




## Metodologia de criação de um banco de dados linguísticos: desafios e contribuições para o processo de ensino-aprendizagem

Methodology for creation of a linguistic database: challenges and contributions to the teaching-learning process

**Raimundo Gouveia da Silva**  <https://orcid.org/0000-0003-4582-6140>  
Instituto Federal de Educação, Ciências e Tecnologia do Acre – IFAC  
raimundo.silva@ifac.edu.br

**Iandra Maria Weirich da Silva Coelho**  <http://orcid.org/0000-0003-3513-962X>  
Instituto Federal de Educação, Ciências e Tecnologia do Amazonas – IFAM  
iandrawcoelho@gmail.com

### Resumo

Este artigo apresenta questões teórico-metodológicas relacionadas à criação de um Banco de Dados Linguísticos, composto por amostras da Reserva Extrativista Cazumbá Iracema, localizada no Acre e discute os principais desafios encontrados e contribuições para o processo de ensino e aprendizagem da Língua Portuguesa. A metodologia de coleta e organização desse Banco de Dados ancora-se nos pressupostos teóricos dos padrões sociolinguísticos, nos fundamentos empíricos da Teoria da Variação e Mudança Linguística e na metodologia de coleta e manipulação de dados em sociolinguística. A implementação da proposta envolve o uso de *softwares* que podem ser utilizados no âmbito educativo. Os resultados evidenciam contribuições do uso dessa amostra para a criação de propostas de ensino, com foco na língua em uso, identificação dos fatores socioculturais que influenciam o surgimento e permanência da variação linguística e realização de pesquisas no âmbito das línguas naturais.

**Palavras-chave:** Ensino. Banco de Dados. Língua Portuguesa. Variedade Linguística.

### Abstract

This paper presents theoretical and methodological questions related to the creation of a Linguistic Database, made up of samples from the Cazumbá Iracema Extractive Reserve, located in the state of Acre, and discusses and discusses the main challenges encountered found and contributions to the teaching and learning process of Portuguese. The methodology for collecting and organizing this database is based



on the theoretical assumptions of sociolinguistic patterns, the empirical foundations of the Theory of Linguistic Variation and Change, and the methodology for collecting and manipulating data in sociolinguistics. The implementation of the proposal involves the use of software that can be used in education. The results show contributions of the use of this sample for the creation of teaching proposals, focusing on the language in use, identification of the sociocultural factors that influence the emergence and permanence of linguistic variation and researches in the scope of natural languages.

**Keywords:** Teaching. Linguistic Database. Portuguese language. Linguistic variation.

## Considerações iniciais

Este artigo traz um recorte da dissertação intitulada “Variação linguística na língua portuguesa: uma proposta de ensino dos modos verbais com uso de banco de dados linguísticos”, realizada no Mestrado Profissional em Ensino Tecnológico, do Instituto Federal de Educação, Ciências e Tecnologia do Amazonas.

Por tratar-se de um Curso de Pós-Graduação na área de Ensino, na modalidade profissional, a pesquisa realizada teve como principal intuito a produção e aplicação de conhecimentos orientados para a prática e resolução de problemas, por meio do planejamento, implementação e avaliação de processos e produtos educacionais (BARROS; VALENTIM; MELO, 2005).

A partir dessa pesquisa, foram desenvolvidos dois produtos educacionais. O primeiro refere-se a um banco de dados linguísticos, oriundo de entrevistas realizadas no núcleo da Reserva Extrativista Cazumbá Iracema, localizado no Acre (doravante Banco Reci), que serviu como material de referência para a elaboração de uma proposta de ensino-aprendizagem dos paradigmas verbais, indicativo e subjuntivo da língua portuguesa. O segundo é um tutorial para uso do Banco Reci.

Nesse cenário, temos como principal objetivo evidenciar questões teórico-metodológicas, apresentando os principais elementos para a coleta e organização do Banco Reci, assim como reflexões sobre a importância do uso desse tipo de ferramenta, constituída como um produto educacional para o processo de ensino-aprendizagem da Língua Portuguesa.

As discussões levam em conta, especialmente a significância em abordar o contexto da variação linguística, seja no âmbito regional e/ou nacional, na perspectiva de propor novas metodologias e uso de diferentes recursos tecnológicos para trabalhar com conteúdos escolares.

Justifica-se abordar tal temática levando em conta que no Brasil, a compilação e uso de Banco de Dados, com foco no processo de ensino-aprendizagem da língua materna ainda é embrionária, com poucos estudos que destacam tal relevância (SIMÕES; MELO, 2008 apud MONGUILHOTT et al., 2016). Dessa forma, buscamos evidenciar algumas das principais contribuições da criação e uso de um Banco de Dados de fala para o ensino da língua materna, no intuito de apresentar elementos que possam evidenciar a necessidade e a significância em constituir amostras que podem servir como acervo para a criação e desenvolvimento de propostas de ensino, concebidas para diferentes conteúdos escolares, a partir do material sociolinguístico



coletado, além de servir como “fonte privilegiada para a descrição do português brasileiro” (FREITAG; MARTINS; TAVARES, 2012, p. 918).

Os procedimentos teóricos-metodológicos que envolvem a coleta e organização do Banco Reci, estão amparados nos padrões sociolinguísticos de Labov (2008); concepções relacionadas à sociolinguística variacionista (FREITAG, MARTINS; TAVARES, 2012); fatores que implicam na metodologia de coleta e manipulação de dados em sociolinguística (FREITAG, 2014), assim como aspectos técnicos na coleta de dados linguísticos orais (OLIVEIRA JUNIOR, 2014).

O artigo encontra-se dividido em duas seções. Na primeira, discorremos sobre a pesquisa sociolinguística e a criação de Banco de Dados no Brasil. Na segunda seção, elucidamos as principais etapas utilizadas para a concepção e criação do Banco de Dados Linguísticos Reci, que incluem: i) a concepção do Banco Reci; ii) a adoção dos pressupostos teóricos; e iii) a definição dos passos metodológicos para a coleta e manipulação dos dados. Além disso, apresentamos também uma discussão com relação as principais contribuições e desafios encontrados no processo metodológico de construção do produto educacional.

## **A pesquisa sociolinguística e a criação de Bancos de Dados**

Na pesquisa sociolinguística brasileira, o Banco de Dados é considerado a espinha dorsal, uma vez que em um Banco de Dados podemos encontrar tendências de correlação entre as variáveis estudadas na comunidade pesquisada, ou realizar uma comparação com outras comunidades em virtude da padronização na metodologia da coleta de dados (FREITAG, 2015).

Um Banco de Dados é um conjunto de informações armazenadas sobre pessoas, lugares, de maneira estruturada. Esta organização tem por finalidade estabelecer o maior número possível de informação e trazer mais eficiência durante a realização de uma pesquisa. Nesse sentido, Simões e Melo (2008, p. 10), destacam que:

É justamente reconhecendo a importância de nossa diversidade cultural e linguística e os múltiplos olhares investigativos que ressaltamos a importância do Banco de Dados como fonte de pesquisa e memória disponibilizada. Recurso imprescindível quando direcionamos nossas inquietações para a expansão da língua portuguesa e para a inclusão das diferentes comunidades.

Com base nas considerações de Simões e Melo (2008), justificamos a criação e uso de tais Bancos, no âmbito das pesquisas sociolinguísticas que, por meio de estudos empíricos sobre a heterogeneidade da fala em uma comunidade, tem por objeto de estudo os padrões de comportamento demonstrados, através da fala e que podem contribuir para o surgimento e a manutenção do português brasileiro, uma vez que as “mudanças na língua podem assim, estar correlacionadas com mudanças na posição dos subgrupos com os quais o falante se identifica” (LABOV, 2008, p. 327).

Com isso, temos uma comunidade de fala que tem por princípio o agrupamento de pessoas que compartilham costumes, crenças, valores e, principalmente, relações interpessoais, independentemente dos traços linguísticos. Para Labov (2008), esse agrupamento de pessoas não compartilha necessariamente as mesmas variáveis



linguísticas, mas os mesmos valores dessas variáveis, sendo reconhecidos por seus pares como membros dessa comunidade.

A constituição de uma amostra que busca considerar tais variedades do português brasileiro leva em conta características acerca da seleção dos seus informantes e cada projeto que constitui seu Banco de Dados busca atender a um conjunto de procedimentos metodológicos padronizados, a fim de possibilitar a realização de novos estudos e a descrição mais acurada da amostra que pode ser utilizada para a análise de diferentes fenômenos linguísticos.

Nesse sentido, no que diz respeito à constituição do Banco Reci, cabe destacar que o mesmo foi organizado de acordo com características sociodemográficas, enfatizando um padrão regular de estratificação socioeconômica das variáveis e as correlações entre tais variáveis linguísticas e categorias sociais primárias, como sexo, idade, escolaridade, entre outros.

A organização desse produto teve como principal intuito servir como uma amostra para auxiliar no processo de ensino-aprendizagem de Língua Portuguesa. De maneira geral, com relação à prática pedagógica, “alguns professores estão acostumados ao hábito de sistematização gramatical e utilizando modos ultrapassados de concepção de ensino de Língua Portuguesa” (BRITO; MATTOS; PISCIOTTA, 2003, p. 20). Nesse cenário, destacamos como um dos desafios do ensino da língua, a dificuldade em definir qual linguagem deve ser usada em diferentes tipos de situações, bem como seus níveis de formalidade (ANTUNES, 2007).

Essa discussão leva em conta a necessidade de abordar diferentes variáveis sociais em sala de aula, tais como: faixa etária, sexo (masculino, feminino), escolaridade, *status* econômico, escolaridade etc., fatores inerentes às variáveis linguísticas (WEINREICH; LABOV ; HERZOG, 2006), a fim de demonstrar como esses fatores influenciam nas variedades dialetais. Nesse percurso, a língua passa a valorizar a heterogeneidade dos fatores sociais, deixando de ser livre e aleatória, tendo em vista os vários fatores que contribuem para o surgimento e a manutenção do modo de falar.

Dessa forma, quando o aluno tem a consciência do seu modo de falar pode ter mais autonomia para apropriar-se das regras da norma culta e usá-las quando for necessário (CALLIAN; BOTELLHO, 2014), possibilitando a partir dessa realidade linguística, potencializar uma aprendizagem autônoma, desafiadora, prazerosa e significativa, pois, pode contribuir na formação de cidadãos críticos, criativos e reflexivos sobre sua própria língua e variedades.

Isso implica ressaltar a relevância em trabalhar com ênfase na variação linguística, pois, além de proporcionar um rico espaço de aprendizagem, é “inerente à língua”, e faz-se necessário, portanto, “discutir sobre ela em sala de aula” (PEREIRA; SILVA; SILVA, 2013, p. 108).

Nessa perspectiva, considerando essa prática docente, ressaltamos a importância da criação de propostas de ensino-aprendizagem que tenham como ponto de partida o vernáculo natural dos estudantes, com o objetivo de compreender melhor os fatores que colaboram para o surgimento e a manutenção da linguagem. Para isso acontecer de forma satisfatória e autônoma é necessário que os “professores de Língua Portuguesa tenham conhecimento de que a linguística faz parte da vida de cada um e saibam como trabalhá-la dentro da sala de aula, de maneira a estimular o interesse de seus alunos” (PEREIRA, SILVA; SILVA, 2013, p. 111).



Dessa forma, faz-se necessário um trabalho diferenciado com uma abordagem mais reflexiva de se trabalhar com a língua, num contexto de investigação, de forma que o aluno possa se tornar um pesquisador de sua própria linguagem, identificando os elementos que influenciam seu falar cotidiano. Nesse sentido, o professor tem o papel de “fazer da sala de aula um laboratório de linguagem e atribuir aos alunos o papel de investigadores linguísticos” (GÖRSKI; COELHO, 2009, p. 84).

## **Percurso metodológico para a concepção e criação do Banco de Dados Linguísticos Reci**

Nessa seção, buscamos evidenciar a concepção, elaboração, validação e metodologias empregadas na construção do Banco Reci, construído como um dos produtos educacionais oriundos de uma pesquisa aplicada.

O Banco Reci teve como propósito constituir uma amostra significativa de dados linguísticos oriundos de povos tradicionais. O propósito principal foi coletar material que serviu como fonte para elaboração de uma proposta de ensino para a Língua Portuguesa. Contudo, vale ressaltar que a amostra pode ser utilizada para a constituição de outras propostas de ensino-aprendizagem.

De maneira geral, a ideia inicial da pesquisa era desenvolver um produto educacional que pudesse contribuir, especialmente, com professores e estudantes de comunidades tradicionais, que trabalham com a língua portuguesa desde uma perspectiva prescritiva. O produto educacional surgiu então, desde uma concepção formativa para servir como fonte para a criação de uma proposta de ensino para o estudo dos paradigmas verbais indicativo e subjuntivo, com ênfase no trabalho da variação linguística.

Outras contribuições visam: i) auxiliar no trabalho pedagógico; ii) potencializar o ensino da língua materna no contexto da sala de aula; iii) fomentar novas formas de ensinar e aprender a língua materna; e iv) servir como fonte para a elaboração de processos e produtos didáticos relacionados ao ensino-aprendizagem de fenômenos da Língua Portuguesa.

Acreditamos que o Banco Reci também pode contribuir na ampliação dos Bancos de dados de fala, especialmente como referência para a região norte, como fontes privilegiadas para a descrição da língua. Pode auxiliar, além de professores, pesquisadores que queiram realizar pesquisas em diferentes áreas e propiciar uma visão ampla das variedades linguísticas no âmbito nacional, com foco especialmente para a coleta e divulgação da fala dos povos tradicionais e para a constituição dos perfis sociolinguísticos dessas variedades que preconizam a língua, compreendida desde uma perspectiva variacionista e mutável.

Isso implica ampliar as amostras de dados e bancos independentes, tais como: i) Bancos de Dados do Rio de Janeiro, PEUL - ii) Banco de Dados VARSUL - Variação linguística na região Sul do Brasil e iii) O Banco de Dados VALPB - Variação linguística no Estado da Paraíba, entre outros, a fim de propiciar uma visão ampla das variedades do português, que podem auxiliar em diferentes linhas de investigação, na perspectiva da sociolinguística variacionista.





Aspectos inovadores da criação e aplicação desse produto estão relacionados ao ensino e a adaptação de uma amostra autêntica da variedade linguística de povos tradicionais, contemplando uma variedade do português brasileiro aos conteúdos curriculares.

Mencionadas as contribuições e possibilidades de uso, destacamos também as principais questões teórico-metodológicas para a concepção, elaboração, validação e divulgação desse produto educacional.

De maneira sintética, destacamos entre os principais procedimentos, a realização de uma coleta e organização do Banco Reci, tomando como variáveis a faixa etária, o sexo e a escolaridade, com base nos padrões sociolinguísticos, nos fundamentos empíricos para uma teoria da mudança linguística e na sociolinguística variacionista. Na coleta dos dados linguísticos, foram entrevistados 36 informantes, divididos em três células sociais, formando os critérios para a obtenção dos dados: sexo (masculino e feminino); idade (a partir dos 15 anos); escolaridade (Educação Básica e não alfabetizado).

Cabe ressaltar que os procedimentos levaram em conta algumas recomendações para a realização do trabalho de campo sociolinguístico que serviram como um roteiro instrucional para direcionar a coleta, posteriormente transcrita e divulgada em formato eletrônico e impresso. Entre as recomendações básicas, seguimos:

- i) Seleção do local para coleta dos dados: na escolha da localidade, considera-se a existência de variedades linguísticas que possibilitam ao pesquisador observar, por meio delas, a construção de uma identidade dos indivíduos;
- ii) Seleção dos sujeitos da pesquisa: escolha coerente dos informantes da comunidade para constituição da amostra com a pesquisa que se quer realizar;
- iii) Elaboração do roteiro de perguntas: pode-se tomar como base outros questionários já adotados por pesquisadores na compilação de bancos de dados;
- iv) Submissão do projeto de pesquisa e aprovação pelo Comitê de Ética;
- v) Preenchimento da ficha social pelos informantes: As fichas sociais reúnem informações sobre idade, estado civil, local de nascimento, quantidade de filhos, religião, profissão, atividades sociais realizadas pelo informante nas horas de lazer, entre outras questões socioeconômicas e culturais.
- vi) Coleta de dados de fala: entrevistas realizadas segundo metodologia laboviana, estratificadas segundo padrões sociais pré-estabelecidos, tomando como base alguns critérios comumente empregados para constituição de uma amostra nas pesquisas sociolinguísticas que são as variáveis: sexo, faixa etária e escolaridade, direcionando o enquadramento e agrupamento dos informantes em células sociais. As entrevistas individuais são baseadas em um roteiro de perguntas.
- vii) Entrega e assinatura do Termo de Consentimento Livre e Esclarecido contendo informações sobre a pesquisa (COELHO, 2018, p. 169-70).

Com relação às etapas utilizadas, destacamos a primeira delas, que se refere à concepção e a constituição do Banco Reci, que foi realizada por meio de entrevistas semiestruturadas, com base nas seguintes variáveis: faixa etária, escolaridade e sexo.

A proposta de compilação desse Banco ancora-se nos pressupostos teóricos dos padrões sociolinguísticos de Labov (2008) e fundamentos empíricos da Teoria da Variação e Mudança Linguística de Weinreich, Labov e Herzog (2006). A coleta e organização dessa amostra estão delineadas na metodologia de coleta e manipulação de dados em sociolinguística de Freitag (2014).

Nesse sentido, a pesquisa foi concebida num cenário em que “Bancos de Dados Linguísticos de fala (especialmente os que seguem a orientação da sociolinguística



variacionista) têm sido fonte privilegiada para a descrição do português brasileiro” (FREITAG; MARTINS; TAVARES, 2012, p. 918).

Com base nos pressupostos da pesquisa sociolinguística, selecionamos o núcleo da Reserva Extrativista Cazumbá Iracema por acreditarmos que no meio rural encontra-se, provavelmente, uma das falas mais conservadoras, o que, por sua vez, poderá servir de objeto de análise para outros pesquisadores contribuindo, dessa forma, para a compreensão dos fatores socioculturais e de traços e perfis sociolinguísticos das variedades faladas nesta comunidade.

Entre as considerações avaliadas antes da coleta, foi necessário selecionar a comunidade de fala, a determinação das variáveis da análise e o número de informantes. A amostra em questão foi coletada a partir de três células sociais que formam os critérios da pesquisa e são categorias tradicionalmente consideradas pela sociolinguística variacionista: sexo (masculino e feminino), faixa etária (a partir dos 15 anos), escolaridade (Educação Básica e não alfabetizado), além de outras características gerais: ter nascido na comunidade, não ter se ausentado por mais de dois anos no período da adolescência e ser aceito pelos moradores do núcleo como membro da comunidade, como exposto no Quadro 1.

Quadro 1. Critérios socioculturais para a elaboração da entrevista

<b>FATORES SOCIOCULTURAIS</b>	
<b>Sexo</b>	Masculino e Feminino
<b>Faixa etária</b>	Faixa I - 12 informantes de (15 a 20 anos). Faixa II - 12 informantes de (25 a 40 anos). Faixa III - 12 informantes de (a partir dos 45 anos).
<b>Continuação Escolaridade</b>	Faixa I – Informantes cursando o Nível Médio na comunidade. Faixa II – Informantes com Ensino Fundamental II incompleto. Faixa III – Informantes com Ensino Fundamental I incompleto ou sem instrução escolar.
<b>Características Gerais</b>	Ter nascido na comunidade e não ter se ausentado por mais de 2 anos. Ter mais de 14 anos. Pais naturais da comunidade ou de outras comunidades, mas com mais de cinco anos morando na reserva. Intervalo de quatro anos entre as faixas etárias. Serem aceitos pelos outros pares como membro da comunidade de fala.

Fonte: Os autores.

Esses cuidados metodológicos adotados levam em conta que em uma pesquisa para criação de um Banco de Dados, os pesquisadores não podem organizar os informantes e elaborar os questionários apenas para alcançar seus objetivos, como afirmam Monguilhott et al. (2016, p. 114):

Os pesquisadores não realizam uma coleta propriamente para ser fonte de suas investigações particulares, mas os projetos em todo o país constituem seus Bancos de Dados considerando aspectos mais amplos justamente para dar conta de investigações nos diferentes níveis da língua e com diferentes abordagens da sociolinguística, considerando também suas interfaces.



Nessas circunstâncias, todas as pesquisas sociolinguísticas adotam padrões metodológicos que visam descobrir e elencar o máximo de informações possíveis de seus informantes, destacando os fatores internos e externos que colaboram para o surgimento e a permanência de uma comunidade de fala, já que são essas informações que podem ser objetivos de investigações por outros pesquisadores.

A escolha do perfil dos informantes teve como base alguns critérios, à luz da teoria laboviana. Foram selecionados falantes nascidos e residentes na comunidade do núcleo da Reserva Extrativista Cazumbá Iracema, com pais naturais também da comunidade, ou de outras comunidades, porém, com mais de cinco anos residindo na reserva, a fim de dificultar o aparecimento de influências de outras nacionalidades nos resultados.

Para obtenção dos dados, foram entrevistados 36 membros do núcleo da Reserva, estratificados em três grupos. No primeiro grupo fazem parte 12 informantes na faixa etária de 15 a 20 anos, cursando o nível médio na respectiva comunidade, preferencialmente, seis membros do sexo masculino e seis do sexo feminino. Para o segundo grupo de informantes, optamos por 12 membros do núcleo da reserva, na faixa etária de 25 a 40 anos, com o Ensino Fundamental II incompleto, sendo também dividido por sexo. E o terceiro compreende 12 pessoas de ambos os sexos, a partir dos 45 anos. Nesse grupo, entrevistamos informantes sem instrução escolar e/ou pessoas com o Ensino fundamental I incompleto.

De acordo com a amostragem exemplificada, finalizamos com um agrupamento de dezoito células sociais, considerando os procedimentos clássicos labovianos, (LABOV, 2008) que determinam a multiplicação entre os fatores das variáveis sociais. Assim, teríamos 2 sexos x 3 níveis de faixa etária x 3 níveis de escolaridade, totalizando 18 células sociais.

Vale ressaltar que a amostra ideal para essa estratificação deveria ser de cinco (5) indivíduos em cada célula social (MOLLICA; BRAGA, 2004), o que resultaria em um total de 90 informantes. Nesse viés, Guy e Zilles (2007 apud VITÓRIO, 2014, p. 54), destacam que:

Em uma pesquisa de cunho variacionista, o ideal é selecionar quatro ou cinco informantes em cada célula, para evitar, durante o momento da entrevista e constituição do *corpus* da pesquisa, um comportamento linguístico idiossincrático ou enviesado [...].

Embora a literatura aconselhe um número ideal para amostragem desse tipo de pesquisa, devido à escassez de tempo e recursos, tornou-se inviável a coleta de dados com esse número de indivíduos em cada célula, em virtude “a tantos obstáculos falta de financiamentos, dificuldades em se conseguir informantes com certos perfis, perda de entrevistas por problemas técnicos, etc.” (ARAÚJO; ALMEIDA, 2014, p. 44). Dessa forma, vale ressaltar que no Brasil, os Bancos de Dados costumam ser constituídos a partir de 2 a 3 informantes por célula social (TAVARES; MARTINS, 2014). Nesse sentido, visando minimizar os possíveis obstáculos e estando em consonância com as diretrizes nacionais no que tange ao número de informantes, decidimos levar em conta, para o levantamento dos dados, o número mínimo de informantes por célula, critério já amplamente adotado em Bancos de Dados como o PEUL, VARSUL e VALPB.

Além dos procedimentos metodológicos mencionados para o levantamento dos dados, também foi necessário: i) escolher o objeto de pesquisa; ii) verificar o número





de habitantes do núcleo da reserva; iii) solicitar a autorização dos gestores da comunidade para a realização das entrevistas; iv) buscar pessoas que tivessem maior contato com os moradores para facilitar as entrevistas; e v) submeter o projeto ao Comitê de Ética e ao SISBio; v) Dividir os informantes pelos critérios socioculturais da entrevista; e vi) especificar as variáveis que nortearam as entrevistas.

Vale ressaltar que para a realização da coleta de dados no núcleo da Reserva Extrativista Cazumbá Iracema e sua divulgação, por meio da criação do site [www.bancoreci.com.br](http://www.bancoreci.com.br), submetemos o projeto para a apreciação do Comitê de ética. Além disso, solicitamos autorização ao Instituto Chico Mendes de Conservação da Biodiversidade- ICMBio. No momento da coleta dos dados na Reserva, torna-se imprescindível a apresentação dessa autorização, pois é ela que ratifica a liberação dos órgãos responsáveis pela fiscalização. Em virtude de o *lôcus* de pesquisa ser uma reserva extrativista e seus sujeitos povos tradicionais, fez-se necessário também submeter o projeto ao Ministério do Meio Ambiente, com um pedido de autorização para atividade com finalidade científica no sistema de autorização e informação em biodiversidade – SISBio.

Com relação ao perfil dos informantes, os critérios adotados para essa escolha levaram em conta novamente a teoria laboviana. Para tanto escolhemos falantes nascidos e residentes na comunidade do núcleo da Reserva Extrativista Cazumbá-Iracema, com pais naturais também da comunidade, ou de outras comunidades, porém, com mais de cinco anos residindo na reserva.

Outro procedimento utilizado para a constituição do produto educacional foi a concepção e elaboração do roteiro da entrevista. Cabe mencionar que a entrevista foi utilizada como estratégia de investigação para a obtenção dos dados e apesar de reconhecermos suas inúmeras vantagens, também reconhecemos sua fragilidade, especialmente no contexto de estilo pergunta/resposta, uma vez que, de maneira geral, os informantes tendem a monitorar-se, não deixando fluir o seu verdadeiro vernáculo.

Com o propósito de extrair o máximo possível dos entrevistados, foram criados temas para subsidiar o diálogo com uma sequência lógica de perguntas, definido por Labov (2008, p. 26) como “conversação natural espontânea”. Por essa razão, elaboramos um questionário com vinte e cinco (25) perguntas, divididas em cinco (5) seções. Nessas seções, analisamos o envolvimento dos informantes na comunidade, na família – redes sociais, na infância, no trabalho – ocupação e sua produção linguística, com algumas perguntas que enfatizaram o uso de construções de base modal.

Para a coleta dessas entrevistas, utilizamos um gravador portátil, uma vez que o mesmo possui cartões de memória que captam menores ruídos durante a gravação, posto que é necessário ter muita atenção no momento da gravação, principalmente em relação a possíveis fontes de ruídos, como: pessoas com conversas paralelas, animais domésticos ou outros, aviões, geladeiras, ventiladores. Mesmo com todos esses cuidados, há momentos que o ruído é inevitável.

O problema de gravar a fala em ambientes naturais é uma questão técnica. O desenvolvimento de gravadores profissionais a pilha tornou possível obter excelentes resultados em campo. Com um bom microfone, um pesquisador de campo pode conseguir ótimas gravações sob condições barulhentas ao encurtar a distância entre a boca do informante e o microfone (LABOV, 2008, p. 239).



Os recursos tecnológicos oportunizam a coleta de um grande volume de fala natural bem gravada, uma vez que na sociolinguística a qualidade dos dados vai depender das características das gravações em campo. Entretanto, o pesquisador precisa ser dinâmico, criativo e usar os recursos tecnológicos disponíveis para minimizar os efeitos negativos dos ruídos quando realizar a coleta dos dados.

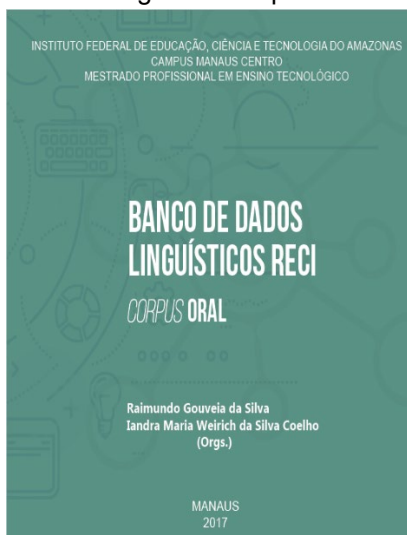
A última etapa foi a transcrição dos dados, realizada unicamente em uma linha, na qual registramos a sintaxe real da fala dos informantes do núcleo da Reserva Extrativista Cazumbá-Iracema, considerando as hesitações, interrupções e demais fatores, além da ortografia oficial e os sinais de pontuação.

A transcrição dos dados é um trabalho minucioso, que precisa de muita atenção para que a transcrição seja feita de acordo com a fala dos informantes, uma vez que “a análise adequada da variável linguística é o passo mais importante da investigação Sociolinguística” (LABOV, 2008, p. 93).

Além dessas estratégias, também foi utilizado um sistema gráfico com o objetivo de representar, na escrita, o discurso autêntico produzido na modalidade oral. A literatura reconhece dois sistemas que podem ser utilizados com estratégias de transcrição de dados: a escrita-padrão e a escrita modificada.

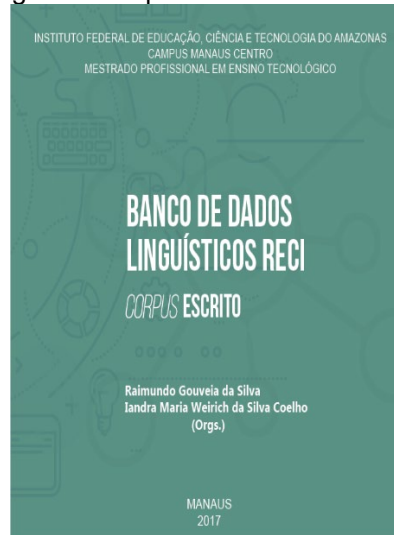
Ao concluir o processo de pesquisa, dividimos o Banco Reci compilado em duas amostras: um *corpus* oral e um escrito (Figuras 1 e 2). A amostra oral possui mais de três horas de gravações, digitalizadas e armazenadas eletronicamente e conta com 36 informantes, do núcleo da Reserva Extrativista Cazumbá Iracema, localizado no Acre. O *corpus* escrito, compilado durante os anos de 2016 e 2017, possui um acervo de 30.798 palavras (constituído de amostras de fala, que foram gravadas, transcritas e, posteriormente, armazenadas eletronicamente). Essa divisão foi justificada pela necessidade de manter, no *corpus* oral, o anonimato dos informantes, vinculando uma amostra de caráter restrito.

Figura 1. Corpus oral



Fonte: Os autores.

Figura 2. Corpus escrito



Fonte: Os autores.

A fim de ampliar o acesso ao produto educacional e à proposta desenvolvida a partir dele, criamos o *site* [www.bancoreci.com.br](http://www.bancoreci.com.br). Tivemos como propósito principal compartilhar essa amostra autêntica com outros docentes, principalmente de língua portuguesa, tendo em vista que foi concebida para auxiliar no trabalho pedagógico, potencializar novas formas de ensinar e aprender a língua materna, assim como contribuir com a comunidade científica para estudos descritivos da variação linguística, constituir e ampliar os bancos de dados sociolinguísticos e fomentar as pesquisas dessa natureza.

A criação desse *site* implica a divulgação da pesquisa realizada e compartilhamento de informações. O *site* contém parte da transcrição de algumas entrevistas, uma parte da proposta de ensino elaborada a partir do Banco Reci e as produções técnico-científicas oriundas da pesquisa, por meio de uma galeria de imagens, entre outros recursos (Figura 3)

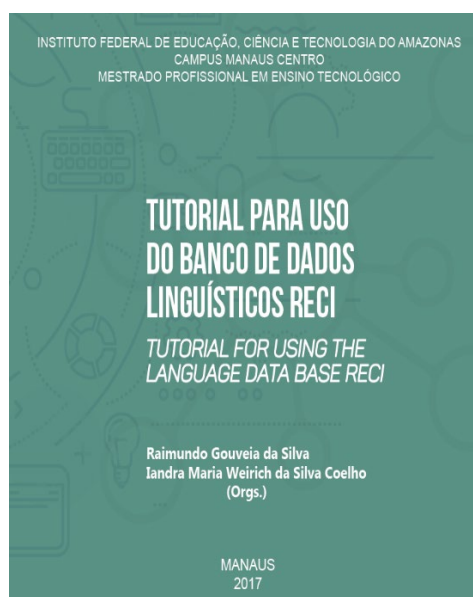
Figura 3. *Site* do banco Reci



Fonte: [www.bancoreci.com.br](http://www.bancoreci.com.br)

A fim de facilitar o acesso aos recursos do *site*, também foi criado um segundo produto educacional, um tutorial (Figura 4) que visa orientar o uso do Banco Reci.

Figura 4. Tutorial para uso do Banco Reci



Fonte: Os autores.

O tutorial visa orientar o uso do Banco de Dados Linguísticos, composto por amostras de fala do núcleo da Reserva Extrativista Cazumbá-Iracema, localizado no estado do

Acre. Foi concebido, principalmente, para professores de Língua Portuguesa e evidencia dados linguísticos que podem ser utilizados para propor novas maneiras de trabalhar com diferentes conteúdos escolares.

De maneira geral, a criação do *site*, bem como seu tutorial, que evidenciam a participação dos segmentos dos povos tradicionais na realidade sociolinguística brasileira, em relação a diferentes pontos de vista sob a constituição da realidade pluricultural da Língua Portuguesa no Brasil.

Cabe ainda ressaltar nesse estudo os principais desafios encontrados e que devem ser levados em conta por outros professores e pesquisadores que tenham interesse em utilizar as estratégias apresentadas para a concepção e construção de novas amostras.

Em síntese, destacamos alguns aspectos que levam em conta desde o planejamento e logística relacionada à coleta de dados.

- I. Logística na coleta dos dados: o acesso para chegar às comunidades tradicionais dificulta a coleta, como é o caso do núcleo da Reserva Extrativista Cazumbá-Iracema que está localizada em uma área de difícil acesso na zona rural de Sena Madureira.
- II. Localização dos informantes: dificuldades para encontrar tanto os jovens do sexo masculino de 15 a 20 anos que estivessem cursando o Ensino Médio, como os informantes do sexo feminino de 25 a 40 anos, com o Ensino Fundamental II incompleto.
- III. Coleta das entrevistas sem ruídos: as gravações foram realizadas ao ar livre, sujeitas a todos os tipos de ruídos. Por essa razão, tornou-se necessário realizar, antes de cada entrevista, um estudo dos possíveis barulhos do local. Além de algumas vezes parar com as entrevistas em virtude dos ruídos de cachorros, galinhas, pássaros, etc.
- IV. Transcrição dos dados: a transcrição foi feita com base na linguagem autêntica, ou seja, da forma real em que os informantes falaram. Dessa forma, foi necessário ouvirmos as entrevistas várias vezes, com o objetivo de sermos o mais fiel possível nas transcrições dos dados.

## Considerações finais

Neste artigo, apresentamos os principais elementos teóricos-metodológicos relacionados à criação do Banco de Dados Linguísticos Reci, composto por amostras de fala de populações tradicionais do núcleo da Reserva Extrativista Cazumbá-Iracema, localizada no Acre, as contribuições que esse produto educacional pode oferecer no processo de ensino-aprendizagem da Língua Portuguesa e os principais desafios encontrados para sua concepção e desenvolvimento.

Uma das principais contribuições refere-se ao potencial desse produto educativo para estimular o uso autêntico da língua no contexto de sala de aula e diminuir os preconceitos linguísticos, levando em conta o trabalho e reflexão sobre as variedades



da língua materna, assim como os conhecimentos socioculturais e linguísticos dos alunos.

Dessa forma, destacamos a importância do uso de dados reais e a ênfase do estudo das variantes linguísticas no trabalho com a Língua Portuguesa, assim como as variáveis que influenciam os vários usos que se faz da língua, dependendo da situação comunicativa. Cabe ressaltar que a relevância do uso dessa amostra para a criação de propostas de ensino, com foco na língua em uso, não significa que temos o intuito de desvalorizar a norma culta, mas ressaltar a importância de inserir no contexto da sala de aula novas possibilidades de ensino, de acordo com a realidade sociocultural e linguística dos alunos.

A possibilidade de novas práticas e intervenções que possam levar em conta a língua em suas diferentes variedades reflete uma importante contribuição social, relacionada principalmente, com os possíveis aportes que as práticas de ensino dessa natureza pretendem alcançar com este tipo de comunidades tradicionais. Trabalhar com dados autênticos desse público-alvo leva em conta não apenas as ressalvas com relação à importância da variedade da nossa língua materna, mas também de aspectos como a valorização da região amazônica, especialmente com relação à conservação dos recursos naturais, com destaque especial para a relação estreita que há entre estes povos e o ecossistema Amazônico e a vinculação da população que habita tradicionalmente esse diverso ecossistema.

Por meio dos dados coletados, é possível conhecer um pouco mais a realidade linguística brasileira e os aspectos sociais que a circunscrevem, permitindo evidenciar características importantes e dar a conhecer a localidade pesquisada a partir de suas falas, conhecer sobre sua história, a demarcação de territórios, o modo de lidar com os recursos naturais, entre outros fatores.

Nessa perspectiva, a amostra também pode ser significativa para outras áreas que discutem questões relacionadas ao uso da linguagem. Isso leva em conta aspectos relacionados ao contato entre as línguas; as variáveis sociolinguísticas; reflexões sobre a participação dos segmentos dos povos tradicionais na realidade sociolinguística brasileira, em relação a diferentes pontos de vista sob a constituição da realidade pluricultural da Língua Portuguesa no Brasil, bem como a manutenção da identidade étnica dos povos, a perda linguística; fenômenos relacionados ao uso da língua materna; entre outros fatores, possibilitando trabalhos no domínio dos estudos linguísticos, bem como na área da Antropologia, História, entre outros.

Por fim, sinalizamos esse contexto de pesquisa como desdobramento para trabalhos futuros que incluem a ampliação do Banco Reci com dados linguísticos dos cinco polos que compõem a Reserva Extrativista Cazumbá Iracema e de outras reservas extrativistas do Estado do Acre, com o objetivo de coletar e divulgar as variáveis linguísticas das reservas do Estado, a fim de aprofundar os estudos no âmbito da morfologia, fonética, fonologia, sintaxe e semântica e possibilitar outras investigações na área do ensino ou da perspectiva sociolinguística.

Dessa forma, “a ampliação de banco de dados é uma tarefa que nunca se esgota, pois precisamos acompanhar a trajetória da variação e da comunicação linguística ao longo do tempo” (MONGUILHOTT et al., 2016, p. 114).

## Referências





ANTUNES, I. **Muito além da gramática**: por um ensino de línguas sem pedras no caminho. São Paulo: Parábola, 2007.

ARAÚJO, S. S. F.; ALMEIDA, N. L. F. O projeto a língua portuguesa no semiárido baiano – fase 3: critérios de constituição e da amostragem do banco de dados. In: FREITAG, R. M. K. **Metodologia de coleta e manipulação de dados em sociolinguística**. São Paulo: Blucher, p. 27-47, 2014.

BARROS, E. C.; VALENTIM, M. C.; MELO, M. A. A. O debate sobre o mestrado profissional na Capes: trajetória e definições. **Revista Brasileira de Pós-graduação**, Brasília, v. 2, n. 4, p. 124-138, 2005. Disponível em: <<http://ojs.rbpg.capes.gov.br/index.php/rbpg/article/view/84/80>>. Acesso em: 7 mar. 2020.

BRITO, V. E.; MATTOS M. J.; PISCIOTTA, H. **PCNs de Língua Portuguesa**: a prática em sala de aula. São Paulo: Arte e Ciência, 2003.

CALLIAN, G. R.; BOTELHO, L. S. A análise linguística e o ensino de língua portuguesa: em busca do desenvolvimento da competência comunicativa. **Revista Eletrônica da Faculdade Metodista Granbery**, 16, p. 1-21, 2014. Disponível em: <http://re.granbery.edu.br/artigos/NDk4.pdf>. Acesso em: 05 mar. 2020.

COELHO, Iandra Maria Weirich da Silva. Metodologia de coleta de dados: uma proposta de protocolo para compilação de bancos de dados linguísticos. In: SOUZA, Ana Cláudia Ribeiro de et al. **Formação de professores e recursos e estratégias de ensino**. Curitiba: Appris, 2018

FREITAG, R. M. K.; MARTINS, M. A.; TAVARES, M. A. Bancos de dados sociolinguísticos do português brasileiro e os estudos de terceira onda: potencialidades e limitações. **Alfa: Revista de Linguística**, 56(3), 2012.

FREITAG, R. M. K. 2015. Desafios teóricos-metodológicos da sociolinguística variacionista. In: PARREIRA, Maria Cristina et al. **Pesquisas em Linguística no século XXI**: perspectivas e desafios teóricos-metodológicos. São Paulo: Cultura Acadêmica, p. 29-43, 2015. Disponível em: <https://www.fclar.unesp.br/Home/Instituicao/Administracao/DivisaoTecnicaAcademica/ApoioaoEnsino/LaboratorioEditorial/serie-trilhas-linguisticas-n27.pdf>. Acesso em: 11 fev. 2020.

FREITAG, R. M. K. (Org). **Metodologia de coleta em manipulação de dados em sociolinguística**. São Paulo: Blucher, 2014.

GÖRSKI, E. M.; COELHO, I. L. Variação linguística e ensino de gramática. **Working papers em Linguística**, 10(1), 73-91, 2009.

LABOV, W. 2008. **Padrões sociolinguísticos**. São Paulo: Parábola Editorial, 392 p.

MOLLICA, M. C.; BRAGA, M. L. **Introdução à sociolinguística**: o tratamento da variação. Editora Contexto, 2003.

MONGUILHOTT, I. O. S. et al. Metodologia de dados em escolas da rede pública e privada de ensino de Florianópolis. In: FREITAG, R. M. K.; SEVERO, C. G.; GÖRSKI,





E. M. **Sociolinguística e política linguística: olhares contemporâneos**. São Paulo: Blucher, p. 113-136, 2016.

OLIVEIRA Jr, M. Aspectos técnicos na coleta de dados linguísticos orais. In: FREITAG, R. M. K. (Org.), **Metodologia de coleta e manipulação de dados em sociolinguística**. São Paulo: Blucher, p. 20-26, 2014.

PEREIRA, B. G.; SILVA, J.; SILVA, J. B. M. Como trabalhar variação linguística e gramática em sala de aula: uma reflexão. **Revista Ribanceira**, v.1(1), p. 107-119, 2013.

SIMÕES, D.; MELO, E. M. A relevância dos Bancos de Dados para o ensino da Língua Portuguesa. **Revista Prolíngua**, 2(2), p. 12-24, 2008. Disponível em: <http://www.periodicos.ufpb.br/index.php/prolingua/article/viewFile/13396/7604>. Acesso em: 05 mar. 2020.

TAVARES, M. A.; MARTINS, M. A. O banco de dados fala-natal: uma agenda de trabalho. In: FREITAG, R. M. K. (Org.), **Metodologia de coleta e manipulação de dados em sociolinguística**. São Paulo: Blucher, p. 71-78, 2014.

VITÓRIO, E. G. S.; LIMA A. A língua falada em alagoas: coleta e transcrição dos dados. In: FREITAG, R. M. (Org.). **Metodologia de coleta e manipulação de dados em sociolinguística**. São Paulo: Blucher, p. 49-60, 2014.

WEINREICH, U.; LABOV, W.; HERZOG, M. **Fundamentos empíricos para uma teoria da mudança linguística**. São Paulo: Parábola, 2006.

---

**Recebido:** 27/08/19

**Aprovado:** 15/05/20

**Como citar:** SILVA, R. G.; COELHO, I. M. W. S. Metodologia de criação de um banco de dados linguísticos: desafios e contribuições para o processo de ensino-aprendizagem. **Revista de Estudos e Pesquisa sobre Ensino Tecnológico (EDUCITEC)**, v. 6, e090520, 2020.

**Direito autoral:** Este artigo está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.

