

## ALiBWEB: ESTADO DA ARTE E PERSPECTIVAS FUTURAS

### ALiBWEB: STATE OF ART AND FUTURE PERSPECTIVES

Daniela Barreiro Claro | [Lattes](#) | [dclaro@ufba.br](mailto:dclaro@ufba.br)

Universidade Federal da Bahia

Josane Moreira de Oliveira | [Lattes](#) | [josanemoreira@hotmail.com](mailto:josanemoreira@hotmail.com)

Universidade Estadual de Feira de Santana | Universidade Federal da Bahia

Marcela Moura Torres Paim | [Lattes](#) | [marcelamtpaim@yahoo.com.br](mailto:marcelamtpaim@yahoo.com.br)

Universidade Federal Rural de Pernambuco | Universidade Federal da Bahia

**Resumo:** O presente artigo dá a conhecer o sistema denominado ALiBWeb, que se encontra em construção, à comunidade acadêmica. O processo de informatização do Projeto Atlas Linguístico do Brasil (ALiB) começou em 2007, quando aconteceram as reuniões iniciais sobre análise de requisitos para o desenvolvimento de um banco de dados e, conseqüentemente, de um sistema que possibilitasse gerenciar os dados armazenados. Assim, será mostrada a importância do ALiBWeb, que permitirá a organização dos dados e sua inserção no banco, facilitando o seu armazenamento e, principalmente, garantindo maior segurança com a informatização e disponibilizando-os a fim de socializá-los, para que possam servir de base para análises linguísticas de dados orais de natureza geolinguística.

**Palavras-chave:** Projeto Atlas linguístico do Brasil; Informatização; Sistema ALiBWeb.

**Abstract:** This article presents to the academic community the web system, called ALiBWeb, which is under construction. The automation of the Linguistic Atlas of Brazil (ALiB) project started in 2007, with meetings to analyze the requirements for developing the database and, consequently, the web system, which manages such data. ALiBWeb enables the management of data easing the storage and guaranteeing greater data security to make it available. Consequently, the data will be socialized, so that they can serve as a basis for linguistic analyzes with oral data of geolinguistic nature.

**Keywords:** Linguistic Atlas of Brazil Project; Automation; ALiBWeb System.

## 1 Introdução

Apresenta-se, neste estudo, o banco de dados e o sistema ALiBWeb (em construção) do Projeto Atlas Linguístico do Brasil (ALiB), desenvolvido em âmbito nacional, que já publicou os dois primeiros volumes (CARDOSO et al., 2014a; CARDOSO et al., 2014b), estando em andamento os volumes 3, 4, 5 e 6.

Trata-se de um projeto nos moldes da Geolinguística Pluridimensional contemporânea (THUN, 2000), que integra a Dialectologia e a Sociolinguística, uma vez que, ao lado da variável diatópica, inclui parâmetros sociolinguísticos (LABOV, 2008 [1972]), considerando as variáveis sociais sexo, faixa etária e escolaridade.

Este artigo busca descrever a importância da informatização do banco de dados do referido projeto para disponibilizar os dados que foram coletados em campo, a fim de socializar o conhecimento adquirido, permitindo que pesquisas sejam aprimoradas na área da Dialectologia. Nesse sentido, um sistema web (ALiBWeb) está sendo desenvolvido para projetar os dados armazenados no banco de dados com o intuito de os pesquisadores de áreas afins e próximas (história, antropologia, sociologia, ciência da computação, entre outras) terem novos subsídios para o conhecimento da variação linguística nas diversas áreas brasileiras. Com essa informatização, os dados atualizados poderão ser disponibilizados através do ALiBWeb e, assim, servir para uma melhor interpretação do caráter multidialectal do português do Brasil, podendo servir para estudo de variação na escola, com base em dados reais oriundos de pesquisa de campo, coletados *in loco*.

Além desta seção introdutória, este artigo é composto por quatro seções que abordam: o Projeto ALiB (seção 2), o encontro da Dialectologia e da Ciência da Computação (seção 3); o banco de dados e o sistema ALiBWeb (seção 4); as perspectivas futuras do banco de dados (seção 5); e, por fim, as considerações finais (seção 6), seguidas das referências.

## 2 O Projeto ALiB

Antes de relatar como começou o processo de informatização do Projeto ALiB, é importante expor os objetivos e os aspectos metodológicos do referido projeto, cujas bases foram lançadas no Seminário *Caminhos e Perspectivas para a Geolinguística no Brasil*, realizado em 1996, em Salvador-BA (Brasil).

O Projeto ALiB tem como meta a produção de um atlas geral do Brasil, no que se refere à língua portuguesa, e tem por objetivos, entre outros: (i) documentar e descrever a língua portuguesa falada no Brasil; (ii) estabelecer isoglossas para a delimitação de áreas

as dialetais brasileiras; e (iii) oferecer dados para análises linguísticas e para subsidiar o ensino de língua portuguesa.

Para concretizar o desejo de produzir o atlas, os pesquisadores constituíram um Comitê Nacional, responsável por gerenciar as atividades do empreendimento, tendo como uma de suas metas manter a unidade teórico-metodológica do projeto e garantir a execução das atividades de modo conjunto e coordenado. Esse Comitê está constituído atualmente por 13 membros pesquisadores vinculados a dez universidades públicas brasileiras<sup>1</sup>.

O projeto, na sua essência de âmbito linguístico, pois se propõe a documentar, descrever e interpretar a diversidade do português brasileiro, tem, exatamente por esse caráter, uma evidente interface com ramos distintos do conhecimento organizado, decorrente do fato de que a história de uma língua é a história do próprio povo que a fala. Esse caráter de que se reveste o Projeto ALiB tem duas evidentes implicações: por um lado, inspira e fundamenta a sua concepção na pluralidade do conhecimento; por outro, permite que, dos resultados que venha a oferecer, se beneficie amplo espectro das ciências na atualidade.

Quanto ao primeiro dos aspectos, a concepção do projeto conduziu a que se buscase fundamentação em diferentes campos do conhecimento. A definição da rede de pontos para levantamento de dados, por exemplo, requereu aprofundado conhecimento no campo da história, da antropologia, da demografia, da geografia (física, humana e política) e, sobretudo, dos estudos culturais, da economia de cada região, do desenvolvimento social e político, do aspecto religioso. A seleção de localidades reflete não apenas o interesse linguístico, mas também o perfil sócio-histórico das zonas mapeadas, que, se, por um lado, é um indicador de importância para a visão de língua, por outro, encerra um relevante feixe de correlações sócio-histórico-culturais.

Nesse sentido, a rede de pontos do Projeto ALiB é composta de 250 localidades, aí incluídas as capitais brasileiras (com exceção de Brasília e Palmas, pelo seu pouco tempo de formação). Para a seleção das localidades foram considerados os seguintes critérios: (i) distribuição geográfica; (ii) história da cidade; (iii) demografia; (iv) características socioculturais; e (v) perfil linguístico.

Para a seleção de informantes, fez-se necessário um estudo da formação demográfica do Brasil, da constituição da sociedade, dos aspectos sociológicos e antropológicos que marcam a composição da população brasileira, a que se acrescenta um conhecimen-

---

<sup>1</sup> Cf. informações sobre a composição do Comitê Nacional de coordenação do Projeto ALiB no *site* do projeto: <https://alib.ufba.br/content/comit%C3%AA-nacional>.

to da realidade social na qual se inserem os entrevistados. Isso levou a que se buscasse harmonizar diferentes variáveis sociais – como idade, sexo, escolaridade – para que se pudesse obter uma amostra linguística representativa da diversidade linguística do país. Dessa forma, foram entrevistados 1.100 informantes (oito em cada capital e quatro em cada cidade do interior), de ambos os sexos, distribuídos por duas faixas etárias (Faixa I – de 18 a 30 anos e Faixa II – de 50 a 65 anos), totalizando cerca de 3.300 horas de gravação. Nas capitais, os informantes são estratificados em dois níveis de escolaridade (fundamental e universitário); nas demais localidades, todos os informantes têm apenas o nível fundamental.

O Projeto ALiB nasceu, assim, evidenciando uma profunda correlação com outros ramos do conhecimento científico. A sua implantação, em 1996, propiciou a retomada da discussão da metodologia relativa aos trabalhos de natureza geolinguística e à elaboração de instrumentos de trabalhos adequados às necessidades de coleta de dados empíricos, fatos tão importantes ao desenvolvimento da Geolinguística no Brasil, a ponto de já se interpretar o Projeto ALiB como marco de uma nova fase na Dialetologia Brasileira, a quarta, definida por Mota e Cardoso (2006) – se se admitem as três fases anteriormente propostas por Ferreira e Cardoso (1994), em ampliação às duas anteriormente propostas por Nascentes (1952; 1953) –, ou a quinta, para Teles (2019).

Considera-se, portanto, uma forma de expansão desse conhecimento específico a transferência de tecnologia dele decorrente na área da pesquisa geolinguística, a partir do aproveitamento das discussões metodológicas ocorridas em reuniões e *workshops* nacionais e regionais, em palestras e em minicursos ministrados pelos pesquisadores que integram a equipe do Projeto ALiB.

Entre os pontos que podem ser arrolados como de transferência de tecnologia, destacam-se:

- (i) Questionário Linguístico – os questionários do ALiB (COMITÊ NACIONAL..., 2001) envolvem vários níveis linguísticos: fonética/fonologia, incluindo a prosódia; morfossintaxe; léxico; e pragmática. Além disso, há ainda discursos semidirigidos, questões metalinguísticas e a leitura de um texto. Pela sua amplitude com relação aos diversos níveis de estudo da língua de que trata, os questionários têm servido de base para a elaboração de questionários específicos, em pesquisas diversas, para trabalhos de pós-graduação e para a realização de atlas linguísticos regionais, fornecendo elementos que

funcionam como facilitadores de novas investigações sobretudo daquelas de iniciativa de jovens investigadores;

- (ii) Implementação de atlas regionais<sup>2</sup>, com o conseqüente crescimento de grupos de pesquisa na área da Dialetoologia e o aumento do número de trabalhos de pós-graduação sobre aspectos da Dialetoologia brasileira;
- (iii) Formação de jovens pesquisadores, com uma procura cada vez maior de estudantes de graduação e de pós-graduação pelo trabalho com dados do ALiB e pela orientação acadêmica de pesquisadores do projeto. O Comitê Nacional do Projeto ALiB tem procurado, continuamente, agregar jovens pesquisadores e estudantes de graduação e de pós-graduação à investigação científica<sup>3</sup>. A titulação obtida por estudantes, os trabalhos por eles desenvolvidos, assim como a presença no quadro de professores das instituições de ensino superior são dados comprobatórios da ação dos membros do Comitê na formação de jovens pesquisadores.

Cumpra detalhar um pouco os questionários do ALiB. O questionário fonético-fonológico tem 159 perguntas (aí incluídas as questões de prosódia – para investigar as interrogativas, por exemplo) e abarca fenômenos como a realização das vogais pretônicas e postônicas, o /s/, o /r/ e o /l/ em coda silábica, a realização de ditongos, a realização de /t, d, n, l/ diante de [i], dentre outros fenômenos variáveis. O questionário semântico-lexical abrange 14 áreas semânticas (acidentes geográficos, fenômenos atmosféricos, astros e tempo, atividades agropastoris, fauna, corpo humano, ciclos da vida, convívio e comportamento social, religião e crenças, jogos e diversões infantis, habitação, alimentação e cozinha, vestuário e acessórios e vida urbana) e tem 202 perguntas. O questionário morfossintático conta com 49 perguntas, envolvendo fenômenos como a marcação do gênero e do número, o uso de artigo antes de antropônimos, o uso dos pronomes, os tempos verbais, a negação, dentre outros. Há ainda quatro questões de pragmática (como o falante se dirige a um interlocutor mais jovem ou mais velho e de ambos os sexos), quatro temas para discurso semidirigidos (relato de uma experiência pessoal, comentário e discurso reportado, por exemplo) e seis questões metalinguísticas (qual a língua que se fala, se há pessoas que falam diferente, por exemplo), além do texto “A parábola dos sete vimes” para leitura.

<sup>2</sup> No que se refere aos atlas regionais desenvolvidos a partir da implantação do Projeto ALiB e com a utilização de metodologia desenvolvida pelo Projeto, sugere-se a consulta a Aguilera e Romano (2016).

<sup>3</sup> Sobre a produção acadêmica realizada com base no *corpus* do Projeto ALiB, sugere-se a consulta ao *site* oficial do Projeto, item “publicações”, disponível em: <https://alib.ufba.br/>.

O resultado imediatamente esperado do Projeto ALiB é, evidentemente, a produção do próprio atlas, cujos volumes iniciais, Introdução e Cartas Linguísticas I, foram publicados em 2014. Esse conjunto de dados espelha, na sua amplitude maior, as variedades do português falado no Brasil. A cartografia dos dados, com base na georreferencialidade, objetiva, além de mapear os dados linguísticos: (i) mostrar coordenadas seguidas no povoamento do país, desfazendo dúvidas sobre roteiros de penetração ou oferecendo elementos comprobatórios de levas de povoamento fixadas nesses locais ou que por eles transitaram; (ii) assinalar o papel de acidentes geográficos na difusão de hábitos linguísticos – como se pode ver examinando, por exemplo, o papel dos rios – ou no isolamento de fenômenos que se detêm por trás de montanhas ou incrustados em vales.

No que se refere à educação, mais especificamente ao ensino de língua portuguesa, é altamente significativa a contribuição do Projeto ALiB, cujos resultados propiciarão um melhor equacionamento do ensino-aprendizagem à diversidade de cada região, uma vez que, descritas as peculiaridades de cada área e caracterizada a variedade de uso da língua ali dominante, pode-se construir um modelo de ensino do vernáculo mais eficaz e baseado em usos reais.

A esses aspectos relacionados, com os quais não se pretendeu esgotar a indicação das possibilidades de interdisciplinaridade do projeto, mister se faz destacar o que advém de um atlas para os estudos linguísticos.

Embora amplo em sua proposta, o Projeto ALiB tem limitações, como, por exemplo, a quantidade de pontos da rede, a quantidade de informantes por localidade, as faixas etárias extremas e a variável escolaridade apenas nas capitais. Isso se justifica pela decisão de optar por um projeto exequível num país de dimensões continentais e com uma população numerosa, com uma equipe relativamente pequena – mas muito bem preparada – e com os poucos recursos recebidos para deslocamento, hospedagem e aquisição de material. Ainda assim, ressalte-se que foram percorridos 277.851 km, do Oiapoque ao Chuí!

Incluir a faixa etária intermediária (entre 31 e 49 anos), por exemplo, implicaria ampliar em um terço todo o trabalho e tempo de pesquisa de campo. Aumentar o número de informantes por célula implicaria mais trabalho e tempo ainda, o que inviabilizaria a execução do projeto. Encontrar informantes com nível universitário de escolaridade em algumas cidades do interior do Brasil, sobretudo em alguns estados do Norte e do Nordeste, por exemplo, ainda é difícil num país com tantas desigualdades sociais, com sérios problemas na educação pública e com pouco acesso da população à universidade.

Esses recortes necessários, que limitam a amostra do ALiB, objetivam mostrar uma fotografia da diversidade da língua portuguesa falada no Brasil. Como toda foto capta um

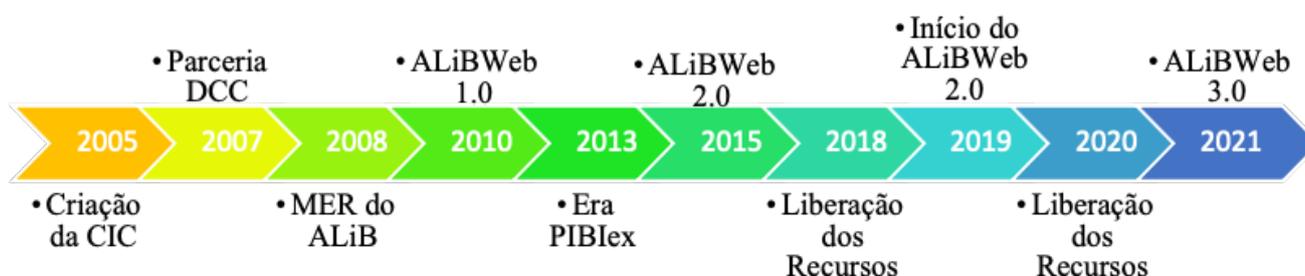
instante de um *flash* (não é um vídeo), cabe aos interessados, a partir do que é possível ver numa foto, a curiosidade de voltar a campo e verticalizar a pesquisa de determinado(s) fenômeno(s) em determinada(s) localidade(s) registrado(s) pelo ALiB. O que se quer mesmo é instigar a continuidade e o aprofundamento da pesquisa empírica, o que, aliás, já tem sido feito por pesquisadores que se inspiraram nos primeiros resultados do ALiB.

Por outro lado, tem sido alto o investimento na informatização do projeto, incluindo um sistema denominado ALiBWeb e um banco de dados com o intuito de salvar, gerenciar e disponibilizar as informações coletadas. Para tanto, foi necessário estabelecer um diálogo com a Ciência da Computação, do qual se fala na seção seguinte.

### 3 Quando a Dialetoлогия e a Ciência da Computação se encontram

A informatização do Projeto ALiB ocorreu em várias etapas, sendo iniciada em 2005 quando da criação da Comissão de Informatização e Cartografia (CIC). Essa comissão tinha como principal objetivo informatizar o ALiB, facilitando as consultas aos dados catalogados pelos inquiridores do projeto e auxiliando a geração das cartas linguísticas. A Figura 1 descreve a linha de tempo desta informatização, principalmente no que concernem os aspectos relacionados à Ciência da Computação.

**Figura 1** – Linha de Tempo da Informatização do ALiB



Fonte: Elaborada pelas autoras.

Em 2007, iniciou-se a parceria com o Departamento de Ciência da Computação (DCC) da Universidade Federal da Bahia, tendo um membro integrante na CIC. Durante todo o ano de 2007 e o ano de 2008, esse membro participou das diversas reuniões do Projeto ALiB com o intuito de compreender a sua essência, o seu objetivo e subsidiar a criação do banco de dados do ALiB através do Modelo de Entidades e Relacionamentos (CHEN, 1990).

Em 2010, o Projeto ALiB teve o primeiro projeto de pesquisa aprovado no Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) para a realização das atividades referentes à CIC. O objetivo desse projeto foi o desenvolvimento do ALiBWeb, na sua primeira versão, com o intuito de facilitar as consultas aos dados catalogados. Inicialmente, as interfaces prioritárias foram as administrativas, porém houve a necessidade de criar um *importador* para que os inquéritos realizados no Word fossem importados para o banco de dados do ALiB de uma maneira semi-automatizada. Durante os dois anos subsequentes, o foco do projeto foi no desenvolvimento dos novos módulos correspondentes à política de acesso e às cartas linguísticas. Porém, com o final do projeto e a descontinuidade do *framework* utilizado no desenvolvimento do projeto, a atualização e manutenção do mesmo para garantir um desempenho satisfatório foi comprometida e, assim, novas iniciativas foram necessárias.

Em 2013, devido à situação financeira em que o país estava condicionado, a solução para a manutenção do sistema ALiBWeb foi através dos diversos projetos de Programa de Bolsas de Iniciação à Extensão (PIBEx) oferecidos pela UFBA.

Um novo panorama surgiu no final de 2014 e em 2015, quando o Projeto ALiB teve outro projeto de pesquisa aprovado para o desenvolvimento do ALiBWeb, desta vez através da Fundação de Amparo e Pesquisa da Bahia (FAPESB). Apesar da aprovação do projeto, a liberação dos recursos só ocorreu por volta de 2018, quando houve um movimento para a contratação de uma empresa especializada em desenvolvimento de *software*. Em tempo, uma análise aprofundada do ALiBWeb, na sua versão 1.0, foi desenvolvida com o objetivo de averiguar o que poderia ser reutilizado e o que deveria ser desenvolvido com tecnologia mais atualizada. Após essa análise, foi recomendado que o sistema ALiBWeb fosse desenvolvido em uma linguagem de programação mais leve e com fácil interatividade entre os usuários. Assim surgia a segunda versão do ALiBWeb, que iniciou o seu desenvolvimento em 2019.

O avanço do desenvolvimento do ALiBWeb na segunda versão foi impactado novamente pela crise financeira e a liberação da segunda parcela do projeto em desenvolvimento só ocorreu após mais de um ano da requisição. Isso impactou no ritmo do projeto e no desenvolvimento de algumas das suas funcionalidades, que foram retomadas no segundo semestre de 2020, durante a pandemia do coronavírus que assolava o país.

Ainda em 2019, o ALiB teve outro projeto de pesquisa aprovado no CNPq com recursos bem reduzidos, mas que permitirá o desenvolvimento do Mapa Falante, correspondendo à terceira versão do ALiBWeb.

Na seção seguinte, descreve-se o estado atual do sistema ALiBWeb e detalha-se a estrutura do banco de dados que comporta os dados do ALiB.

#### **4 O estado atual do banco de dados e do sistema ALiBWeb**

Tecnologicamente, o ALiBWeb é um sistema inovador em âmbito internacional, visto que se utiliza de tecnologias relevantes da área de Ciência da Computação com o intuito de melhor desenvolver e publicar os dados inseridos no banco de dados do Projeto ALiB. O próprio ALiBWeb já é, pela sua própria natureza, uma contribuição tecnológica para ambas as áreas: a Dialectologia e a Ciência da Computação.

O ALiBWeb é especificado em módulos que permitem gerenciar as transcrições dos inquéritos; assim como os informantes e as suas respostas. Além dessas funcionalidades, possui módulos de auditoria, autorização e autenticação de usuários. Esse sistema foi desenvolvido em Ruby on Rails (SMART, 2016). Na fase atual, está sendo feita a validação e os ajustes nos módulos, envolvendo a interação dos usuários finais. Quanto à análise linguística e às cartas geradas pelo ALiBWeb, estas correspondem às cartas que já foram publicadas, o que indica que os testes apresentam resultado positivo.

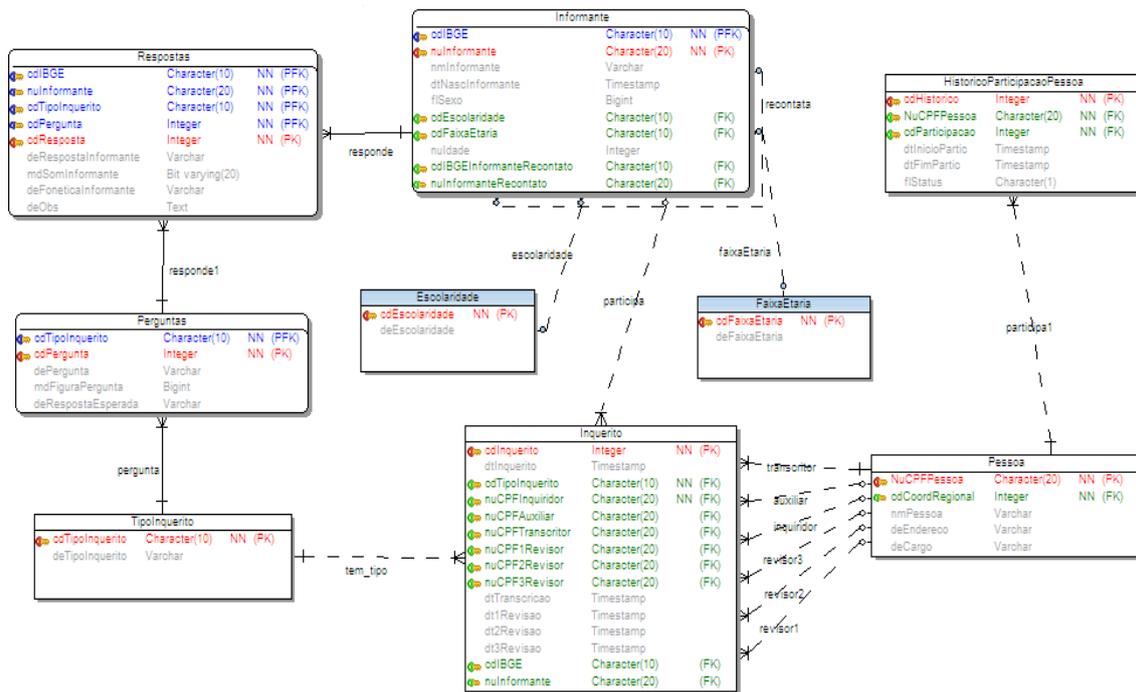
##### **4.1 O banco de dados**

O banco de dados do ALiB foi conceituado e modelado através do Modelo de Entidades e Relacionamentos (MER) (CHEN, 1990) e contém mais de 60 entidades e relacionamentos referentes ao Projeto ALiB. O mapeamento relacional foi realizado para o Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL (GONZAGA, 2007). Os dados do ALiB foram importados através de um módulo do ALiBWeb denominado Importador de Inquéritos, no qual os inquéritos transcritos em documentos no formato Word foram incluídos no sistema ALiBWeb.

A reorganização de alguns módulos culminou com uma evolução no MER do ALiB. Assim, uma nova versão do MER foi gerada através da Engenharia Reversa, ou seja, a partir do banco de dados PostgreSQL. Devido a sua grandiosidade, a última versão do MER do ALiB pode ser visualizada no endereço: <<https://drive.google.com/ALiB-MER>>.

Algumas entidades e relacionamentos do MER do ALiB foram destacados na Figura 2.

**Figura 2** – Principais entidades do MER do ALiB



Fonte: Elaborada pelas autoras.

Para os interessados na área de Modelagem de Dados, dentre as principais entidades, destacam-se: *Inquérito*, *Informante*, *Pessoa*, *Respostas*, *Perguntas* e *Tipos de Questionário*. A construção de um banco de dados relacional em um Sistema Gerenciador de Banco de Dados Relacional (SGBD-R) envolve três etapas principais. A primeira delas, que é o Projeto Conceitual do banco de dados, é a criação do Modelo de Entidades e Relacionamentos (MER) com o intuito de conceber o banco de dados requerido e ter as diversas interações com o usuário. A segunda etapa, denominada de Projeto Lógico, corresponde ao mapeamento do MER para um SGBD-R específico; no caso do ALiB, o mesmo foi mapeado para o PostgreSQL. Nesse nível, as entidades do MER passam a ser mapeadas para relações e as instâncias de entidades passam a ser as tuplas<sup>4</sup>. Por fim, a terceira etapa é a criação das tabelas no banco de dados relacional com as chaves primárias (*primary keys*) e as chaves estrangeiras (*foreign keys*), garantindo a integridade referencial (ELMASRI; NAVATHE, 2019).

Realizado o mapeamento para o SGBD-R, observam-se, em vermelho, as chaves primárias de cada relação, que permitem identificar unicamente as tuplas. Em verde, as chaves estrangeiras permitem garantir a restrição de integridade referencial nas tuplas de

cada relação. E, por fim, em azul, destacam-se as chaves primárias, que também são chaves estrangeiras na relação. A cardinalidade entre cada relação no MER está mapeada em cada relacionamento binário descrito na Figura 2. As cardinalidades apresentadas são 1:1, 1:N e M:N (ELMASRI; NAVATHE, 2019). Essas cardinalidades definem qual o tipo de relacionamento e o tipo de participação que cada relação tem com o relacionamento.

Uma vez mapeado o banco de dados para o PostgreSQL (GONZAGA, 2007), iniciou-se o desenvolvimento do sistema web denominado ALiBWeb, que será detalhado na subseção seguinte.

#### 4.2 O sistema ALiBWeb

O sistema ALiBWeb possui uma área de acesso público e uma área de acesso restrito e estará, em breve, disponível gratuitamente na Internet, permitindo a geração de cartas linguísticas como produto final. As cartas são geradas por linguistas e disponibilizadas publicamente.

Através do ALiBWeb, o usuário pode selecionar os dados de que precisa utilizando filtros. Assim, ele pode selecionar os dados a partir de perguntas ou respostas aos questionários e escolher o retorno dos dados (em valores absolutos e em percentuais) por localidade, sexo, faixa etária e/ou escolaridade. É possível ainda manipular os dados realizando agrupamentos. Por exemplo, para a pergunta 202 – “Como se chama um lugar pequeno, com um balcão, onde os homens costumam ir beber \_\_\_\_\_ (cf. item 182) e onde também se pode comprar alguma outra coisa” (COMITÊ NACIONAL..., 2001, p. 48) –, são respostas documentadas, dentre outras, *bar*, *barzinho*, *boteco* e *bodega*. O usuário pode, então, agrupar *bar* e *barzinho*, por exemplo, se for do seu interesse, como uma única lexia. O sistema então contará *bar* e *barzinho* como uma única variante em relação às demais e ajustará o total de ocorrências e os percentuais.

Nos testes do ALiBWeb, é possível documentar as respostas para a pergunta 191 – “Como se chama aquilo que as mulheres passam no rosto, nas bochechas, para ficarem mais rosadas?” (COMITÊ NACIONAL..., 2001, p. 37). São respostas registradas para essa questão *ruge*, *blush* e *carmim*, como mostra a carta L26, publicada no volume 2 do *Atlas linguístico do Brasil* (CARDOSO, 2014b, p. 331), reproduzida na Figura 3, em que o vermelho é *ruge*, o azul é *blush* e o amarelo são outras variantes, dentre as quais *carmim*.

**Figura 3** – Denominações diatopicamente representativas para *ruge* – Carta L26 –

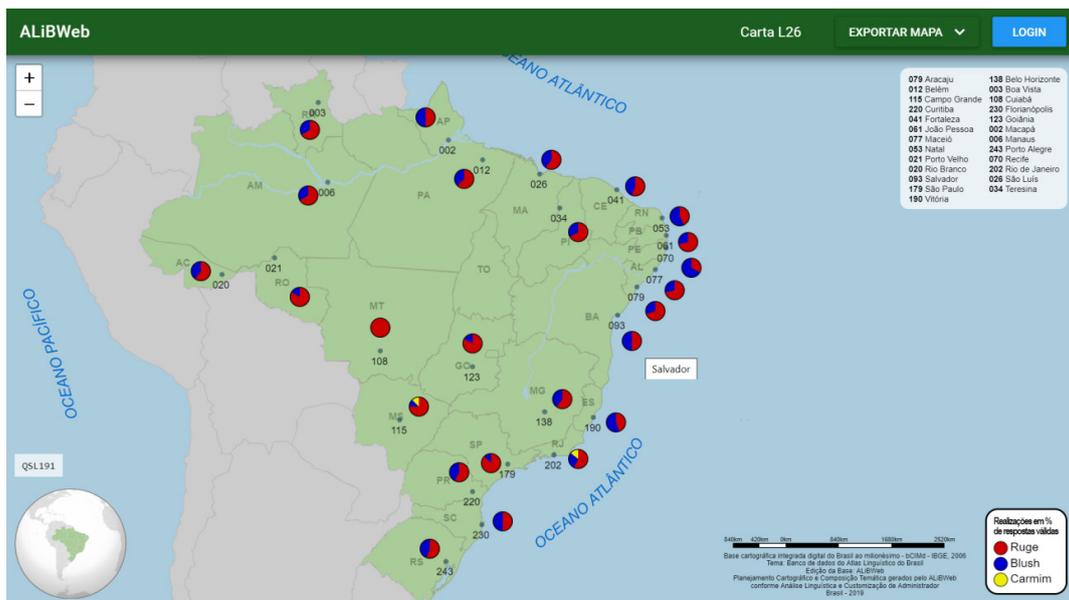
capitais de estado



Fonte: Cardoso et al. (2014b, p. 331).

Após a seleção dos dados e o retorno dos resultados numéricos, no ALiBWeb, pode-se gerar uma carta linguística, como ilustrado na Figura 4, a seguir, em que o vermelho é *ruge*, o azul é *blush* e o amarelo são outras variantes, dentre as quais *carmim*:

Figura 4 – Carta gerada pelo ALiBWeb



Fonte: Carta gerada pelas autoras no ALiBWeb.

Comparando as cartas apresentadas nas figuras 4 e 5, vê-se que os resultados são

equivalentes, ou seja, o sistema ALiBWeb reproduz fielmente as cartas já publicadas no volume 2 do ALiB, o que atesta a sua funcionalidade no que diz respeito ao acesso ao banco de dados. Além de gerar as cartas, o sistema apresenta os dados com o número de ocorrências e os percentuais de cada variante selecionada, além de permitir filtros e possíveis agrupamentos dos dados.

Além do acesso aos dados e da geração das cartas linguísticas, o usuário, desde que seja autorizado de acordo com a Política de Acesso do ALiB, pode ainda acessar a ficha do informante e outras informações, como data da realização do inquérito, inquiridores etc. Também é possível editar a legenda das cartas (cores e nomes dos agrupamentos) e escolher a ordem de colocação das esferas, se por maior frequência absoluta ou por maior número de localidades em que a forma ocorreu.

### **5 O que o futuro nos reserva?**

A equipe do Projeto ALiB reconhece as suas limitações no que tange à constituição de uma amostra nacional: (i) o número de localidades; (ii) a quantidade de informantes por localidade (oito nas capitais e quatro nas cidades do interior); (iii) as faixas etárias extremas (Faixa 1 – de 18 a 30 anos e Faixa 2 – de 50 a 65 anos); (iv) a consideração da escolaridade apenas nas capitais (no interior todos os informantes são de nível fundamental); (v) a má qualidade de alguns áudios; (vi) a má condução de algumas entrevistas; (vii) a não obtenção de algumas respostas; (viii) a necessidade de reformulação de algumas perguntas, que, hoje, implicam preconceitos.

É importante também ressaltar duas grandes dificuldades enfrentadas pela equipe do projeto: (i) a enorme extensão do país; e (ii) a falta de financiamento suficiente para a pesquisa.

Todavia as perspectivas do projeto são alentadoras. Com os ajustes que estão sendo feitos no sistema ALiBWeb, novos módulos devem ser agregados, tornando viável a incorporação dos áudios para a elaboração de mapas sonoros. Há muitas análises linguísticas em andamento que serão apresentadas nos próximos volumes que serão publicados. Os dados integrantes do banco de dados do Projeto ALiB têm servido como fonte de material para análises também na área da Fraseologia e na área da Ciência da Computação, ampliando seu escopo inicial. E, finalmente, aventa-se a possibilidade de ampliação da amostra, com a execução de novos inquéritos.

Enfim, a contribuição social do Projeto ALiB reside não só na produção da ciência

(na pesquisa de professores universitários, de estudantes de pós-graduação e de graduação) mas também na sua divulgação ao público em geral. E o sistema ALiBWeb permite o acesso e a manipulação dos dados do banco do ALiB para a geração e a visualização de cartas linguísticas com os resultados do projeto como produto, colaborando com o conhecimento e a difusão da diversidade linguística do país, rica e diversificada em usos, mas ainda pouco (re)conhecida pela sociedade extramuros das universidades brasileiras.

## 6 Considerações finais

O Projeto ALiB destaca-se pela contribuição social e pelo aporte que pode trazer ao estudo da diversidade da língua portuguesa, permitindo o conhecimento da variação espacial, explicitando as diferenças e convergências que se registram no território nacional, relacionando áreas dialetais a áreas socioculturais e oferecendo um conjunto de dados linguísticos que venham a contribuir para o conhecimento e a descrição da diversidade linguística do país e para o fornecimento de dados para o aperfeiçoamento do ensino do português.

Os volumes iniciais publicados do *Atlas linguístico do Brasil* são de fundamental importância para o entendimento da variação linguística com base na realidade da língua, de forma a eliminar preconceitos e discriminações sociais, mostrando como convivem diferenças e convergências, de acordo com as normas das diferentes comunidades, reconhecendo, porém, a validade da existência de variedades, conforme as normas das comunidades.

Além disso, os volumes 1 e 2 do atlas fornecem elementos para a construção da história da língua portuguesa no Brasil, quer pela indicação dos caminhos seguidos, quer pela natureza das opções de mudança feitas, quer pelo estabelecimento de camadas caracterizadas linguística, social e geograficamente, quer, ainda, pela referência a resultados de contato com outras línguas ou à adoção de empréstimos linguísticos. Assim, os atlas permitem a atualização de dicionários da língua portuguesa bem como auxiliam a construção de uma gramática voltada para a diversidade do português brasileiro.

Cientificamente, o banco de dados do Projeto Atlas Linguístico do Brasil terá contribuições tanto para a área da Dialectologia quanto para a área da Ciência da Computação, visto que já há dissertações de mestrado e teses de doutorado sendo desenvolvidas em ambas. Espera-se que, com a informatização desse atlas, novas teses em ambas as áreas possam ser desenvolvidas, contribuindo para o crescimento da pesquisa na área de Dialectologia e, também, na área de Processamento de Linguagem Natural, especifica-

mente na Extração da Informação.

O resultado imediatamente esperado do ALiBWeb é, evidentemente, o Atlas Falante, no qual será possível a audição da voz do informante no momento em que foi inquirido, além das cartas linguísticas dinâmicas, geradas de acordo com os requisitos do usuário do ALiBWeb. O desenvolvimento do ALiBWeb tem sido longo por causa da escassez de recursos financeiros, dificultando o desenvolvimento mais rápido desse sistema. No entanto a trajetória percorrida até o presente momento revela a perseverança em desenvolver esse sistema, que possibilitará o gerenciamento de dados orais de natureza geolinguística importantes para a academia e para a sociedade em geral.

## Referências

- AGUILERA, Vanderci de Andrade; ROMANO, Valter Pereira (Org.). *A geolinguística no Brasil: caminhos percorridos, horizontes alcançados*. Londrina. EDUEL, 2016.
- CARDOSO, Suzana Alice et al. *Atlas linguístico do Brasil*. v. 1: Introdução. Londrina: EDUEL, 2014a.
- CARDOSO, Suzana Alice et al. *Atlas linguístico do Brasil*. v. 2: Cartas linguísticas I. Londrina: EDUEL, 2014b.
- CHEN, Peter. *Modelagem de dados – a abordagem entidade-relacionamento*. São Paulo: Makron Books, 1990.
- COMITÊ NACIONAL do Projeto ALiB. *Atlas lingüístico do Brasil: questionários 2001*. Londrina: EDUEL, 2001.
- ELMASRI, Ramez; NAVATHE, Shamkant. *Sistemas de banco de dados*. 7. ed. São Paulo: Pearson Education, 2019.
- FERREIRA, Carlota; CARDOSO, Suzana Alice. *A dialetologia no Brasil*. São Paulo: Contexto, 1994.
- GONZAGA, Jorge Luiz. *Dominando o PostgreSQL*. Rio de Janeiro: Ciência Moderna, 2007.
- LABOV, William. *Padrões sociolinguísticos*. Trad. de Marcos Bagno, Marta Scherre e Caroline Cardoso. São Paulo: Parábola, 2008 [1972].
- MOTA, Jacyra Andrade; CARDOSO, Suzana Alice. Sobre a dialetologia no Brasil. In: MOTA, Jacyra Andrade; CARDOSO, Suzana Alice (Org.). *Documentos 2*. Projeto Atlas

Lingüístico do Brasil. Salvador: Quarteto, 2006. p. 15-26.

NASCENTES, Antenor. Études dialectologiques du Brésil. *ORBIS – Bulletin International de Documentation Linguistique*, Louvain, t. 1, n. 1, p. 181-184, 1952.

NASCENTES, Antenor. Études dialectologiques du Brésil. *ORBIS – Bulletin International de Documentation Linguistique*, Louvain, t. 2, n. 2, p. 438-444, 1953.

SMART, Mark. *Ruby on rails 5: web app development for beginners*. Seiten/Verlag: Createspace, 2016.

TELES, Ana Regina Torres Ferreira. *Cartografia e georreferenciamento na geolinguística: revisão e atualização das regiões dialetais e da rede de pontos para a elaboração do Atlas Lingüístico do Brasil formuladas por Antenor Nascentes*. 2018. 483f. Tese (Doutorado em Língua e Cultura) – Universidade Federal da Bahia, Salvador, 2018.

THUN, Harald. La géographie linguistique romane à la fin du XX<sup>e</sup> siècle. In: CONGRÈS INTERNATIONAL DE LINGUISTIQUE ET DE PHILOGIE ROMANES, 1998, Bruxelles. *Actes...* Tübingen: Max Niemeyer, 2000, v. 3, p. 367-388.



Data de submissão: 29/10/2020  
Data de aceite: 29/06/2021