

REGISTRO DE PROJETO

 OPEN ACCESS

EDITORES

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

AVALIADORES

- Adriana Lucena (UNP-UNINTA)
- Tiago Torrent (UFJF)

SOBRE OS AUTORES

- Tony Berber SARDINHA
Administração do Projeto, Curadoria de Dados, Conceptualização, Software, Aquisição de Financiamento, Supervisão, Escrita - Rascunho Original.
- Sandra MADUREIRA
Curadoria de Dados, Conceptualização, Software, Aquisição de Financiamento, Supervisão, Escrita - Análise e Edição.
- Beth BRAIT
Curadoria de Dados, Conceptualização, Aquisição de Financiamento, Supervisão, Escrita - Análise e Edição.
- Maria Cecília Perez de Souza-e-Silva
Conceptualização, Aquisição de Financiamento, Supervisão.
- Zuleica Camargo
Contribuiu igualmente com Afrânio Gonçalves Barbosa. Papéis: conceituação, metodologia.
- Renata Lamberti SPAGNUOLO
Curadoria de Dados, Conceptualização, Software, Aquisição de Financiamento, Supervisão.
- Arianne Alfonso Brogini BRAZ
Curadoria de Dados, Metodologia, Recursos.

DATAS

- Recebido: 31/07/2021
- Aceito: 30/08/2021
- Publicado: 11/09/2021

COMO CITAR

SARDINHA, T.B.; MADUREIRA, S.; BRAIT, B.; SOUZA-E-SILVA, M.C.P.; CAMARGO, Z.; SPAGNUOLO, R.L.; BRAZ, A.A.B. (2021). Portal multimodal/multilíngue para o avanço da Ciência Aberta nas Humanidades. *Cadernos de Linguística*, v. 2, n. 4, e406.

PORTAL MULTIMODAL/ MULTILÍNGUE PARA O AVANÇO DA CIÊNCIA ABERTA NAS HUMANIDADES

Tony Berber SARDINHA  

Pontifícia Universidade Católica de São Paulo (PUC-SP)

Sandra MADUREIRA  

Pontifícia Universidade Católica de São Paulo (PUC-SP)

Beth BRAIT  

Pontifícia Universidade Católica de São Paulo (PUC-SP)

Maria Cecília Perez de SOUZA-E-SILVA  

Pontifícia Universidade Católica de São Paulo (PUC-SP)

Zuleica CAMARGO  

Pontifícia Universidade Católica de São Paulo (PUC-SP)

Renata Lamberti SPAGNUOLO  

Pontifícia Universidade Católica de São Paulo (PUC-SP)

Arianne Alfonso Brogini BRAZ  

Pontifícia Universidade Católica de São Paulo (PUC-SP)

RESUMO

Embora práticas de ciência aberta tenham sido adotadas em diversos campos do conhecimento ao longo dos anos, em larga medida estão concentradas em campos como as ciências naturais, exatas e computação. Nas Ciências Humanas, a migração para a ciência aberta ainda é incipiente. Neste trabalho, apresentamos o Portal multimodal/multilíngue para o Avanço da Ciência Aberta nas

Humanidades, apoiado pelo CNPq por meio do Edital 25/2020, destinado a ajudar a preencher essa lacuna. O Portal prevê fornecer uma série de conteúdos relacionados à ciência aberta nas Humanidades, incluindo dados verbais, visuais, verbo-visuais e verbo-gestuais, por meio dos quais possam ser desenvolvidas pesquisas, nacionais e internacionais. O Portal foi concebido em torno de quatro vetores, a saber: acesso aberto, dados abertos, fonte aberta e ciência cidadã / humanidades cidadãs. O trabalho apresenta, em relação a cada um dos vetores, a problemática envolvida bem como as ações pretendidas. Além disso, são apresentados os princípios de design do Portal, visando à sua usabilidade e acesso. O movimento de abertura da ciência exige novas posturas e novas práticas, que o Portal pretende estimular, pois entende que esse movimento é em última análise responsável pela democratização da ciência.

ABSTRACT

Although open science practices have been widely used in several knowledge areas for many years, open science has had greater acceptance in fields such as the natural sciences, exact sciences and computing (STEM). In the Human Sciences, the migration to open science is still in its infancy. In this work, we present the project to build the Multimodal/multilingual Portal for the Advancement of Open Science in the Humanities, supported by CNPq through Call 25/2020, designed to help fill this gap. The Portal intends to provide a range of open science materials for the Humanities, including verbal, visual, verbal-visual and verbal-gesture data, to stimulate both national and international research in the Humanities. The Portal was designed around four vectors, namely: open access, open data, open source and citizen science/citizen humanities. This paper presents a discussion of the problems and planned actions for each of the vectors. In addition, the paper outlines the major design principles underlying the implementation of the Portal. Open science requires a new set of postures and practices, which the Portal intends to encourage. In line with open science, its ultimate goal is democratizing science.

PALAVRAS-CHAVE

Portal de Ciência Aberta; Acesso Aberto; Dados Abertos; Fonte Aberta; Humanidades Cidadãs.

KEYWORDS

Open Science Portal; Open Access; Open Data; Open Source; Citizen Humanities.

INTRODUÇÃO

Neste trabalho, apresentamos o Portal multimodal/multilíngue para o Avanço da Ciência Aberta nas Humanidades, apoiado pelo CNPq por meio do Edital 25/2020. O Portal está sendo desenvolvido pelo Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem (LAEL) da Pontifícia Universidade Católica de São Paulo, em conjunto com o Laboratório Integrado de Análise Acústica de Cognição (LIAAC-LAEL-PUCSP) e o Centro de Pesquisa, Recursos e Informação (CEPRIL). O projeto de desenvolvimento teve início no primeiro semestre de 2021, e a versão beta do Portal será lançada no primeiro semestre de 2022.

O objetivo geral do Portal é colaborar para a ciência aberta, constituindo um portal multimodal e multilíngue, de grande alcance nas Humanidades, que reúna e disponibilize, do ponto da tecnologia e da inovação, um amplo conjunto de dados verbais, visuais, verbo-visuais e verbo-gestuais, por meio dos quais possam ser desenvolvidas pesquisas, nacionais e internacionais, que atinjam pessoas em diferentes situações de trabalho e/ou em situação de vulnerabilidade, quer do ponto de vista físico, psicológico ou social.

Entendemos as Humanidades como uma área de conhecimento que abrange as variadas formas de linguagem, as vivências do homem em sociedade e suas reflexões sobre o mundo que o cerca. O escopo de seus estudos pode ser definido como:

The term 'humanities' includes, but is not limited to, the study and interpretation of the following: language, both modern and classical; linguistics; literature; history; jurisprudence; philosophy; archaeology; comparative religion; ethics; the history, criticism and theory of the arts; those aspects of the social sciences which have humanistic content and employ humanistic methods; and the study and application of the humanities to the human environment with particular attention to reflecting our diverse heritage, traditions, and history and to the relevance of the humanities to the current conditions of national life (NATIONAL ENDOWMENT FOR THE HUMANITIES, 2021).

Por sua vez, a ciência aberta é entendida como:

Uma nova abordagem para o processo científico baseado no trabalho cooperativo e novas formas de difusão do conhecimento usando tecnologias digitais e novas ferramentas colaborativas [...] deslocando-se das práticas padrão de publicação de resultados de pesquisa em publicações científicas em direção ao compartilhamento e uso do conhecimento disponível em estágio anterior do processo de pesquisa (EUROPEAN COMMISSION, 2016).

Segundo Nielsen (2012), os pilares da ciência aberta são: acessibilidade; capacidade de fazer descobertas; reutilização; reprodutibilidade; transparência; compreensão; colaboração; controle de qualidade; avaliação; validação; inovação; e o bem comum.

Subjacente a esses pilares, o Portal reconhece quatro grandes princípios da Ciência Aberta (OPEN SCIENCE & THE FIELD OF HUMANITIES CONFERENCE, 2018): acesso aberto (*open access*); dados abertos (*open data*); fonte aberta (*open source*); e ciência cidadã (*citizen science*). Cada um desses princípios promove a abertura da ciência de uma

maneira. O acesso aberto prevê o acesso imediato e gratuito ao usuário final à comunicação científica. A política de dados abertos preconiza o acesso gratuito e descomplicado a dados de pesquisa a todos, sem restrição de direitos autorais ou outros mecanismos restritivos. A fonte aberta permite que o código de programas de computador, *scripts* e aplicativos seja compartilhado por meio de licenças que prevejam a possibilidade de o código ser alterado por outros usuários. E por fim, a ciência cidadã estimula o engajamento do público na pesquisa científica, não somente os especialistas.

1. VETORES ORGANIZADORES DO PORTAL

Tendo em vista esses pilares e princípios, o Portal foi organizado em torno de quatro grandes vetores da ciência aberta nas Humanidades:

1. Acesso aberto (*Open Access*)
2. Dados abertos (*Open Data*)
3. Fonte aberta (*Open Source*)
4. Ciência cidadã / Humanidades Cidadãs (*Citizen Science / Citizen Humanities*)

1.1. ACESSO ABERTO

O primeiro vetor, acesso aberto, tem a função de impulsionar a criação e divulgação da produção científica da modalidade aberta, incluindo revistas e livros com acesso gratuito na rede. Essa modalidade tem tido grande crescimento com a popularização da *web*, desde os anos 1990. Já existem revistas da área de Humanidades de acesso aberto, como a própria *Cadernos de Linguística*, DELTA, Bakhtiniana e os periódicos patrocinados pela SciELO (Scientific Electronic Library Online), da FAPESP. Na SciELO, estão listados 93 periódicos na área de Ciência Humanas e 15 na área de Linguística, Letras e Artes, totalizando 108 periódicos.

Muitos outros periódicos de acesso aberto existem disponíveis em *sites* das Universidades responsáveis, os quais empregam majoritariamente a plataforma OJS (Open Journal System) de edição eletrônica. Embora não haja números definitivos a respeito, o acesso aberto a periódicos científicos nacionais das Humanidades parece ser a norma atualmente. No exterior, a adoção do acesso aberto tende a ser em menor escala, devido ao fato de a publicação de artigos e livros ser dominada por grandes empresas que têm nas assinaturas e na venda de publicações a sua fonte de renda. Mesmo assim, há diversas revistas das Humanidades com acesso aberto no exterior, como Digital

Humanities Quarterly (DHQ), destinada às Humanidades Digitais, Open Linguistics e Glossa, ambas de Linguística, bem como a Language Science Press, editora especializada em acesso aberto.

É preciso ressaltar que, embora o acesso ao leitor seja gratuito, muitos periódicos de acesso aberto exigem o pagamento de taxa de publicação (APC, Article Processing Charge), que os autores devem desembolsar para que seu artigo seja publicado. No país, algumas Universidades e agências de fomento dispõem de verba para o pagamento dessa taxa. Esse modelo baseado em APC vem sendo apresentado como uma possibilidade para garantir a sustentabilidade de periódicos nacionais por consórcios como a SciELO, tendo em vista a escassez de recursos públicos.

Mesmo com a adoção do acesso aberto pelas grandes editoras e consórcios que lideram a publicação científica no país e no exterior, a publicação em periódicos e em livros leva um tempo considerável, devido às etapas necessárias para trâmite da produção, incluindo a revisão por pares, alterações e editoração eletrônica. Aos olhos da ciência aberta, esse tempo entre a produção e a divulgação do conhecimento é um entrave para sua circulação. Embora, em certos periódicos, o prazo entre a submissão e a publicação possa ser reduzido em condições especiais (*fast track*), podendo chegar a apenas semanas, em vez de meses (ou anos) como é de costume em vários campos, diversas áreas fora das Humanidades adotam veículos mais ágeis e afinados com a ciência aberta, tais como *working papers*, *conference papers* e *preprints*, como destino primário do *output* de projetos de pesquisa.

Áreas como a computação, física, matemática e biologia têm dado preferência aos depositórios de *working papers* e *conference papers*, para permitir a publicação veloz do conhecimento científico. Depositórios como arXiv, medRiv e bioRxiv concentram quase dois milhões de publicações abertas. Esses depositórios não fazem o controle de qualidade da pesquisa, no entanto muitas publicações ao chegarem aos depositórios já passaram por revisão por pares em encontros científicos. Durante a pandemia do novo coronavírus, esses depositórios têm tido um papel fundamental na divulgação de conhecimento atualizado.

Embora a modalidade de ciência aberta seja de ampla aceitação em muitas áreas, nas Humanidades, incluindo a Linguística, ainda é pouco adotada¹. Um dos depositórios pioneiros na Linguística no país é o DIRECT Working Papers, que nos anos 1990 distribuiu os Working Papers do Projeto 'Em Direção à Linguagem dos Negócios' (CEPRIL, LAEL, PUCSP) pelo *site* do LAEL.

1 Em relação à Linguística, é importante frisar que o campo da Linguística Computacional destoa das demais sub-áreas da Linguística, na medida em que incorpora práticas muito maduras de acesso aberto e dados abertos já há bastante tempo.

As ações do Portal incluem a divulgação tanto de revistas de acesso aberto como de depositórios de publicações das Humanidades de modo a incentivar tanto a leitura quanto a produção de trabalhos na modalidade de acesso aberto. O Portal trará informação sobre a prática de cobrança de APC, para que os autores sejam alertados da necessidade de pagamento, visto que essa prática, embora usual em muitos campos do conhecimento, ainda é pouco conhecida nas Humanidades.

1.1. DADOS ABERTOS

O segundo vetor, dados abertos, tem a finalidade de disseminar a prática do livre acesso a dados de pesquisa. Em campos como as ciências naturais, computação e engenharias, a prática de compartilhar dados de pesquisa é bastante difundida. Contudo, em nenhum outro campo a disponibilização de dados de pesquisa é mais difundida do que na ciência de dados: existem grandes depositórios para compartilhamento de conjuntos de dados (data sets). Depositórios como Kaggle são frequentemente usados em competições entre equipes que disputam prêmios desenvolvendo soluções de programação de computador para efetuar a mineração dos dados (*text mining*) da maneira mais eficiente. Via de regra, esses depositórios guardam conjuntos volumosos de dados (big data), acessados por computador.

Além das plataformas de compartilhamento de dados, existem plataformas cujo propósito é auxiliar grupos de pesquisa a organizarem os dados na nuvem, compartilhando-os entre os integrantes da equipe e eventualmente tornando o acesso aberto. Um exemplo é a Open Science Framework (OSF), destinada a uso transparente de recursos de pesquisa. Embora não haja dados sobre a adoção da OSF entre pesquisadores das Humanidades, nossa impressão é que sua adoção ainda é incipiente. A exceção nesse caso são as Humanidades Digitais², que empregam naturalmente dados digitais e cujas equipes de pesquisa têm formação em computação, onde o processo de adoção da ciência aberta como um todo está em um estágio muito mais acelerado do que nas Humanidades (e mesmo em outros campos).

Nas Humanidades, os dados de pesquisa ainda tendem a ser vistos como propriedade de um indivíduo ou de um grupo, em vez de um recurso valioso que deve ser compartilhado para o avanço da ciência. As razões para essa diferença são muitas, incluindo o fato de a curadoria de dados nas Humanidades ser na maioria dos casos realizada manualmente (*hand and eye*). Além disso, muitos pesquisadores trabalham individualmente, em vez de em equipes, tornando a aquisição manual dos dados ainda mais laboriosa. Mesmo quando há

² Pesquisadores no campo da Linguística Computacional tendem a adotar a OSF e a publicar rotineiramente em equipes.

o desejo de compartilhamento, muitas vezes há impedimentos de natureza ética, por exemplo quando o pesquisador não obtém a tempo a permissão devida junto a suas fontes de dados. Outras vezes essa permissão não é possível de obter, dada a existência de informações de cunho pessoal ou sensível nos dados. Esses problemas podem ser evitados quando é feito planejamento para a abertura dos dados desde o início da coleta, permitindo que o pesquisador colete os dados em consonância com as exigências da ciência aberta, evitando imprevistos. Infelizmente, esse planejamento não é costumeiro nas Humanidades, podendo impedir a adesão à ciência aberta.

O Portal pretende atuar por meio desse vetor para a mudança de cultura da área, de tal modo que mais pesquisadores reutilizem dados existentes e enxerguem os benefícios do compartilhamento de dados. Para evitar o impedimento do compartilhamento devido à falta de permissão por conselhos de ética, serão veiculados no Portal tutoriais para planejamento da coleta. O Portal trará informações sobre dados disponíveis em plataformas de compartilhamento, além de dados oferecidos por pesquisadores diretamente ao Portal. Assim, o Portal não será apenas um centro de redistribuição de dados de outros centros, mas também uma fonte de dados inéditos. Serão incluídas, entre outras informações, o link para acesso aos dados, origem, histórico, formato, extensão, composição e público-alvo. Somente dados cuja distribuição atenda os requisitos éticos necessários serão disponibilizados.

1.3. FONTE ABERTA

O terceiro vetor, fonte aberta, refere-se à produção, disponibilização e adoção de software *open source*, isto é, cujo código pode ser distribuído e modificado. Existem muitos exemplos de software de fonte aberta, para as variadas soluções, como Mozilla Firefox, Linux, Python, R e LibreOffice. O software de código aberto geralmente tem um correspondente de código fechado -- nesses casos, o código fonte é propriedade intelectual, protegida por lei, e alterar e/ou distribuir o código é crime. Devido ao alto custo das licenças, a chamada pirataria de software tornou-se prática comum, principalmente em países de baixa renda. O modelo de software aberto evita a distribuição ilegal e os problemas decorrentes dessa prática. Atualmente, a qualidade do software de código aberto em muitos casos não fica aquém dos equivalentes comerciais, tornando esse tipo de programa uma alternativa realista, sendo adotados por uma quantidade crescente de usuários.

Além de ferramentas de uso geral, como navegadores, sistemas operacionais, pacotes estatísticos e de edição de texto, há um grande número de ferramentas de código aberto para as mais variadas finalidades de pesquisa, muitas das quais de grande relevância para a pesquisa nas Humanidades, como etiquetadores (*part-of-speech taggers*), analisadores sintáticos (*parsers*) e anotadores para as mais variadas finalidades.

O software de código aberto geralmente é produzido e mantido por desenvolvedores trabalhando em conjunto em diferentes lugares do mundo ao mesmo tempo. O uso de tecnologia computacional é uma premissa do desenvolvimento científico e na medida em que essa tecnologia seja gratuita, com atualizações contínuas e gratuitas, tanto a comunidade científica quanto a sociedade são beneficiadas.

Em relação a esse vetor, o Portal pretende contribuir para a difusão do software aberto nas Humanidades oferecendo links para o download, tutoriais de instalação e catálogos contendo descrição do propósito, público-alvo, nível de dificuldade de uso e instalação, etc. Os programas serão classificados em tipos, para permitir que o usuário tenha uma visão geral das opções disponíveis.

1.4. CIÊNCIA CIDADÃ, HUMANIDADES CIDADÃS

O quarto e último vetor, ciência cidadã, refere-se à participação do público em geral, de qualquer idade e formação, na pesquisa científica, em conjunto com a comunidade acadêmica. O público participante é geralmente composto de voluntários que dedicam parte de seu tempo colaborando com pesquisadores e acadêmicos em projetos de interesse para a comunidade. Tais iniciativas são um tipo de *crowdsourcing*, na medida em que grupos de indivíduos coletam e/ou fornecem dados pela Internet ou redes digitais, com ou sem ganho financeiro.

Tanto a comunidade científico-acadêmico quanto a comunidade em geral são beneficiadas com a ciência cidadã. Os benefícios para a comunidade científica vão desde a ampliação da coleta de dados até a formação de futuros pesquisadores, passando pela economia de recursos. E para a comunidade em geral, incluem a criação de vínculo entre o cidadão 'comum' e o pesquisador 'profissional', o envolvimento de cientistas em questões de impacto para a comunidade e a aplicação do resultado de pesquisas para o bem da sociedade.

Existem muitas iniciativas de ciência cidadã pelo mundo, como o Citizen Science (<https://www.citizenscience.gov>), projeto patrocinado pelo governo dos EUA que incentiva a participação da comunidade em atividades de pesquisa junto a órgãos federais e o EU-Citizen Science, conduzido pela União Europeia, que busca envolver a comunidade em projetos desenvolvidos nos países membros. Via de regra, os projetos de ciência cidadã referem-se a questões da alçada das ciências naturais -- os temas abarcados pelos projetos da EU-Citizen Science são:

Agricultura e Ciências Veterinárias; Animais; Arqueologia e Cultura; Astronomia e Espaço; Biodiversidade; Biogeografia; Biologia; Pássaros; Ciências Químicas; Clima e tempo; Ecologia e Meio Ambiente; Educação; Ciência gastronômica; Genética; Geografia; Geologia e Ciências da Terra; Saúde e Medicina; Cultura indígena; Ciências da Informação

e Computação; Insetos e polinizadores; Monitoramento de espécies de longo prazo; Oceano, Água, Mar e Terra; Natureza e ar livre; Gestão de recursos naturais; Física; Política de ciência; Ciências Sociais; Som; Transporte.

Como se percebe, apenas cinco dos 29 temas não estão ligados às ciências naturais (Arqueologia e Cultura, Educação, Cultura indígena, Política de ciência, Ciências Sociais).

No Brasil, a ciência cidadã está sendo introduzida por órgãos federais, como o SiBBr (Sistema Nacional de Informação sobre a Biodiversidade Brasileira), braço nacional do Global Biodiversity Information Facility (GBIF), além de Universidades e agências de financiamento de pesquisa. Contudo, também aqui o foco tem sido nas ciências naturais, particularmente na conservação do meio-ambiente.

Devido a esse viés das ciências naturais, surge um movimento chamado Humanidades Cidadãs (*Citizen Humanities*), para alavancar o engajamento da sociedade com a pesquisa nas Ciências Humanas. Segundo Heinisch et al. (2021, p. 98):

Citizen humanities is the term for citizen 'science' in the humanities. While (citizen) science comprises natural sciences, such as biology, chemistry, and physics, (citizen) humanities encompass fields such as languages, literature, history, philosophy, and art.

Entre projetos de Humanidades Cidadãs, existem iniciativas relacionadas a anotação colaborativa, transcrição, categorização de conteúdo, tradução, georreferenciação, coleta, curadoria e organização de conteúdo (HEINISCH et al., 2021, p. 100).

O Portal reconhece que ainda há um longo caminho a percorrer até que as Humanidades Cidadãs gozem do nível de difusão e do grau de impacto de seus equivalentes de outras áreas. As Humanidades Cidadãs continuam a ser um campo pouco conhecido no país e no mundo, porém com grande potencial de crescimento, haja vista a importância das Ciências Humanas para a sociedade. Assim, além de apresentar as iniciativas existentes, o Portal pretende criar conteúdo e eventos para conscientizar tanto pesquisadores quanto alunos e a comunidade em geral da existência das Humanidades Cidadãs e de sua relevância, mostrando oportunidades de engajamento da sociedade com a pesquisa em Humanidades em geral e com a Linguística em particular. Muitas vezes, projetos de pesquisa desenvolvidos por docentes e discentes em instituições do país qualificam-se como parte das Humanidades Cidadãs, mesmo que os praticantes não tenham se dado conta. O Portal pretende identificar tais iniciativas e catalogá-las como potenciais candidatas às Humanidades Cidadãs. Ao mesmo tempo, o Portal pretende ser um ponto de contato entre pesquisadores e a comunidade, visando ao estabelecimento de parcerias para projetos em comum dentro da área.

2. PRINCÍPIOS DE DESIGN DO PORTAL

Para a criação do Portal na *web*, serão adotados princípios do design de *website* (GARETT et al., 2016) e de acessibilidade seguindo o eMAG, Modelo de Acessibilidade em Governo Eletrônico (<http://emag.governoeletronico.gov.br>).

Para Garrett et al. (2016), a principal consideração deve ser usabilidade (*usability*): segundo a International Standardization Organization (2018), usabilidade é uma medida da eficiência e satisfação com que os usuários conseguem fazer as tarefas desejadas. De acordo com Garrett *et al.* (2016, p. 3), a usabilidade envolve vários critérios, tais como:

- (1) organização: o *site* deve ser organizado de forma lógica;
- (2) utilidade: o conteúdo prestado pelo *site* tem utilidade para o usuário;
- (3) navegação: o usuário consegue acessar o conteúdo do *site* com eficiência;
- (4) grafismo: os elementos gráficos, tais como cores, ícones, ilustrações, são agradáveis e valorizam o *site*;
- (5) propósito: o *site* tem um propósito claro e bem-definido;
- (6) familiaridade: o *site* permite que os usuários se acostumem com o seu formato e consigam navegar mais facilmente à medida que retornam;
- (7) atualização: o conteúdo e os links são atuais;
- (8) simplicidade: o *site* é desobstruído e de aparência simples;
- (9) objetividade: o *site* fornece informações diretas e objetivas;
- (10) confiabilidade: os dados fornecidos pelo *site* são confiáveis;
- (11) consistência: as diversas partes do *site* têm design compatível umas com as outras;
- (12) precisão: as informações são precisas,
- (13) velocidade: o *site* carrega no navegador do usuário com rapidez;
- (14) segurança: o *site* utiliza protocolos seguros de transmissão de dados pessoais, se houver;
- (15) interação: o *site* permite que o usuário interaja com o conteúdo e demais usuários.
- (16) customização: o *site* permite que o usuário customize sua experiência;
- (17) legibilidade: as fontes e os textos são fáceis de ler e entender;
- (18) eficiência: os usuários conseguem encontrar informação eficientemente;
- (19) seletividade: os usuários conseguem selecionar informação com facilidade;
- (20) aprendizagem: o *site* consegue auxiliar no aprendizado dos usuários.

Com base na revisão da literatura de múltiplos estudos que investigaram o design de *sites*, Garrett et al. (2016) apontam os critérios mais recorrentes dentre os acima, em ordem de importância:

1. Navegação
2. Grafismo
3. Organização
4. Utilidade
5. Propósito
6. Simplicidade
7. Legibilidade

Segundo Rosen e Purinton (2004), há três fatores principais que influenciam o sucesso do design de *websites*:

1. Coerência
2. Complexidade
3. Legibilidade

O fator de coerência abrange os seguintes elementos:

1. As informações são organizadas logicamente
2. É intuitivo
3. Tem redação clara e de qualidade
4. O conteúdo é variado
5. A navegação é simples
6. Faz o usuário querer explorar o conteúdo

Já o fator de complexidade engloba os seguintes quesitos:

1. Emprega elementos visuais satisfatoriamente

2. As imagens e demais elementos não-textuais são integrados
3. Utiliza diversos tipos de material gráfico e multimodal

Por fim, o fator de legibilidade inclui os elementos seguintes:

1. Assemelha-se a outros *sites*, evitando surpresas e demora na utilização
2. Tem uma identidade própria
3. Possui elementos únicos que os distinguem dos demais

Esses critérios serão adotados no desenho do *website* que abrigará o Portal.

Em relação à acessibilidade, empregaremos as diretrizes do Modelo de Acessibilidade em Governo Eletrônico (eMAG), adotado pelo governo federal do Brasil, que por sua vez é baseado no WCAG (Web Content Accessibility Guidelines). O eMAG apresenta uma série de princípios de design de websites tendo em mente a inclusão de usuários com necessidades especiais. A equipe do Portal implementará essas diretrizes a fim de permitir a acessibilidade do conteúdo a todos os usuários.

3. MÓDULOS INICIAIS DO PORTAL

Atrelados aos vetores organizacionais e guiados pelos princípios de design apresentados, o Portal comportará uma série de módulos de conteúdo. Esses módulos serão criados ao longo da duração do Projeto, podendo sofrer alteração, tendo em vista a natureza dinâmica da ciência aberta. No momento, estão previstos os seguintes módulos iniciais:

1. Módulo de corpora eletrônicos. Corpora são coletâneas de texto, fala ou multimodais, coletadas criteriosamente a fim de representar um domínio de uso da língua (BERBER SARDINHA 2004). O módulo de corpora eletrônicos do Portal será voltado para a catalogação e distribuição de corpora existentes, de vários tipos e línguas. O LAEL, com longa tradição na Linguística de Corpus, tornou-se ao longo dos anos um centro de desenvolvimento e distribuição de corpora. Entre os corpora desenvolvidos pelo Programa estão:

Sigla	Nome	Quantidade de palavras	Conteúdo	Financiador
CB	Corpus Brasileiro	1 bilhão	Português escrito e falado	FAPESP, CNPq
Br-ICLE	Brazilian sub-corpus of the International Corpus of Learner English	250 mil	Inglês escrito de aprendizes brasileiros da língua	CNPq, PIPEQ PUCSP
CBVR	Corpus Brasileiro de Variação de Registro	5 milhões	Português escrito e falado	CNPq
CRVR-E	Corpus for Register Variation Research in English	10 milhões	Inglês escrito e falado	CNPq, PIPEQ
SMEC	Social Media English Corpus	1,4 milhão	Inglês das redes sociais	CNPq, PIPEQ
CoronaCorpus	Coronavirus Sample Corpus	825 milhões	Inglês escrito	CNPq, PIPEQ
PUCSPAMC	São Paulo Catholic University Annotated Metaphor Corpus	1 milhão	Inglês escrito e falado	CNPq
COBRA-7	Multi-Level Corpus of Brazilian English Learners	527 mil palavras	Inglês escrito de aprendizes brasileiros da língua	CAPES

Tabela 1. Alguns corpora disponibilizados pelo LAEL.

Esses corpora, que já são distribuídos a pesquisadores e grupos qualificados por meio de licenças de uso pelo CEPRIL, serão os primeiros a compor o acervo de materiais do módulo. Além desses, muitos outros corpora disponíveis serão agregados à coleção. Por fim, corpora coletados para as pesquisas docentes e discentes do LAEL e de outros Programas serão adicionados ao módulo.

2. Módulo de produções de tradutores intérpretes de LIBRAS-Língua-Portuguesa/Língua Portuguesa/LIBRAS, registrados por meio de materiais audiovisuais, para estudo das especificidades em diferentes esferas da atuação humana, aí incluída a artística, hoje uma realidade na comunidade surda. Dada a realidade intersemiótica dos processos de tradução e de interpretação envolvendo línguas de sinais, esse segmento do portal teria papel importante para compreender de que forma essas atividades articulariam a dimensão verbal (gestual linguística) a outras dimensões languageiras humanas para promover, de forma inclusiva, a comunicação de pessoas surdas. E permitiria ao LAEL enfrentar o desafio da demanda, cada vez maior, de intérpretes em busca do LAEL por uma formação científica. Por outro lado, esse segmento do portal também estará aberto a outras línguas de sinais, assim como a materiais que contemplem, por exemplo, a interpretação para surdos-cegos.
3. Módulo de práticas intersemióticas e de discursos que circulam na contemporaneidade, como os de inclusão, meio ambiente e qualidade de vida, o que implica o estudo das normas e da atividade de trabalho de diferentes atores sociais. Constituir esse segmento do portal permitirá ao LAEL dar maior visibilidade àquilo que se realiza nas atividades das pessoas envolvidas e também incorporar a cultura produzida pelo coletivo, de tal maneira que essa produção passe a ter reconhecimento como parte do

processo de desenvolvimento como liberdade, que consiste em reforçar/aumentar as capacidades dos indivíduos. Pensar o binômio liberdades/capacidades fundamenta-se na compreensão de que o aumento das capacidades não se dissocia dos discursos que circulam em uma sociedade em um dado momento sociohistórico, os quais constroem e, muitas vezes, cristalizam questões políticas e sociais.

4. Módulo voltado para a investigação das expressões faciais e vocais de tal modo a: classificar perfis vocais e movimentos da face no discurso oral e relacioná-los a efeitos de sentido; relacionar características das expressões vocal e facial para identificar e distinguir características paralinguísticas e extralinguísticas dos participantes; pesquisar correlatos acústicos e perceptivos. O potencial da análise das expressões faciais para revelar estados cognitivos (atenção e concentração, por exemplo), afetivos/emotivos/atitudinais, primitivos emocionais (valência e ativação do organismo, por exemplo) e expressão de sentidos, embora imenso é ainda pouco conhecido e demanda pesquisa. Os sistemas automatizados de codificação facial identificam em tempo real os movimentos da face e viabilizaram a agilização da codificação dos movimentos da face por meio das Unidades de Ação (AUs), abrindo perspectivas para a investigação em tempo real do papel da expressão facial na expressão de emoções, atitudes e modalidades em situações de comunicação falada. O ambiente de reconhecimento de padrões, baseado em Inteligência Artificial (IA), respalda a codificação automática dos movimentos faciais. Tecnologias de redes neurais, de Aprendizagem por Máquina (Machine Learning) e de Aprendizagem Profunda (Deep Learning) estão envolvidas. Para fornecerem métricas confiáveis e precisas, os algoritmos são treinados utilizando grandes bancos de dados e se pautam pela consideração de variáveis determinadas pela etnia, idade e sexo. Os dados coletados para armazenamento nos bancos de dados representam expressões faciais espontâneas do mundo real. Utilização do sistema para avaliação auditiva de ajustes de qualidade e dinâmica vocal denominado Vocal Profile Analysis Scheme (VPA) de base fonética. Para a análise de parâmetros acústicos, é utilizado o ProsodicEvaluator, script desenvolvido por Plínio Barbosa do IEL UNICAMP para ser rodado no PRAAT. Para a anotação dos dados e a sincronização entre os gestos faciais e vocais, o ELAN, software livre do Max Planck Institute for Psycholinguistics é utilizado.

4. COMENTÁRIOS FINAIS

O projeto do Portal Multimodal/Multilíngue para o Avanço da Ciência Aberta nas Humanidades visa a ajudar na mudança de paradigma do fazer científico nas Ciências

Humanas, incluindo a Linguística, contribuindo com esforços de entidades (como a ABRALIN) e instituições governamentais e privadas no Brasil e no exterior. A abertura da ciência é um movimento sem volta, que exige novas posturas e novas práticas, trazendo benefícios para todos os envolvidos. O Portal vem se juntar a esse movimento, enfocando especificamente as Ciências Humanas, uma esfera do conhecimento ainda pouco aderente à essa mudança de paradigma do fazer científico.

O Portal é mais amplo que a recolha dos dados da Universidade que o abriga, não sendo um repositório institucional. Responde, com suporte no desenvolvimento de pesquisa dos estudos das linguagens, à demanda da academia em adotar práticas que visam o compartilhamento de informações sobre desenvolvimentos teóricos, métodos de análise e ferramentas, o compartilhamento de dados, a abertura à cooperação nacional e internacional entre pesquisadores, a imediata disponibilização dos resultados de pesquisa e ao impacto das tecnologias na sociedade.

5. AGRADECIMENTOS

Agradecemos o CNPq pelo apoio financeiro ao Portal, na forma de bolsas de estudo de mestrado e de doutorado e os dois pareceristas pela leitura e sugestões de melhoria do texto. Os autores seguintes agradecem o CNPq pelo apoio à pesquisa por meio de bolsas Produtividade em Pesquisa: Tony Berber Sardinha, Sandra Madureira, Beth Brait, Maria Cecília Perez de Souza-e-Silva e Zuleica Camargo. Os autores seguintes agradecem o CNPq pelo apoio à pesquisa por meio de bolsas de estudo: Renata Lamberti Spagnuolo e Arianne Alfonso Brogini Braz.

REFERÊNCIAS

BERBER SARDINHA, Tony. *Linguística de Corpus*. 1. ed. Barueri: Manole, 2004.

GARETT, Renee; CHIU, Jason; ZHANG, Ly; & YOUNG, Sean D. A Literature Review: *Website Design and User Engagement*. *Online Journal of Communication and Media Technologies*, Nicosia, Chipre, v. 6, n. 3, p. 1-14, jul. 2016. DOI: <https://doi.org/10.29333/ojcm/2556>. Acesso em: 18 jun. 2021.

HEINISCH, Barbara; OSWALD, Kristin; WEISSPFLUG, Maike; SHUTTLEWORTH, Sally *et al.* Citizen Humanities. *In*: VOHLAND, Katrin; LAND-ZANDSTRA, Anne; CECCARONI, Luigi; LEMMENS, Rob; PERELLÓ, Josep; PONTI, Marisa; SAMSON, Roeland e WAGENKNECHT, Katherin. *The Science of Citizen Science*. Cham, Switzerland: Springer, 2021. p. 97-118. DOI: <https://doi.org/10.1007/978-3-030-58278-4>

NATIONAL ENDOWMENT FOR THE HUMANITIES. *About the National Endowment for the Humanities*, 2021. Fundação de apoio a pesquisa. Disponível em: <https://www.neh.gov/about>. Acesso em: 18 jun. 2021.

NIELSEN, M. *Reinventing Discovery: The New Era of Networked Science*. 1. ed. Princeton, NJ: Princeton University Press, 2012.

OPEN SCIENCE & THE FIELD OF HUMANITIES CONFERENCE. *Open Science in the Field of Humanities*, 2018. Congresso de Ciência Aberta nas Humanidades. Disponível em: <http://www.ub.edu/openscienceandthehumanities/open-science/>. Acesso em: 18 jun. 2021.

EUROPEAN COMMISSION. *Open Innovation, Open Science, Open to the World: A vision for Europe*. 2016. Política de ciência aberta da União Europeia. Disponível em: <https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe>. Acesso em: 18 jan. 2020.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. *ISO 9241-11:2018*. 2018. Ergonomics of human-system interaction -- Part 11: Usability: Definitions and concepts. Disponível em <https://www.iso.org/standard/63500.html>. Acesso em: 18 jun. 2021.

ROSEN, Debora E.; PURINTON, Elizabeth. Website design: Viewing the web as a cognitive landscape. *Journal of Business Research*, Amsterdã, Holanda, v. 57, n. 7, p. 787-794, jul. 2004. [https://doi.org/10.1016/S0148-2963\(02\)00353-3](https://doi.org/10.1016/S0148-2963(02)00353-3). Acesso em: 18 jun. 2021.