

# Introdução a pesquisas de sons não nativos

RONALDO LIMA JR.  
UBIRATÃ KICKHÖFEL ALVES  
FELIPE FLORES KUPSKE

## 1. POR QUE ESTUDAR SONS NÃO NATIVOS?

As discussões e percepções sobre o que caracteriza ‘sotaque’ vão muito além do objeto de estudo da Linguística. Linguistas ou não, todos os indivíduos mostram-se capazes de dizer, a seu modo, se a fala de outro indivíduo “tem ou não sotaque”. Esse sotaque, a que se refere o indivíduo leigo, pode dizer respeito tanto a diferenças dialetais das distintas variedades que constituem a língua materna (L1) de um indivíduo, bem como a diferenças que caracterizam a produção da fala em uma língua não nativa, sendo esse último aspecto o objeto de discussão da presente obra. O fato é que, independentemente do seu grau de instrução em Linguística, é muito mais fácil para o falante de uma língua apontar ‘sotaque’ em outros falantes do que em si. Nesse sentido, a percepção do que constitui ‘sotaque’, para o leigo, corresponde a uma questão de diferença entre a fala percebida e a do indivíduo que aponta tal diferença. Em outras palavras, a noção de ‘sotaque’, à primeira vista, poderia ser considerada como de senso comum: assim como os indivíduos se julgam capazes de opinar sobre política, religião e futebol, também todos são capazes de determinar quem “tem ou não tem sotaque”.

Ainda que as discussões referentes às capacidades comunicacionais de um indivíduo possam vir a ser abordadas à luz do senso comum, estudar a fala multilíngue a partir de um método bem delimitado, com base em uma teoria que sustente as afirmações a serem feitas, constitui uma tarefa de caráter indiscutivelmente científico. No caso da produção de sons de línguas não nativas, esse tipo de estudo é bastante desafiador, pois implica problematizar questões que, a princípio, poderiam ser vistas como “corriqueiras”, ou de “senso comum”. Para além de discutir se um sotaque é “muito diferente” ou “próximo” do nosso, ou se “é bonito” ou “feio” aos ouvidos de um indivíduo, o investigador da área volta-se ao entendimento de questões que envolvem o processamento cognitivo do sistema de sons, as condições físicas e os construtos mentais para a sua produção e percepção, o impacto que tal fala com sotaque pode ter para a comunicação, além de buscar meios de contribuir para a discussão acerca de um limiar a partir do qual as produções orais com sotaque não exerçam dificuldades na inteligibilidade da fala e no processo comunicacional, ao mesmo tempo em que seja reconhecida como caracterizadora do indivíduo e merecedora de respeito por parte da sociedade.

Do aqui dito, vemos que o estudo dos sons das línguas não nativas invoca uma série de saberes, advindos de diferentes subáreas da Linguística e de áreas afins. Para Colantoni, Steele e Escudero (2015), por exemplo, pesquisadores e profissionais em Linguística, Psicologia, Educação, Fonoaudiologia e Ciência da Computação têm demonstrado cada vez mais interesse pelo estudo da fala em uma língua não nativa, dada a sua crescente contribuição para modelos de linguagem e cognição humana. Além disso, julgamos importante considerar, com especial atenção, o caráter aplicado da área. A partir desse viés aplicado, são investigadas diferentes estratégias de sala de aula para garantir que a fala em língua não nativa seja mais

inteligível, o que leva, por exemplo, à discussão sobre a não necessidade de uma mera “erradicação” do sotaque estrangeiro.

Nesse sentido, entender como os sons das línguas não nativas são percebidos, processados e produzidos é um passo importante para que possamos alimentar a área de ensino de línguas, sobretudo, é claro, o campo da oralidade. Ao entendermos o processamento de línguas não nativas, somos capazes, por exemplo, de (re) pensar práticas de ensino, bem como contribuir para o desenvolvimento de materiais voltados para o ensino de novas línguas. Consideramos que tais contribuições atingem um âmbito maior do que o ensino do domínio fonético-fonológico *per se*. De fato, as considerações sobre o ensino dos sons constituem parte importante das discussões sobre o próprio ensino de línguas adicionais no âmbito de todas as suas habilidades, especialmente se considerarmos que o componente referente aos sons da língua, por ser parte da língua em uso, não deve ser isolado dos demais componentes, tais como a sintaxe, a morfologia, a semântica e a pragmática.

## 2. O PAPEL DA ‘FALA’ NOS ESTUDOS DA LINGUAGEM

Para entendermos o deslocamento da ‘fala’ da margem da Linguística para, como alguns defendem, o lócus do desenvolvimento, devemos, primeiramente, entender a origem da distinção entre ‘língua’ e ‘fala’. Com o delineamento de um objeto de estudo, o *Curso de Linguística Geral* (CLG, Saussure 2006 [1916]) instaura a Linguística como ciência independente. Para Saussure, a linguagem pode ser tomada como o produto da soma da ‘língua/*langue*’ e da ‘fala/*parole*’, sendo a primeira definida como “o conjunto de convenções ao qual um corpo social recorre ao fazer uso da faculdade da linguagem” e como “o produto social proveniente de tal faculdade”

(Saussure 2006: 17). Já a ‘fala’, por sua vez, é definida, na gênese da Linguística, como “o lado executivo da linguagem, que é sempre individual” (2006: 21). Dessa forma, a linguagem possuiria um lado social e um lado individual, perpetuando-se, então, a separação entre o que é, para Saussure, essencial, a ‘língua’, e o que é “acessório”, a ‘fala’ (2006: 16). À luz da dicotomia ‘língua-fala’, a distinção entre Fonética e Fonologia começa a ser construída.

Muito embora o CLG já defina Fonética como uma ciência histórica, que “analisa acontecimentos, transformações e se move no tempo” (Saussure 2006: 43), é Trubetzkoy (1969 [1939]), à luz do Estruturalismo do Círculo Linguístico de Praga, que define a Fonética, como comumente ainda tomada hoje, como o estudo dos sons da ‘fala’ em termos saussureanos. Aqui, solidifica-se a separação entre Fonética, que investigaria os sons propriamente ditos das línguas, por exemplo, como eles são produzidos e percebidos, e Fonologia, que trabalharia com a organização desses sons em sistemas linguísticos.

Contudo, muito embora essa separação entre Fonética e Fonologia, entre ‘fala’ e ‘língua’, seja tradicionalmente tomada como categórica, sobretudo por ter sido perpetuada por modelos amplamente aceitos e adotados (e.g., Bloomfield 1961 [1933]; Chomsky 1957; Chomsky & Halle 1968), para outros autores (e.g., Ohala 1983; Browman & Goldstein 1992; Bybee, 2001; Albano 2001; 2020; Pierrehumbert 2003; Munson; Edwards & Beckmann 2005), a ‘fala’ possui um papel central no desenvolvimento da linguagem e, assim, as áreas da Fonética e da Fonologia deveriam ser unificadas, representando os dois lados, um micro e outro macro, de uma mesma moeda (Browman & Goldstein 1993). Essa proposta de integração da ‘fala’ à língua ancora-se na crescente evidência levantada por estudos experimentais, desenvolvidos a partir da década de 1980, acerca da importância do detalhe fonético, por exemplo, no processamento, no acesso e na aquisição da linguagem (Brescancini &

Gomes 2014). Em outras palavras, há evidências empíricas de que a ‘língua’, em termos saussureanos, é também dirigida pela ‘fala’, que tem sido desmarginalizada.

Além das mudanças paradigmáticas relacionadas ao *status* da ‘fala’ dentro dos estudos linguísticos, mesmo aqueles que ainda mantêm a distinção categórica entre ‘fala’ e ‘língua’ têm percebido que os dados de sons da fala, sobretudo em relação aos sons não nativos, são importantes fontes empíricas para discutirmos todas as áreas da Linguística. Estudos têm evidenciado que o desenvolvimento dos sons de línguas não nativas, por exemplo, é capaz de alterar a forma como bilíngues adultos produzem e percebem a L1, tanto a curto (e.g., Chang 2019) quanto a longo prazo (e.g., Schmid 2002), estando esses bilíngues imersos em uma comunidade de L2 dominante (Kupske 2016) ou não (Schereschewsky; Alves & Kupske 2017). Tais estudos têm implicações para a linguística teórica como um todo, ao tensionar a natureza, a organização e a maleabilidade de informações linguísticas no cérebro. Assim, a área do desenvolvimento dos sons não nativos vem questionando, por exemplo, a própria definição de ‘língua’.

Ademais, como afirma Alves (2008), os dados de línguas não nativas são capazes de dar suporte ou até mesmo refutar os postulados estabelecidos por modelos de análise linguística. São dados de sons não nativos, por exemplo, que fazem com que Alves (2008) problematize e proponha reformulações a um modelo de formalização linguística. Como podemos perceber, a pesquisa sobre os sons não nativos deve ser entendida como um componente fundamental para a pesquisa em linguagem. Em outras palavras, os estudos com a fala não nativa não somente se alicerçam teoricamente a partir dos postulados da Linguística Geral, mas, também, contribuem com insumos empíricos e teóricos para que se possam repensar pressupostos bastante arraigados na área.

### 3. QUESTÕES GERAIS DE METODOLOGIA DE PESQUISA

Muito embora os métodos mudem em função das bases teóricas adotadas e dos objetos e perguntas de cada pesquisa, nesta seção, buscamos apresentar, de maneira geral, alguns dos principais passos metodológicos na condução e na análise de experimentos da área, que são relevantes para todos os capítulos deste volume. Novamente, destacamos que, em cada capítulo, os leitores encontrarão seções específicas de metodologia referentes ao fenômeno em foco. Por esse motivo, aconselhamos a leitura das considerações a seguir, independentemente do fenômeno com o qual o futuro pesquisador trabalhará.

#### 3.1 OBJETIVOS, PERGUNTAS E HIPÓTESES DE PESQUISA

A palavra-chave para a formulação inicial de uma pesquisa é ‘leitura’. Toda pesquisa busca levantar respostas, por meio de procedimentos racionais e sistemáticos, a um problema relacionado à grande área de investigação. Dessa forma, a pesquisa é sempre dirigida por um problema de pesquisa claro, isto é, por uma ‘situação’ não resolvida ou completamente compreendida dentro de nosso domínio de atuação. Nesse sentido, a leitura é essencial. Muito embora problemas de pesquisa possam ser suscitados de nossas próprias experiências, muitos deles já podem ter sido (pelo menos parcialmente) respondidos por estudos anteriores. No caso da investigação dos sons não nativos, os problemas de pesquisa estão relacionados, em linhas gerais, à produção e à percepção da fala: por que alguns sons da L2 são mais difíceis de serem aprendidos? Por que alguns imigrantes apresentam alterações em suas L1s e outros não? Como testar os efeitos da instrução explícita de pronúncia?

Com a formulação ou identificação de um problema de pesquisa, é preciso recortá-lo em função da realidade do investigador. Por exemplo, um projeto de trabalho de conclusão de curso (TCC), com um ano de duração média, e um projeto de tese de doutorado, com quatro anos de duração média, não podem ter as mesmas pretensões de pesquisa, mesmo que se debruçam sobre o mesmo problema de pesquisa, pelo simples fato de possuírem extensões de tempo distintas para a condução do estudo. Assim, devemos pensar de que forma podemos contribuir para a solução ou compreensão do problema em função de nosso contexto (e.g., duração da pesquisa, equipamentos disponíveis, quantidade de participantes de pesquisa disponíveis, etc.). Nesse momento, então, devemos formular as perguntas de pesquisa, indicando os objetivos gerais e específicos, bem como as hipóteses de pesquisa, quando for o caso. Estudos analíticos (ou explicativos) necessitam de hipóteses, isto é, respostas especulativas e provisórias para as nossas perguntas de pesquisa, que serão verificadas ao final do projeto. A seção final de um relatório de pesquisa (TCC, artigo, dissertação, tese, etc.) deve conter as respostas para as perguntas do estudo; a informação de quais hipóteses foram rejeitadas, parcialmente confirmadas ou totalmente confirmadas; bem como informar se os objetivos, geral e específicos, foram alcançados.

É crucial que as perguntas, os objetivos e as hipóteses estejam alinhados entre si, convergindo para o escopo do estudo a ser conduzido. É essencial também que as perguntas, objetivos e hipóteses sejam formulados antes da condução do estudo e que não sejam alterados durante ou após a coleta e análise dos dados, pois a definição de perguntas e hipóteses *post-hoc*<sup>1</sup>, além de antiéti-

---

1 Prática conhecida como HARKing (*Hypothesis After Results are Known*), que faz alusão a *hacking*.

ca, minimamente distorce e potencialmente infla os resultados e as discussões, gerando um relatório de pesquisa distante da realidade. Perguntas, objetivos e hipóteses *post-hoc* fazem parte das más condutas que têm levado grandes áreas de estudo, como, recentemente, a psicologia, a uma ‘crise de replicabilidade’, justamente por gerarem resultados não replicáveis e, conseqüentemente, não generalizáveis (e.g., Anvari & Lakens 2018; Diener & Biswas-Diener 2020). Como procedimento antídoto a essa prática, o movimento da ‘ciência aberta’ tem motivado e valorizado o ‘pré-registro’ de pesquisas, isto é, o cadastro público, em páginas eletrônicas específicas para esse fim<sup>2</sup>, das perguntas, objetivos e hipóteses de um estudo, bem como dos passos metodológicos a serem empregados na análise, antes que ele seja conduzido, justamente para se evitar a alteração desses elementos tão importantes ao longo da coleta e análise, de acordo com os resultados prévios.

É por isso que a fase de elaboração das perguntas, objetivos e hipóteses da pesquisa é de suma importância, e deve ser baseada, como dito no início desta seção, em leitura abundante para se chegar a um conhecimento próximo ao exaustivo do estado da arte da área a ser pesquisada.

---

2 O *Open Science Framework* (<https://osf.io/>) é atualmente o mais frequentemente utilizado para esse fim.

### 3.2. ÉTICA

Com o projeto delineado, os pesquisadores devem submetê-lo a um Comitê de Ética em Pesquisa (CEP) competente. Além de norteados por uma metodologia bem desenhada, os pesquisadores devem obedecer a normas de conduta ética, para que possam proteger seus participantes, as instituições envolvidas no projeto e, é claro, a si mesmos.

A publicação da Resolução N° 466/2012<sup>3</sup>, aprovada pelo Conselho Nacional de Saúde (CNS) em dezembro de 2012, apresenta as diretrizes a serem consideradas por todos os projetos de pesquisa que envolvam seres humanos. Essa resolução baliza o funcionamento da Comissão Nacional de Ética em Pesquisa (CONEP), bem como de todos os Comitês de Ética em Pesquisa (CEP) a ela vinculados, na tentativa de assegurar os direitos e os deveres dos participantes da investigação. Por exemplo, desde 2012, é obrigatório que todos os participantes da pesquisa, ou seus representantes legais, recebam esclarecimentos sobre todos os procedimentos a serem conduzidos durante as etapas da pesquisa, bem como sobre todos os possíveis riscos e benefícios oriundos de sua inclusão no experimento. É importante mencionar a fundamental contribuição prestada pela posterior Resolução CNS/MS 510/2016, que regulamenta questões de ética em pesquisa dentro das áreas das ciências humanas, dentro da qual se encontram nossos estudos. A resolução em questão, fruto de diversas solicitações históricas de pesquisadores das ciências humanas, especifica algumas questões pontuais referentes a estudos como os realizados no âmbito da Linguística, de modo que sua leitura, além da Resolução CNS 466/2012, consti-

---

3 A Resolução N° 510, de 07 de abril de 2016, trata especificamente das ciências humanas e sociais.

tui etapa fundamental para o entendimento dos direitos e deveres éticos de participantes e pesquisadores.

Conforme a Resolução CNS 466/2012, a fase relativa à coleta de dados apenas poderá ser iniciada após a aprovação do projeto por um CEP. Contudo, para terem os seus projetos analisados por um CEP competente, os pesquisadores devem cadastrar suas pesquisas, conforme Norma Operacional CNS/MS nº 001/2013, na Plataforma BRASIL<sup>4</sup>. Apenas projetos com a documentação completa e devidamente alimentados na plataforma serão conduzidos para a apreciação de um CEP local. Assim, com o projeto terminado, a primeira etapa é fazer o seu cadastro na plataforma. Geralmente, as instituições contam com páginas eletrônicas para os seus CEPs e, neles, os leitores poderão encontrar mais detalhes, bem como dicas para a apreciação dos projetos. De qualquer modo, informações detalhadas podem também ser encontradas na Plataforma Brasil. No Quadro 1, apresentamos os documentos atualmente solicitados para as pesquisas que investigam sons não nativos nas propostas apresentadas neste volume.

#### QUADRO 1

Documentos solicitados para inserção na Plataforma Brasil

DOCUMENTO	DESCRIÇÃO
Projeto de Pesquisa Completo	Deve ser anexado à Plataforma. Na atual configuração, o proponente da pesquisa deverá 'copiar', na plataforma, as informações/seções do projeto. A apresentação de um cronograma e de um orçamento é obrigatória.

4 <http://conselho.saude.gov.br/plataforma-brasil-conep?view=default>. Nesse mesmo site, os leitores encontrarão o Manual do Pesquisador, contendo todas as orientações para a inclusão do projeto na Plataforma.

## INTRODUÇÃO A PESQUISAS DE SONS NÃO NATIVOS

Folha ou Formulário de Informações Básicas do Projeto	Documento de controle interno do CEP. O proponente deve fazer o preenchimento e alimentá-lo na plataforma.
Curriculum Lattes	O proponente deve apresentar os currículos de todos os pesquisadores responsáveis pela pesquisa.
Declaração de Participação no Projeto	O proponente deve apresentar a Declaração de Participação no Projeto de todos os pesquisadores.
Folha de Rosto	Esse documento, na atual configuração do sistema, é gerado pela própria Plataforma Brasil. Ele deverá ser assinado e carimbado tanto pelo pesquisador responsável quanto pelo responsável pela instituição de vínculo do pesquisador. O arquivo assinado será inserido na plataforma.
Termo de Consentimento Livre e Esclarecido (TCLE)	Os dados do Termo ou de um Pedido com Justificativa para Dispensa de TCLE serão copiados na plataforma. O documento deve ser construído conforme recomendações da Resolução 466/12.

Fonte: Os Autores.

No Quadro 1, listamos os itens básicos a serem apresentados ao CEP. Todavia, caso a pesquisa envolva a utilização futura de imagens e depoimentos dos participantes, um Termo de Autorização de Uso de Imagem e Depoimentos deve ser confeccionado. Caso a coleta de dados seja realizada dentro de uma instituição específica, uma Autorização para Coleta de Dados deve ser assinada e carimbada pelo responsável pelo local, preferencialmente em uma página contendo o cabeçalho com logotipo da instituição. Além disso, caso a pesquisa seja realizada com menores de idade, além da autorização dos pais de tais participantes, a ser realizada através do TCLE tradicional, tais participantes menores também deverão expressar seu consentimento, a partir de um Termo de Assentimento Livre e Esclarecido (TALE). Caso os leitores tenham dúvidas em relação aos

documentos necessários, sugerimos que entrem em contato com o CEP de suas instituições.

É importante destacarmos que documentos redigidos em língua estrangeira, comuns em pesquisas que investigam os sons não nativos, demandam a apresentação de uma versão traduzida para o português. Em relação à utilização de grupos controle, uma prática comum nos estudos em desenvolvimento de L2, indicamos que, sempre que possível, os proponentes do projeto também prevejam, mesmo que após todas as coletas de dados, algum tipo de trabalho instrucional com esses grupos, para que seus membros não sejam prejudicados.

Para finalizar esta seção, ressaltamos que estudos com participação de pesquisadores internacionais ou que considerem a coleta de dados em países estrangeiros devem levar em consideração as resoluções de ética do país em questão. Por exemplo, no Brasil, não podemos oferecer nenhum tipo de pagamento ou gratificação aos nossos participantes, sendo que, em diversos outros países, essa prática é comum. Assim, projetos interinstitucionais internacionais não devem ferir a legislação vigente de nenhum dos países envolvidos. A depender da legislação do país em questão, o projeto deverá passar pela apreciação ética local.

### **3.3. ESTUDO PILOTO**

Sempre que possível, é importante incluir, nos passos metodológicos, uma coleta piloto e uma análise dos dados piloto. Trata-se de uma coleta em pequena escala a fim de identificar limitações ou falhas nos instrumentos de coleta de dados e/ou nas decisões dos passos metodológicos, dando oportunidade de corrigir essas falhas antes da coleta de dados “oficial”. É comum que, na fase de

planejamento, não se percebiam algumas limitações ou problemas com os instrumentos ou com algumas decisões metodológicas, e que venham à tona apenas no momento da coleta de dados ou da análise propriamente dita. Alguns exemplos podem ser: uma tarefa para os falantes de L2 demasiadamente difícil ou demasiadamente fácil, uma tarefa que acaba não suscitando/eliciando os elementos necessários para a análise, uma tarefa muito longa ou muito cansativa para os participantes, uma falha técnica de configuração ou de operacionalização de equipamentos (e.g., gravadores, microfones, placas de som, alto-falantes, fones de ouvido, *software*, etc.), instruções que não ficaram claras para os participantes, etc.

Imagine coletar algum tipo de dado de 30 aprendizes de L2 (pesquisadores experientes sabem que nem sempre é fácil conseguir 30 alunos voluntários para colaborar com uma pesquisa), e só depois de todos os dados coletados, ao iniciar a análise, o pesquisador percebe algum erro ou limitação na maneira como os dados foram coletados, de modo que a única solução seria coletar tudo novamente. Trata-se de uma situação desesperadora, pois o pesquisador teria que conseguir outros 30 participantes para refazer a coleta de dados, primeiramente pelo constrangimento com os primeiros 30 participantes ao solicitar que cedam o dobro do tempo previsto para a coleta, e segundo porque seus dados teriam agora a influência de já conhecerem, pelo menos parcialmente, o instrumento de coleta, podendo gerar 'efeito de treinamento' nos dados coletados.

Sendo assim, para se evitar uma situação como essa, é aconselhável primeiramente realizar uma coleta com apenas dois ou três participantes de pesquisa para verificar se todos os instrumentos funcionam como planejado, e se geram dados como esperados, que sirvam para a análise proposta. Após esse estudo piloto, caso nenhum ajuste seja feito nos instrumentos ou nos procedimentos de

coleta, os dados do piloto podem ser adicionados aos “oficiais”. Caso qualquer mínimo ajuste seja feito após o estudo piloto, o que costuma ser o caso, esses dados devem ser descartados e apenas os dados “oficiais”, coletados posteriormente, devem ser utilizados para a análise propriamente dita. Por isso é importante coletar os dados piloto com poucos participantes, para não “desperdiçar” a valiosa colaboração de participantes de pesquisa. Por fim, cabe mencionar que a etapa piloto, incluindo o número de participantes recrutados nessa etapa, deve ser prevista, também, na documentação de solicitação de apreciação ética enviada ao Comitê de Ética em Pesquisa.

### 3.4. PREPARAÇÃO E ANÁLISE DOS DADOS

Independentemente da natureza da análise que será conduzida, quantitativa ou qualitativa, de produção ou de percepção, de um estudo de caso ou com grupos experimental e controle, os dados coletados deverão ser organizados para posterior análise. O recomendado é que sejam organizados em uma planilha, com cada variável em uma coluna e cada linha contendo os dados de um caso/uma observação. A tabela a seguir ilustra possíveis colunas e linhas iniciais de uma planilha com dados oriundos de uma análise acústica. Perceba que cada coluna tem uma das variáveis, tanto as independentes/previsoras<sup>5</sup> (idade, grupo, vogal) como as dependentes/de resposta (duração e formantes); e cada linha contém uma obser-

---

5 Variáveis ‘dependentes’ ou ‘de resposta’ são as medições do fenômeno que está sendo observado/avaliado, como as notas de uma prova, as respostas a um teste de percepção, a nota atribuída por um painel de juízes acerca do grau de inteligibilidade de um aprendiz, ou, como no exemplo acima de análise acústica, os parâmetros acústicos. Variáveis ‘independentes’ ou ‘previsoras’ são aquelas que potencialmente afetam ou interferem a dependente/resposta, como nível de proficiência, idade, ser do grupo controle ou do grupo experimental, L1, etc.

vação ou caso, que aqui se refere a cada instância de vogal analisada acusticamente, por isso há repetição do mesmo participante em algumas linhas. Exemplos de outras variáveis previsoras que poderiam entrar nessa planilha são nível de proficiência dos participantes, tempo de estudo da L2, palavra da qual a vogal foi retirada, a repetição da palavra (caso cada uma tenha sido apresentada mais de uma vez), faixa etária e/ou idade dos participantes, contexto de aprendizagem da L2, região dialetal da L1, etc. Exemplos de outras variáveis de resposta são  $f_0$ ,  $F_3$ , intensidade, formantes normalizados (um em cada coluna), etc.

**TABELA 01: Exemplo de tabulação de dados de análise acústica**

participante	idade	grupo	teste	vogal	duração	f1	f2	etc...
Part1	16	controle	pre	i	92	301	1921	
Part1	16	controle	pos1	i	93	294	1912	
Part1	16	controle	pos2	i	89	312	1965	
Part2	22	experimental	pre	i	101	320	1901	
Part2	22	experimental	pos1	i	99	332	1988	
etc...								

Fonte: Os Autores.

A próxima tabela apresenta um exemplo de tabulação para dados oriundos de um julgamento realizado por um painel de juízes. As escolhas dadas aos juízes para avaliarem o nível de proficiência percebido, o grau de inteligibilidade e o grau de sotaque estrangeiro foram codificadas em números ao serem inseridas na planilha.

É bom que essa codificação reflita a expectativa em relação ao desempenho dos aprendizes para facilitar a interpretação dos dados, em que o aumento no valor de nível de proficiência e de inteligibilidade, por exemplo, represente um nível mais alto de proficiência e um maior grau de inteligibilidade.

**TABELA 02: Exemplo de tabulação de dados de avaliação por painel de juízes**

participante	idade	grupo	teste	juiz	nível	inteligibilidade	sota-que	etc...
Part1	16	controle	pre	A	1,2	2,3	3	
Part1	16	controle	pos1	A	1,3	2,5	2,9	
Part1	16	controle	pos2	A	1,4	2,7	2,7	
Part2	17	experimental	pre	A	3	5,1	1,2	
Part2	17	experimental	pos1	A	2,9	5,3	1	
etc...								

Fonte: Os Autores.

Note que nos exemplos das Tabelas 01 e 02 não há nenhum acento, espaço ou caractere especial (como símbolos fonéticos) nas células, isso prevendo a possibilidade de posterior exportação dessa planilha para um programa de análise estatística, já que a análise quantitativa é o próximo passo. Caso a análise seja de fato realizada em um *software* diferente do editor de planilhas, é indicado não configurar aspectos visuais da planilha (como cores, fontes e tamanhos) e é importante não mesclar células.

### 3.4 ANÁLISE INFERENCIAL DOS DADOS

É possível realizar uma análise meramente qualitativa de dados numéricos, apresentando, por exemplo, médias e comentando os tamanhos das diferenças entre elas, juntamente com a apresentação de gráficos. Contudo, raramente uma análise apenas qualitativa dos dados numéricos é suficiente para fazer inferências confiáveis. Por exemplo, suponhamos que um grupo experimental apresente um aumento no nível de inteligibilidade conforme julgamento de um painel de juízes de 0,4 pontos, em média, e que o grupo controle apresente um aumento de 0,2 pontos. Esse resultado parece estar de acordo com a hipótese de que o grupo que recebeu determinada instrução explícita teria um aumento no nível de inteligibilidade maior do que o do grupo controle. No entanto, como podemos saber que essa diferença se deu por efeito da instrução e não ao acaso? Como saber se outro grupo controle não poderia ter um aumento de 0,4 simplesmente por continuar estudando a L2? Para responder a essas perguntas, é necessário conduzir algum tipo de análise inferencial dos dados.

Outro sinal claro dessa necessidade é que absolutamente todos os capítulos que compõem esta obra citam, em determinado momento, a necessidade de se realizarem análises estatísticas dos dados, mas também comentam que o espaço destinado a cada capítulo não comportaria uma explicação detalhada das análises recomendadas. Semelhantemente, o espaço e o objetivo deste capítulo introdutório não permitem explicações detalhadas de análises estatísticas. Mesmo assim, nos sentimos compelidos a deixar registradas algumas observações sobre limitações das práticas mais recorrentes, bem como sugestões de opções mais elaboradas, reconhecendo que este trecho será proveitoso apenas para aqueles já iniciados no tema. Caso você não tenha nenhum conhecimento

de estatística inferencial, pode ignorar os próximos parágrafos e ir direto para o último parágrafo desta seção; porém, primeiramente, deixamos registradas algumas sugestões de materiais sobre estatística específicos para linguistas, que podem ser utilizados tanto para começar como para aprofundar os estudos em análise quantitativa de dados: os manuais de Larson-Hall (2015) e de Loerts, Lowie e Seton (2020) como introduções; os de Gries (2013), Levshina (2015), Winter (2019) e Garcia (2021) para aprofundamento; além dos materiais disponibilizados gratuitamente on-line por Lima Jr, Garcia e Angele (2020), Oushiro (2017), Sonderegger, Wagner e Torreira (2018), Garcia (2019), e Godoy (2019).

A estratégia inferencial mais comum para se fazer comparações entre grupos ou entre tarefas distintas em um mesmo grupo se dá por meio de um teste-*t* ou de uma *análise de variância* (ANOVA), quando há mais de dois grupos (ou duas tarefas) a serem comparados, com posterior testes pareados *post-hoc*, como *Tukey* ou *Scheffe*. Entretanto, há diversas críticas aos testes de hipótese, ou testes de significância, e, principalmente, à ênfase ao *valor-p* (Wagenmakers 2007, Nuzzo 2014, Halsey 2015) para as tomadas de decisão sobre causalidade, dadas (i) a decisão categórica que o *valor-p* impõe, (ii) a aleatoriedade do 0,05 como valor limite para essa decisão, (iii) a possibilidade de se manipular os dados ou realizar múltiplos testes a fim de se alcançar um *valor-p* abaixo de 0,05, e (iv) a existência de estudos com *valor-p* abaixo de 0,05, mas com baixo poder ou tamanho de efeito. Além disso, o *valor-p* apresenta apenas a probabilidade dos dados diante de uma hipótese nula, mas não é capaz de informar sobre a probabilidade do efeito. Sendo assim, propomos três condutas alternativas, apresentadas em ordem crescente de complexidade, a fim de tornar a análise quantitativa dos dados mais robusta, confiável, replicável, e com mais possibilidade de trazer à tona inferências sobre causalidade.

A primeira sugestão é que, no caso da utilização de um teste de hipótese, como teste-t ou ANOVA, além do valor-p, do valor da estatística (t para teste-t e F para ANOVA) e dos graus de liberdade, o pesquisador também reporte o *intervalo de confiança* (95% *confidence interval*) e o *tamanho do efeito* (*effect size* – como *d* de Cohen para testes-t e *eta-quadrado* para ANOVAs). Esse procedimento acrescenta mais informações sobre os dados, trazendo um pouco de gradiência e de dúvida com o intervalo de confiança, e foco no efeito com o tamanho do efeito. Idealmente, o pesquisador ainda poderia informar o poder do(s) seu(s) teste(s).

A segunda sugestão, superior à primeira, seria a utilização de modelos de regressão mais completos (e.g., linear, logístico, multinomial, etc., preferencialmente com efeitos mistos). Modelos de regressão são superiores a testes de significância porque eles podem receber diversas variáveis predictoras ao mesmo tempo, levando todas em consideração, inclusive possíveis interações entre elas. Sendo assim, esses modelos refletem melhor a natureza do aprendiz de L2 e de seu sistema linguístico em desenvolvimento, que é composta por diversos agentes que interagem entre si (i.e., é multifatorial). Além disso, os resultados desses modelos, apesar de também apresentarem valores-p, naturalmente enfatizam os tamanhos do efeito e exploram possibilidades de explicações de causalidade.

A terceira sugestão, superior às demais, mas que exige mais estudo e uma mudança na maneira de se raciocinar sobre inferência e causalidade em estatística, sendo, na verdade, um raciocínio mais intuitivo, é a utilização de análises bayesianas. A inferência bayesiana leva em consideração não apenas os dados coletados, mas também o conhecimento prévio do pesquisador acerca do tema (a probabilidade *a priori*). Além disso, na estatística frequentista, como são chamados os testes e modelos men-

cionados nos parágrafos anteriores, busca-se saber a probabilidade de se encontrar dados pelo menos tão extremos como os coletados diante de uma hipótese nula. Isto é, calcula-se a probabilidade dos dados diante de um parâmetro, sendo o parâmetro inverso ao parâmetro real da pesquisa, ou seja, uma hipótese nula. Já na estatística bayesiana, calcula-se a probabilidade do parâmetro diante dos dados coletados e do conhecimento prévio do pesquisador sobre o fenômeno.

Para fazer esse cálculo e se chegar a uma distribuição de probabilidades do(s) parâmetro(s) (probabilidade *a posteriori*), utiliza-se o teorema de Bayes, daí o nome 'estatística/inferência bayesiana'. Note que o resultado, diferentemente do de análises frequentistas, não é um valor categórico, como o valor-p ou um coeficiente, mas uma distribuição de probabilidades, trazendo gradiente e espaço para dúvidas sobre o resultado. Vale ressaltar também que a inferência bayesiana é muito mais antiga do que a frequentista, proposta no século XVII; porém, sua aplicação prática e generalizada não era possível até poucas décadas atrás por falta de poder computacional para os cálculos. Para aqueles já familiarizados com testes estatísticos e modelos de regressão, sugerimos, como passo inicial em direção a uma análise bayesiana, o manual de Kruschke (2015) ou de McElreath (2020).

Realizada a análise inferencial, o passo seguinte é interpretar os dados à luz da Teoria. A observação dos dados e a ponte entre os aspectos empíricos e teóricos garantirá o sucesso do trabalho, pois de nada adianta uma análise estatística bem feita se o pesquisador não souber discutir os resultados evidenciados pela análise inferencial. O último passo, então, é a redação do relatório de pesquisa. O relatório pode ser um artigo em um periódico, um capítulo de livro, um TCC, uma dissertação ou uma tese. Grande parte do sucesso do trabalho científico diz respeito a como o pesquisador faz

a divulgação do seu trabalho. A escrita científica demanda passos pré-definidos. Tente ser o mais claro possível, de modo a detalhar seus objetivos e suas hipóteses, de modo a revisitá-los na seção de discussão dos resultados, à luz de uma teoria de base. A escrita clara de artigos científicos é parte fundamental do trabalho de pesquisa, e exige bastante afino e prática de qualquer pesquisador.

## **CONSIDERAÇÕES FINAIS**

No presente capítulo, acreditamos ter mostrado não somente a grande complexidade do que consideramos ser uma área interdisciplinar de estudos, mas também a sua pertinência, sobretudo em um mundo em que a maioria dos habitantes são bilíngues ou multilíngues.

Frente a um objeto de investigação tão complexo como o da fala multilíngue, exige-se do pesquisador uma série de saberes multidisciplinares, articulados a partir do método científico. Dessa forma, em um segundo momento deste capítulo, descrevemos, de forma mais pormenorizada, alguns dos passos básicos para a condução de um estudo empírico na área dos sons não nativos. Cabe mencionar que esses passos básicos serão complementados a cada um dos capítulos que seguem, a partir das especificidades de cada uma das questões de pesquisa tratadas de forma mais detalhada ao longo do livro.

Julgamos também necessário mencionar que, ainda que o pesquisador que lida com os sons não nativos possa vir a ter suas preferências por um ou outro tema de pesquisa a ser tratado nos capítulos seguintes, dada a complexidade do processo de desenvolvimento linguístico, exige-se do futuro pesquisador um entendimento global sobre a maioria das questões a serem retratadas nos capítulos a seguir. Por exemplo, aqueles que vierem a se interes-

sar por questões de treinamento perceptual precisarão, necessariamente, entender sobre as teorias perceptuais e o processo de produção da fala, além de considerar estudos como o de instrução e consciência fonológica, que são complementares aos de treinamento. Mais do que isso, ao se pensar no caráter aplicado do tema, dado que a inteligibilidade da fala em L2 caracteriza o grande objetivo do profissional de ensino, é preciso também saber discutir tais conceitos. Em outras palavras, ainda que apresentados com ênfases específicas, todos os temas dos capítulos seguintes mostram-se fortemente interconectados, de modo que, na prática, o pesquisador da área acabe lidando com todas as questões. A leitura de todos os capítulos contribuirá, nesse sentido, para o entendimento da área como um todo, dado que cada um deles acaba por assumir um caráter complementar a outro.

Com a presente obra, não somente esperamos que o leitor se mostre fascinado pelo objeto de estudos referente aos sons das línguas não nativas, mas, também, desejamos fornecer subsídios para que os desafios a serem enfrentados pelo futuro pesquisador possam ser transpostos. Como em todo o fazer científico, os desafios não são poucos. Entretanto, é a partir de tais desafios que nos sentimos instigados a aprender e a pesquisar ainda mais, em meio ao ciclo incessante que é a busca do conhecimento.

## REFERÊNCIAS

ALBANO, E. C. *O gesto e suas bordas: esboço de fonologia acústico-articulatória do português brasileiro*. Campinas, SP: Mercado de Letras, 2001.

ALBANO, E. C. *O gesto auditivo: fonologia como pragmática*. São Paulo: Cortez, 2020.

ALVES, U. K. *A aquisição das sequências finais de obstruintes do inglês (L2) por falantes do Sul do Brasil: análise via Teoria da Otimidade*. Tese (Doutorado em Letras). Porto Alegre: Pontifícia Universidade Católica do Rio Grande do Sul, 2008.

ANVARI, F.; LAKENS, D. The replicability crisis and public trust in psychological science. *Comprehensive Results in Social Psychology*, v. 3, n. 3, p. 266-286, 2018.

BLOOMFIELD, L. *Language*. New York: Holt, Rinehart and Winston, 1961.

BRESCANCINI, C. R.; GOMES, C. A. Apresentação: Fonética versus Fonologia? *Letras de Hoje*, v. 49, n. 1, p. 5-10, 2014. DOI: 10.15448/1984-7726.2014.1.17283.

BROWMAN, C. P.; GOLDSTEIN, L. Articulatory Phonology - an overview. *Haskins Laboratories Status Report on Speech Research*, SR-111/112, p. 23-42, 1992.

BROWMAN, C. P.; GOLDSTEIN, L. Dynamics Articulatory Phonology. *Haskins Laboratories Status Report on Speech Research*, SR-113, p. 51-62, 1993.

BYBEE, J. *Phonology and Language Use*. Cambridge: Cambridge Univ. Press, 2001.

CHANG, C. B. Phonetic Drift. In: SCHMID, M. S.; KÖPKE, B. (Eds.). *The Oxford Handbook of Language Attrition*. Oxford: Oxford University Press, 2019, p. 191-203.

CHOMSKY, N. *Syntactic structures*. The Hague: Mouton & Co., 1957.

CHOMSKY, N.; HALLE, M. *The sound pattern of English*. New York: Harper & Row, 1968.

COLANTONI, L.; STEELE, J.; ESCUDERO, P. *Second language speech: theory and practice*. Cambridge: Cambridge University Press, 2015.

DIENER, E.; BISWAS-DIENER, R. The replication crisis in psychology. In: BISWAS-DIENER, R.; DIENER, E. (Eds). *Noba textbook series: Psychology*. DEF publishers, 2020. Disponível em <http://noba.to/q4cvydeh>. Acesso em 09 de abril de 2021.

GARCIA, G. D. *Introduction to data analysis using R*, 2019. Disponível em [https://guilhermegarcia.github.io/rWorkshop/garcia\\_rWorkshop\\_complete.html](https://guilhermegarcia.github.io/rWorkshop/garcia_rWorkshop_complete.html). Acesso em 09 de abril de 2021.

GARCIA, G. D. *Data Visualization and Analysis in Second Language Research*. New York: Routledge, 2021.

GODOY, M. C. *Introdução aos modelos lineares mistos para os estudos da linguagem*. PsyArXiv, 2019. <https://doi.org/10.17605/OSF.IO/9T8UR>

GRIES, S. Th. *Statistics for linguistics with R: A practical introduction*. Berlin: Walter de Gruyter, 2013.

HALSEY, Lewis G. et al. The fickle P value generates irreproducible results. *Nature methods*, v. 12, n. 3, p. 179-185, 2015.

KUPSKE, F. F. *Imigração, Atrito e Complexidade: A Produção das Oclusivas Surdas Iniciais do Inglês e do Português por Sul-Brasileiros Residentes em Londres*. Tese (Doutorado em Letras). Porto Alegre: Universidade Federal do Rio Grande do Sul, 2016.

KRUSCHKE, J. K. *Doing Bayesian data analysis: a tutorial with R, JAGS, and Stan*, 2a edição. Elsevier, 2015.

LARSON-HALL, J. *A guide to doing statistics in second language research using SPSS and R*. Routledge, 2015.

LEVSHINA, N. *How to do linguistics with R: Data exploration and statistical analysis*. Amsterdam: John Benjamins Publishing Company, 2015.

## INTRODUÇÃO A PESQUISAS DE SONS NÃO NATIVOS

LIMA JR, R. M.; GARCIA, G. D. *Diferentes análises estatísticas podem levar a conclusões categoricamente distintas*. Disponível em: <http://osf.io/ndwbh>. Acesso em 06 de abril de 2021.

LIMA JR, R. M.; GARCIA, G. D.; ANGELE, B. *Introdução a modelos de regressão para linguistas no R*. 2020. Disponível em: <https://guilhermegarcia.github.io/rling.html>. Acesso em 06 de abril de 2021.

LOERTS, H.; LOWIE, W.; SETON, B. *Essential Statistics for Applied Linguistics: Using R Or JASP*. Amsterdam: Macmillan International, Red Globe Press, 2020.

MCELREATH, R. *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press, 2020.

MUNSON, B.; EDWARDS, J.; BECKMAN, M. E. Phonological knowledge in typical and atypical speech-sound development. *Topics in Language Disorders*, v. 25, p. 190-206, 2005.

NUZZO, R. Scientific method: statistical errors. *Nature News*, v. 506, n. 7487, p. 150, 2014.

OHALA, J. J. The origin of sound patterns in vocal tract constraints. In: OHALA, J. J.; MACNEILAGE, P. F. (Eds.). *The production of speech*. Springer, 1983, p. 189-216.

OUSHIRO, L. *Introdução à Estatística para Linguistas*, v.1.0.1 (dez/2017), 2017. Disponível em DOI <http://rpubs.com/oushiro/iel>.

PIERREHUMBERT, J. Probabilistic Phonology: Discrimination and Robustness. In: BOD, R.; HAY, J.; JANNEDY, S. (Eds.). *Probability Theory in Linguistics*. Cambridge, MA: The MIT Press, 2003. p. 177-228.

SAUSSURE, F. *Curso de Linguística Geral*. São Paulo: Cultrix, 2006.

SCHERESCHEWSKY, L. C.; ALVES, U. K.; KUPSKE, F. F. First language attrition: the effects of English (L2) on Brazilian Portuguese VOT patterns in an L1-dominant environment. *Letrônica*, v. 10, n. 2, p. 700-716, 2017. DOI: 10.15448/1984-4301.2017.2.26365.

## INVESTIGANDO OS SONS DE LÍNGUAS NÃO NATIVAS

SCHMID, M. S. *First Language Attrition, Use and Maintenance: The case of German Jews in anglophone countries*. Amsterdam/Philadelphia: John Benjamins, 2002.

SONDEREGGER, M.; WAGNER, M.; TORREIRA, F. *Quantitative Methods for Linguistic Data*. v. 1.0 (out/2018), 2018. Disponível em <http://people.linguistics.mcgill.ca/~morgan/book/>.

TRUBETZKOY, N. *Principles of Phonology*. Berkeley/ Los Angeles: University of California Press, 1969.

WAGENMAKERS, Eric-Jan. A practical solution to the pervasive problems of p values. *Psychonomic bulletin & review*, v. 14, n. 5, p. 779-804, 2007.

WINTER, B. *Statistics for linguists: An introduction using R*. Routledge, 201.

## **ORGANIZAÇÃO**

Felipe Flores Kupske

Ubiratã Kickhöfel Alves

Ronaldo Lima Jr.

## **REVISÃO**

Letícia Pereyron

## **CAPA E PROJETO GRÁFICO**

Estúdio Guayabo

### **Dados Internacionais de Catalogação na Publicação (CIP) (Câmara Brasileira do Livro, SP, Brasil)**

---

Investigando os sons de línguas não nativas [livro eletrônico] :  
uma introdução / Felipe Flores Kupske, Ubiratã Kickhöfel Alves,  
Ronaldo Lima Jr. (org.). -- Campinas, SP : Editora da Abralin,  
2021. -- (Linguística em ação)  
PDF

Vários autores.

Bibliografia.

ISBN 978-85-68990-11-7

1. Aquisição de linguagem 2. Fala 3. Fonética 4. Fonologia  
5. Linguagem e línguas 6. Linguagem e línguas - Estudo e ensino  
7. Linguística I. Kupske, Felipe Flores. II. Alves, Ubiratã Kickhöfel.  
III. Lima Júnior, Ronaldo. IV. Série.

21-81238

CDD-410

---

### **Índices para catálogo sistemático:**

1. Linguística 410

Eliete Marques da Silva - Bibliotecária - CRB-8/9380

DOI 10.25189/9788568990117