escola precisa, assim, comprometer-se com essa variedade de linguagens que se apresenta na TV, nos meios digitais, na imprensa, em livros didáticos e de literatura e outros suportes, tomando-as objetos de estudo a que os estudantes têm direito. (BRASIL, 2016, p. 87)

Ao mesmo tempo, enfatiza o desenvolvimento de senso de pertencimento e identificação com o contexto social:

O processo de descentralização, que caracteriza esse período de vida, amplia a capacidade dos/das estudantes, tanto de desenvolver sua autonomia, como de cultivar a alteridade. Desse modo, ao se potencializarem novas práticas nos diferentes componentes, também se criam condições para que os/as estudantes percebam a pluralidade sociocultural (artística, corporal e linguística) e as relações entre linguagens, identidades e pertencimentos, alargando as possibilidades de ser e de interagir com grupos sociais. (BRASIL, 2016, p. 325)

Estas premissas que norteiam a BNCC na área de Linguagens se materializam como direito de aprendizagem de todo o estudante brasileiro, assim enunciado para o Ensino Fundamental:

Compreender que a variação linguística é um fenômeno que constitui a linguagem, reconhecendo as relações de poder e as formas de dominação e preconceito que se fazem na e pela linguagem e refletindo sobre as relações entre fala e escrita em diferentes gêneros, assim como reconhecer e utilizar estratégias de marcação do nível de formalidade dos textos em suas produções. (BRASIL, 2016, p. 98).

Na prática, este direito de aprendizagem se reflete no desenvolvimento de competências e habilidades:

- Compreender a condição social e histórica da língua como polissistema dinâmico e variável;
- Refletir sobre a variação de registro e sobre a variação regional e social da língua. (BRASIL, 2016, p. 366);
- Refletir sobre o endereçamento dos textos e sobre as escolhas linguísticas adequadas à interlocução proposta. (BRASIL, 2016, p. 351)

Já no Ensino Médio, a BNCC continua enfatizando o respeito à diversidade linguística brasileira como direito de aprendizagem:

No Ensino Médio, aprofundam-se também a análise e a reflexão sobre a língua, no que diz respeito à contraposição entre uma perspectiva prescritiva única, que segue os moldes da abordagem tradicional da gramática, e a perspectiva de descrição de vários usos da língua. Ainda que continue em jogo a aprendizagem da norma-padrão, em função de situações e gêneros que a requeiram, outras variedades devem ter espaço e devem ser legitimadas. (BRASIL, 2018, p. 504).

Ainda mais explícita fica a importância da diversidade linguística brasileira quando são apresentadas as competências e habilidades para o Ensino Médio:

Compreender as línguas como fenômeno (geo)político, histórico, cultural, social, variável, heterogêneo e sensível aos contextos de uso, reconhecendo suas variedades e vivenciando-as como formas de expressões identitárias, pessoais e coletivas, bem como agindo no enfrentamento de preconceitos de qualquer natureza. (BRASIL, 2018, p. 494)

Em suma, a importância da diversidade linguística está explícita nos documentos norteadores da educação brasileira, no sentido de reconhecer que a variação e a diversidade são constitutivas das línguas humanas, que são historicamente situadas. Assim, o papel da escola diante da diversidade e da desigualdade é legitimar e propiciar conhecer todas as variedades linguísticas e combater preconceitos, de modo a promover amplo acesso a vozes/comunidades diversas e respeito por elas, bem como aproximar as pessoas. Por isso é importante conhecer as regularidades das variedades da Língua Portuguesa e dos valores sociais nelas implicados, bem como reconhecer que o domínio dos diferentes usos da linguagem amplia as possibilidades de participação social no exercício da cidadania. Então, por que ainda falhamos tanto com o ensino da Língua Portuguesa, apesar de todos os documentos norteadores preconizarem o tratamento da diversidade linguística brasileira na educação básica?

Apesar das políticas públicas preverem o tratamento da diversidade linguística, reconhecidamente patrimônio imaterial dos brasileiros, a sua implementação ainda é falha, por ausência de suporte pedagógico. Em que pese a existência de um Programa Nacional do Livro Didático (PNLD), o tratamento da diversidade linguística brasileira em nível educacional ainda carece de material de base, constituído dentro dos padrões científicos, e preparados para a apropriação pedagógica.

A “Plataforma da Diversidade Linguística Brasileira” possibilitará que, por exemplo, professores da Educação Básica façam uso desse grande banco de dados com amostras autênticas de língua. Por meio dela, podem elaborar atividades didáticas que mobilizem seus alunos e as respectivas famílias em favor da divulgação das riquezas culturais e linguísticas de sua região. É nesse sentido que a plataforma contribuirá para a plena implementação da BNCC de Linguagens, no que diz respeito a prover subsídios para garantir os direitos de aprendizagem relativos à diversidade linguística brasileira.

3. AGENDA DE TRABALHO

*¿Están las ciencias sociales, en particular la economía, las ciencias políticas y las ciencias de la educación fundadas en una adecuada comprensión de la naturaleza del proceso de aprendizaje humano, de lo que determina la diversidad de las conductas humanas?, y si no lo están ¿podrían llegar a estarlo?**

Em um cenário geral de escassez de recursos e sucessivos cortes e restrições às ciências humanas em especial, a “Plataforma da Diversidade Linguística Brasileira” é uma ação para sustentabilidade de Letras e Linguística, e para tal precisa mobilizar um grande grupo de agentes interessados. A iniciativa gestada no escopo do projeto “SocInTec Brasil – Sociolinguística em (inter)ação: tecnologias sociais para o desenvolvimento da sustentabilidade no Brasil”, do GT de Sociolinguística da ANPOLL, depois ampliada para a Comissão Científica da Área de Sociolinguística da ABRALIN, e agora em uma discussão ampliada envolvendo experiência de um projeto já realizado na Fundação Casa de Rui Barbosa (FCRB), reforça que é possível, mesmo nas áreas de Ciências Humanas e de Linguística, contribuir não só para o movimento da Ciência Aberta, como também contribuir para o compartilhamento de informações com a sociedade e a preservação da nossa memória linguística.

A ampliação da capilaridade da proposta, especialmente para o campo da popularização, alcançando espaços museográficos, como o Museu da Língua Portuguesa (2021) e o Museu Nacional da UFRJ, e outros, como Secretarias de Educação e ONGs, permite maior inserção social da área na comunidade, contribuindo para a educação linguística, ao dar ampla e significativa visibilidade a potencialidades de uso e reuso desse repositório de dados linguísticos em diferentes setores da sociedade.

Apresentado o desenho da proposta e as suas potencialidades e vantagens, a próxima etapa para viabilizar a implantação é a captação de recursos financeiros. Nesse sentido,

8 No início do prefácio de Rolf Behncke C. na obra *El árbol del conocimiento: las bases biológicas del entendimiento humano*, de Humberto Maturana R. e Francisco Varela G. (2009: p. IX); <http://repositorio.uchile.cl/handle/2250/119932>

como destacam Freitag et al (2021), a elaboração de um projeto coletivo para obtenção de apoio financeiro junto a órgãos de fomento precisa contar com o apoio formal de associações da área, como, por exemplo, a Associação Brasileira de Linguística (ABRALIN), para uma consolidação efetiva do patrimônio não de um grupo de pesquisadores, mas de uma instituição de pesquisadores, representativa de toda uma área, com garantias de continuidade e preservação, independentemente de pessoas. Afinal, estamos lidando com um patrimônio que ultrapassa as fronteiras da área da Sociolinguística. Língua enquanto patrimônio interessa não só a todas as áreas da Linguística, mas a toda a humanidade!

REFERÊNCIAS

- ARCHIVING and Language Documentation. Mesa-redonda apresentada por Ana Paula Brandão, Patience Epps, Susan Kung, Denny Moore, Zachary O'Hagan, Jorge Emilio Roses Labrada. [s.l, s.n], 2020. 1 vídeo (2h 21min). Publicado pelo canal da Associação Brasileira de Linguística. Disponível em: <https://www.youtube.com/watch?v=uQY4dZnQKds> Acesso em: 23 jul. 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. Ministério da Educação e do Desporto. *Parâmetros Curriculares Nacionais*. Brasília: MEC/SEF, 1997.
- BRASIL. Ministério da Educação. *PNLD 2019* Língua Portuguesa –guia de livros didáticos. Brasília: Ministério da Educação, Secretária de Educação Básica, 2018.
- BRASIL. Secretaria de Educação Fundamental. *Base Nacional Comum Curricular*. Brasília: MEC/SEF, 2018.
- CASQUEIRO, Mayara Lima; IRFFI, Guilherme Irffi.; SILVA, Cristiano da Costa da. A expansão das Universidades Federais e os seus efeitos de curto prazo sobre os Indicadores Municipais. *Avaliação: Revista Da Avaliação Da Educação Superior*, 25(1). 2020. Disponível em: <http://periodicos.uniso.br/ojs/index.php/avaliacao/article/view/3922> Acesso em: 22 out. 2021.
- FREITAG, Raquel Meister Ko et al. Desafios da gestão de dados linguísticos e a ciência aberta. In: *Cadernos de linguística*. Campinas, SP. Vol. 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 (2h 02min). Publicado pelo Festival de Conhecimento da UFRJ (2021). Disponível em: <https://www.youtube.com/watch?v=ZrXsd5QQns>. Acesso em: 23 jul. 2021.
- GESTÃO de dados linguísticos. Simpósio apresentado por Raquel Meister Ko Freitag, Elisa Battisti, Aluiza Araújo, landra Silva Coelho, Marco Antônio Martins, Marta Farias Sousa e Rodrigo Lopes [s.l.,s.n], 2020. 1 vídeo (2h 30min). Publicado pela Associação Brasileira de Linguística. Disponível em: <https://www.youtube.com/watch?v=S7YS57i7ogs>. Acesso em: 23 jul. 2021.
- GRANDES PROJETOS DA LINGUÍSTICA em rede: Tycho Brahe, PROHPOR e PHPB Mesa-redonda apresentada Charlotte Galves, Tânia Lobo, Sandro Morengo e Izete Coelho. [s.l, s.n], 2020. 1 vídeo (2h 4min). Publicado pelo canal da Associação Brasileira de Linguística. Disponível em: <https://youtu.be/gJgrqArFDlw> Acesso em: 23 jul. 2021.
- GRANDES PROJETOS EM REDE: NURC, Gramática do Português Falado e ALIB. Mesa-redonda apresentada por Dinah Callou, Ataliba de Castilho, Jacyra Mota e Sandro Marengo. [s.l, s.n], 2020. 1 vídeo (2h 08min). Publicado pelo canal da Associação Brasileira de Linguística. Disponível em: <https://www.youtube.com/watch?v=miDdAf7ICFA>. Acesso em: 23 jul. 2021.

GREEN, Cable. Open Licensing and Open Education Licensing Policy. In: Jhangiani, R S and Biswas-Diener, R. (eds.) *Open: The Philosophy and Practices that are Revolutionizing Education and Science*. London: Ubiquity Press, 2017, p. 29-41. DOI: <https://doi.org/10.5334/bbc.c>. License: CC-BY 4.0

MACHADO VIEIRA, M. dos S.; WIEDEMER, M. L.; BARBOSA, J. B. Projeto Mapeamento de Bancos de Dados Sociolinguísticos no Brasil. Subprojeto gerido pelas coordenações do GT de Sociolinguística da ANPOLL e da Comissão Científica da Área de Sociolinguística da ABRALIN e vinculado ao "Projeto SocInTec Brasil - Sociolinguística em (inter)ação: tecnologias sociais para o desenvolvimento da sustentabilidade no Brasil", desenvolvido desde fevereiro de 2021 no GT de Sociolinguística da ANPOLL. Brasil, junho de 2021.

MACHADO VIEIRA, M. dos S.; WIEDEMER, M. L.; FREITAG, R. M. Ko; BARBOSA, J. B.; PERES, E. P.; MOLLICA, M. C. de M. 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 No. 01/2021, agosto de 2021.

MEDEIROS, Ana Lígia Silva et al. Humanidades digitais na fundação casa de Rui Barbosa: um estudo aplicado de seu conceito. *Informação & Tecnologia*, v. 4, n. 2, p. 243-259, 2017. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/101598>. Acesso em: 10 set. 2021.

MEDEIROS, Ana Lígia Silva. Biblioteca Pública do século XXI. In: *CRB-8 Digital*, São Paulo, v. 5, n. 2, p. 49-55, dez. 2012. Disponível em: <http://revista.crb8.org.br>. Acesso em: 23 jul. 2021.

MUSEU DE LÍNGUA PORTUGUESA. Centro de Referência. 2021. Disponível em: <https://www.museudalinguaportuguesa.org.br/mlp/centro-de-referencia/>. Acesso em 15 ago. 2021.

SOCIOLINGUÍSTICA Histórica: caminhos e desafios. Mesa-redonda apresentada por Célia Regina dos Santos Lopes, Rosane de Andrade Berlink e Huda da Silva Santiago. [s.l., s.n], 2020. 1 vídeo (1h 49min 20s). Publicado pelo canal da Associação Brasileira de Linguística. Disponível em: <https://www.youtube.com/watch?v=nE4i53QDacE>. Acesso em: 23 jul. 2021.

VICENTE-SAEZ, Ruben; MARTINEZ-FUENTES, Clara. Open Science now: A systematic literature review for an integrated definition. *Journal of Business Research*, 88, 428-436. 2018.