



PROCEDIMENTOS ESTATÍSTICOS APLICADOS À FALA DO INGLÊS COMO LÍNGUA ESTRANGEIRA

Maria Guadalupe Dourado Rabello¹

Leônidas José da Silva Junior²

RESUMO:

Em virtude de dificuldades enfrentadas em se estruturar, quantificar, analisar, interpretar e reportar resultados de pesquisas em língua estrangeira (L2) de um ponto de vista analítico-sistemático, o presente artigo tem como objetivo mostrar como a matemática aplicada (via procedimentos estatísticos) pode realizar uma interface com a fonética e a fonologia de L2 a partir da análise da produção/não-produção da fricativa interdental não-vozeada /θ/ do inglês por alunos brasileiros de diferentes graus de escolaridade e, desta forma, potencializar as tomadas de decisão quanto à feitura de inferências por parte do pesquisador acerca do fenômeno. Na Metodologia, realizamos um teste de Qui-quadrado tomando o valor proporcional das ocorrências de /θ/ e o grau de escolaridade (ensino fundamental, médio e superior) dos alunos. Nossos resultados apontaram que há uma correlação entre a (não) produção e o grau de escolaridade dos participantes. Com as observações realizadas a partir da estatística de testes, bem como, dadas as limitações deste estudo, podemos concluir que o fator escolaridade foi preponderante para a realização do referido segmento no inglês como L2.

PALAVRAS-CHAVE: Procedimentos estatísticos; Fala em língua estrangeira; Interdisciplinaridade.

ABSTRACT

Due to difficulties faced in structuring, quantifying, analyzing, interpreting and reporting results in foreign language (L2) research from an analytical-systematic point of view, this paper aims to demonstrate how applied mathematics (via statistical procedures) can interface with L2 phonetics and phonology based on the analysis of English interdental fricative /θ/ (non) production by Brazilian speakers from distinct school levels to enhance researcher's decision making regarding the phenomenon. As for the Methodology, we performed a Chi-squared test taking the proportional values of the occurrences of /θ/ controlled for school level (elementary and high school, and higher education) of the students. Our results pointed out that there is a correlation between (non) production and the school level of the participants. With the observations made from the test statistics, as well as, given the limitations of this study, we conclude that school level was determinant for the performance of the mentioned segment in English as L2.

KEYWORDS: Statistical procedures; Foreign language speech; Interdisciplinarity.

1 INTRODUÇÃO

Com base em relatos de docentes, tanto na educação básica, como na educação superior, é comum nos depararmos com dificuldades mais acentuadas no ensino de Estatística e suas subáreas como: descrição, probabilidade e interpretação de dados; ferramentas essenciais para formalização de protocolos de pesquisas em qualquer área da ciência. Em se tratando das ciências humanas e levando-se em conta que os profissionais que atuam nessas áreas têm um certo distanciamento das ciências exatas por diversos fatores, tais como: resistência em aprender conceitos matemáticos remanescentes do período escolar, dedicação à

¹ UNICAP. Email: <guadelupedr@gmail.com>.

² UEPB. Email: <leonidas.silvajr@gmail.com>.



sua área de estudo, etc. (cf. FAYOL, 2012), em um dado momento, esses profissionais precisam fazer uso de métodos estatísticos para que suas pesquisas sejam capazes de gerar inferências acerca do objeto pesquisado.

Segundo Oliveira & Cordani (2016), os professores de matemática não têm tanta familiaridade com a Estatística, tendo em vista que, apenas nas últimas três décadas, as licenciaturas em matemática apresentaram a Estatística em suas grades curriculares (e muitas vezes como optativa). Assim conceitos como aleatoriedade, incerteza e variabilidade nem sempre são enfatizados e discutidos ao longo do conteúdo correspondente.

Os autores ainda expõem que os livros didáticos apresentam uma abordagem instrumental dos conceitos de probabilidade e quando o fazem, geralmente reportam-se à análise combinatória (ferramenta para cálculo), impedindo uma discussão mais ampla de análise de dados e da importância da probabilidade nas análises estatísticas como, por exemplo, calcular a probabilidade da ocorrência de um determinado fenômeno linguístico ocorrer em uma dada comunidade de fala e como interpretá-lo.

Nas áreas humanísticas em que os dados quantitativos desempenham um papel relevante às inferências, como para a Psicologia Social, à Sociometria, a aplicação de métodos estatísticos tradicionais (desenho experimental, amostragem, estimativa, teste de hipóteses), ou mesmo heurísticos (agrupamento e escalonamento de uma quantidade incerta de eventos) são de suma importância. No que tange à Linguística, Sankoff (2001) propõe que é a partir da Teoria da Variação, onde as questões sociais tornam-se mais intimamente ligadas às questões fonético-fonológicas (variação de sotaque e outros aspectos da fala) e às questões morfosintáticas, que um protocolo com base analítico-estatística foi desenvolvido e amplamente adotado.

A Linguística (ainda) é uma das únicas entre as disciplinas científicas, que um número significativo de seus pesquisadores não exige e nem faz uso de metodologias estatísticas e, desse modo, seus resultados podem ser limitados por critérios estatísticos de validação. Sankoff (2001) acrescenta que, os linguistas tradicionalmente concordavam que a estrutura gramatical de uma língua consistia, em grande parte, de entidades ou categorias discretas, cujas relações e restrições de co-ocorrência eram de natureza qualitativa e compartilhadas por todos os falantes da comunidade de fala. Essas estruturas podem então ser deduzidas analisando e comparando enunciados suscitados ou intuídos por qualquer falante nativo da língua (por exemplo, linguistas que servem como sua própria fonte de dados), sem necessidade de qualquer desenho experimental e/ou protocolo formal de base estatística.

Os modelos formais da teoria gramatical têm estruturas discretas de natureza algébrica, algorítmica e/ou lógica. Tais estruturas geralmente envolvem conjuntos de dois ou mais componentes alternantes, como sinônimos, paráfrases ou alofones, que o pesquisador pode determinar estar executando funções linguísticas idênticas ou semelhantes. Ao permitir um certo grau de aleatoriedade na escolha entre esses suplentes, os formalismos gramaticais são convertidos em modelos probabilísticos de desempenho linguístico suscetíveis ao estudo estatístico (CEDERGREN e SANKOFF, 1974).

Assim, o presente artigo tem como objetivo mostrar como a matemática aplicada, por meio da estatística, pode realizar uma interface com uma pesquisa fonético-fonológica de língua estrangeira (L2) e, desse modo, potencializar as tomadas de decisão quanto a feitura de inferências por parte do pesquisador acerca de um determinado fenômeno.

Justificamos o presente artigo em virtude das dificuldades enfrentadas por um relativo número de estudiosos e pesquisadores da área de fala em L2 em estruturar, quantificar, analisar, interpretar e reportar os dados de suas pesquisas. Na Metodologia, analisamos a produção/não-produção do som da fricativa interdental não-vozeada /θ/ a partir de dados orais extraídos da leitura de um pequeno texto da pesquisa de Silva Jr. (2009). Verificamos a



correlação entre as ocorrências de (não) produção por grau de escolaridade (ensino fundamental, ensino médio e ensino superior) do segmento-alvo a partir da técnica estatística de Qui-quadrado.

Nossos resultados apontaram que alunos do ensino superior produzem de forma expressiva o som interdental quando falam inglês e que, com isso, é possível verificar que há uma correlação entre as variáveis. Em contrapartida, os alunos da educação básica (ensino fundamental e médio) apresentaram baixo índice de acertos e não apresentaram diferenças significativas quanto à produção entre eles. Com essas observações preliminares e dada as limitações técnicas oriundas deste estudo, podemos concluir que o fator escolaridade foi preponderante para a realização do referido segmento no inglês como L2.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção, mostraremos as relações científicas estabelecidas entre as áreas de Linguística e Estatística apontando protocolos para o desenvolvimento de uma pesquisa confiável. Mostraremos também como a fonologia de L2 e o surgimento de novas categorias fonéticas, influenciadas pela língua materna (L1) do falante, podem ser interpretadas e mensuradas por modelos de aprendizagem perceptual.

2.1 A RELAÇÃO ENTRE A LINGUÍSTICA E A ESTATÍSTICA

Ao longo da história moderna de sua ciência, a Linguística Formal (Fonética, Fonologia, Morfologia, Sintaxe e Semântica Formal) de modo não incomum, tem feito uso da Matemática, por meio da Estatística, do Cálculo e da Lógica, como uma ferramenta para investigar, postular e quantificar aspectos da língua a partir da determinação de modelagens além de refutar propostas teóricas em função de resultados empíricos. Segundo Hockett (1967), e Partee, ter Meulen e Wall (1993), a Linguística Formal entende linguagem como um sistema matemático baseado em regras.

De acordo com Scherre e Naro (2003), a Estatística entra nesse processo para revelar tendências e correlações inerentes na massa de dados linguísticos, e validá-las, dentro de um determinado grau de certeza. Assim, a Estatística ajuda o pesquisador-linguista a quantificar, resumir e manipular seus *corpora*. Portanto, é através da quantificação que o pesquisador confere suas hipóteses, realiza análises, interpreta os resultados obtidos e afirma o resultado da sua pesquisa. É comum o uso de tabelas e figuras para ilustrar melhor as descrições e estimativas do fenômeno em pauta, sendo possível através de um determinado processo, conferir se há na língua uma variação estável ou uma mudança em curso, por exemplo.

É apenas a partir do trabalho de William Labov, no final dos anos 1960 (LABOV, 1969), que, de fato, passa a haver uma formalização científica na montagem de desenhos experimentais e de protocolos estatísticos para investigar questões de interesse central da teoria linguística. Em seu trabalho, o autor examina, do ponto de vista da linguística gerativa, a posição da cópula do verbo de ligação <be> (ser/estar) na fala do inglês não-padrão afro-americano (*Non-standard Negro English* - NNE) e conclui que de fato, há o apagamento desta cópula. Suas inferências baseiam-se nos modelos sintáticos, com padrões semelhantes ao NNE, observados em outras línguas como húngaro, hebraico e crioulo-francês do Caribe.

Labov (1969) expõe à comunidade linguística a ideia de que duas ou mais articulações distintas de uma determinada forma fonológica podem ocorrer na mesma palavra ou afixo, nos mesmos contextos, sem afetar o significado referencial de um item lexical ou mesmo a função sintática de um determinado afixo ou partícula. É possível prever qual forma ocorrerá em um determinado momento no tempo a partir de um modelo probabilístico pelo qual os



efeitos do contexto linguístico e extralinguístico podem ser determinados com precisão; todavia, o resultado desta análise continua sendo apenas uma probabilidade. A escolha da forma sempre contém um componente do puro acaso, embora isso seja precisamente delimitado (SANKOFF, 2001).

Apresentamos no Quadro 1, um exemplo do trabalho de Labov (1969) em que se observa a distribuição de ocorrências do verbo copular <be> (ser/estar) do inglês falado. Este verbo ocorre como contração no inglês padrão (*Standard English* - SE) e apagamento no NNE.

FORMA REGULAR	SE	NNE	TRADUÇÃO
<i>John is a doctor</i>	<i>John's a doctor</i>	<i>John a doctor</i>	John é médico
<i>We are there</i>	<i>We're there</i>	<i>We there</i>	Estamos lá
<i>I am at home</i>	<i>I'm at home</i>	<i>I at home</i>	Estou em casa

Quadro 1: Variáveis linguísticas para a cópula “be” (em negrito).

Fonte: Adaptado de Labov (1969).

Para compreendermos os métodos que Labov utilizou, é necessário que tenhamos uma noção sobre métodos de investigação empírica. De acordo com Monteiro (2000), o variacionismo parte do pressuposto de que a heterogeneidade manifestada na fala pode ser analisada de forma que o pesquisador deve colher uma boa soma de dados em uma comunidade e que esses dados construirão o material que será submetido a análises estatísticas para a testagem da sua hipótese. O autor menciona ainda que a opção pela pesquisa empírica se liga ao fato de que, sendo a sociolinguística uma ciência social, ela depende da observação do comportamento dos seres humanos.

De acordo com Guy e Zilles (2007), toda pesquisa - dialetal, geográfica ou social - é inerentemente quantitativa, visto que essa metodologia inclui uso de tabelas e Figuras para apresentação de dados, teste de significância, confiabilidade e técnicas analítico-quantitativas. Os autores ainda mencionam que existem três fases importantes no curso de qualquer análise quantitativa, são elas:

- *Coleta de dados:* Aqui, devemos observar aspectos como amostra e confiabilidade em que em estudos de comunidade de fala, onde se usam mais de um pesquisador, há prática de se usar testes de confiabilidade entre os pesquisadores para assegurar que todos estão aplicando o mesmo critério de análise;
- *Redução e apresentação de dados:* As técnicas para redução de dados mais utilizadas provêm da área de Estatística e tais técnicas incluem medidas de tendências centrais como média, mediana e moda, além de medidas de dispersão, como, desvio-padrão e variância. Na apresentação dos dados, o uso de tabelas, Figuras ou mapas são de grande importância;
- *Interpretação e explicação de dados:* identificar e explicar fenômenos linguísticos e não somente produzir números, como por exemplo, medidas estatísticas para resumir os dados. Os fenômenos e sua natureza devem ser explicados através dos números. Os números mostram, de forma codificada em linguagem matemática, o que acontece com um determinado aspecto de variação fonética, por exemplo. Em outras palavras, os números representam um caminho que pode explicar o comportamento linguístico em estudo.

Guy e Zilles (2007) concluem que é através da realização de análises quantitativas que é possível realizar estudos da variação linguística em uma dada comunidade. Quando se fala em variação, esta deve ser compreendida como a alternância entre dois ou mais elementos linguísticos, que não podem ser adequadamente descritos e analisados em termos categóricos



ou estritamente qualitativos. Vejamos na próxima seção, a descrição e as condições de uso da técnica estatística que tomaremos para análise de nossos dados.

2.2 TÉCNICA ESTATÍSTICA EM DADOS SOCIOLINGÜÍSTICOS DA FALA: O TESTE DE QUI-QUADRADO (χ^2)

De acordo com Everitt e Hothorn (2006), a inferência é o processo de se tirar conclusões sobre uma dada população com base em mensurações ou observações feitas a partir de uma amostra de indivíduos desta população. Esse processo é central quando laçamos mão da estatística de testes.

Nesta seção, utilizamos o conjunto de dados descritos no clássico trabalho de Labov ([1972], 2008) para ilustrar a aplicação de um dos testes estatísticos mais comuns na Estatística inferencial: o teste de Qui-quadrado (*chi-squared test*). Sobre a técnica estatística de Qui-quadrado (χ^2), Guy e Zilles (2007) afirmam que essa distribuição é um procedimento não-paramétrico, útil para calcular a probabilidade de uma hipótese nula (H_0) ser verdadeira entre duas variáveis: a de linha x e a de coluna y em que ambas as variáveis são categóricas. Sua estatística é uma medida de divergência entre a distribuição dos dados observados e uma distribuição esperada desses dados em análise.

Assim como na pesquisa de Labov ([1972], 2008), o teste de Qui-quadrado foi implementado ao nosso conjunto de dados o qual apresentaremos nas seções posteriores. O trabalho laboviano que ilustra a utilização da técnica de Qui-quadrado foi realizado em três lojas de departamento na cidade de Nova York em que o autor toma, dentre outras, a variável diastrática (*classe social*) dos consumidores das lojas descritas: *Saks* (status financeiro superior), *Macy's* (status financeiro médio) e *S. Klein* (status financeiro mais baixo) para definir se haveria produção ou apagamento da aproximante retroflexa /ɹ/ em posição pós-vocálica final nas palavras <four^Rth floo^R> (quarto andar) por parte dos vendedores dessas lojas.

Labov realizou 68 entrevistas na *Saks*, 125 na *Macy's* e 71 na *S. Klein*, distribuídas entre 264 falantes. As Tabelas 1 e 2 apresentam os resultados da referida pesquisa. A Tabela 1, apresenta as ocorrências totais de produção e apagamento do /ɹ/, enquanto que a Tabela 2, apresenta a proporção (em %) da produção e apagamento do /ɹ/.

Ocorrências do /ɹ/	Lojas de departamento			TOTAL
	<i>S. Klein</i>	<i>Macy's</i>	<i>Saks</i>	
<i>Apagamento</i>	195	211	93	499
<i>Produção</i>	21	125	85	231
TOTAL	216	336	178	730

Tabela 1: Tabela de contingência 2x3 da distribuição da produção/apagamento do /ɹ/ por loja de departamentos.

Fonte: Adaptado de Labov ([1972], 2008).

Proporção do /ɹ/	Lojas de departamento		
	<i>S. Klein (%)</i>	<i>Macy's (%)</i>	<i>Saks (%)</i>
<i>Apagamento</i>	90.2	62.8	52.2
<i>Produção</i>	9.8	37.2	47.8

Tabela 2: Tabela de proporção da produção/apagamento do /ɹ/ por loja de departamentos em Nova Iorque

Fonte: Adaptado de Labov ([1972], 2008)

Com os valores apresentados na Tabela 3, observemos no Figura 1, a descrição dos dados de produção/apagamento do /ɹ/ da referida pesquisa de Labov.

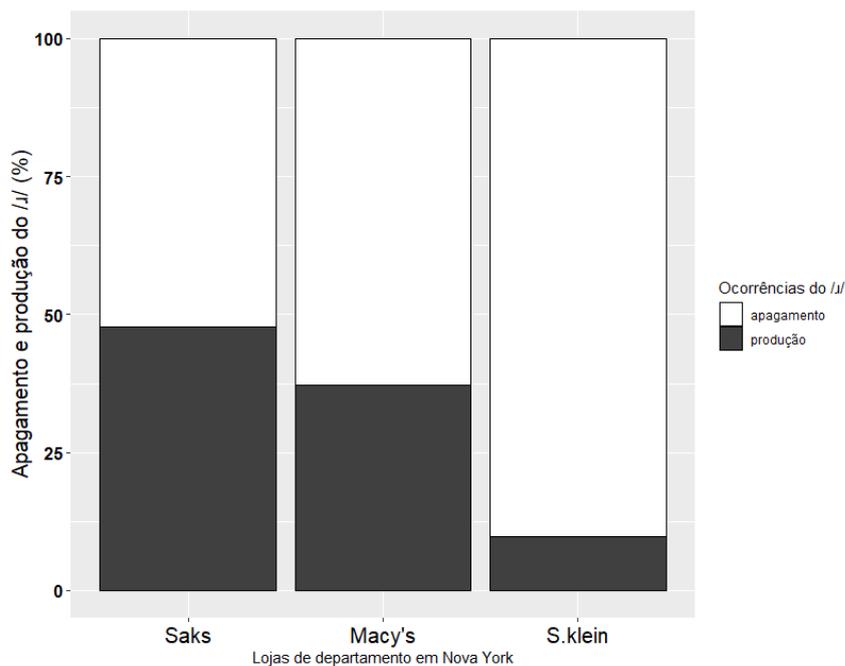


Figura 1: Apagamento e produção do /ɪ/ pós-vocálico.

Fonte: Adaptado de Labov ([1972], 2008).

Vejam na Tabela 3 os resultados dos testes de qui-quadrado a partir dos dados da Tabela 2:

LOJAS DE DEPARTAMENTO	VALOR χ^2	VALOR-P
<i>Saks; Macy's; S.Klein</i>	35,7	<0,01
<i>Saks; Macy's</i>	2,28	=0,13
<i>Saks; S.Klein</i>	35,3	<0,01
<i>Macy's; S.Klein</i>	21,0	<0,01

Tabela 3: Resultados dos testes de Qui-quadrado com as lojas de departamento.

Fonte: Adaptado de Labov ([1972], 2008).

No domínio da metalinguagem matemático-estatística, atentemos também para o fato de como devemos reportar resultados estatísticos em pesquisas. Uma vez que o pesquisador realizou as análises (cf. Tabela 3) e interpretações acerca dos dados, faz-se necessário que os resultados sejam reportados de modo sistemático em publicações, tais como: periódicos, anais de congressos, dentre outros meios. Oushiro (2017, p. 108-109) orienta que a notação convencional (a matemática) deve ser 'lida', por exemplo, com a primeira linha alfanumérica da Tabela 4: [$\chi^2 = 35,7$ (2), $p < 0,01$], i.e., "Qui-quadrado igual a 35,7, com dois graus de liberdade e p menor do que 0.01". Vejamos o que representa cada um dos parâmetros encontrados no resultado do χ^2 .

Como mencionado no início da seção, o teste de Qui-quadrado (35,7) compara os valores observados com os esperados com base na H_0 . Os graus de liberdade (GL) são computados da seguinte forma: $GL = (\text{número de linhas} - 1) * (\text{número de colunas} - 1)$. Quando tomamos a Tabela 1, temos o seguinte cálculo: $GL = (2 - 1) * (3 - 1) = 2$. Assim como nas usuais técnicas de Estatística inferencial, o teste de Qui-quadrado gera um valor de p como podemos observar no parágrafo anterior.



O valor de p (*p-value*), é o valor da probabilidade de se rejeitar ou aceitar a H_0 . No experimento de Labov acima ilustrado, p é igual probabilidade (p) de ocorrer produção ou apagamento do retroflexo. Na prática, a literatura estatística estabelece um valor de significância (*alfa*) de 5% por convenção (cf. BROWN, 1988; GRIES, 2009; TRIOLA, 2014; dentre outros).

- Se o valor de p for menor que *alfa*, ou seja, $p < 0,05$, mais forte é a evidência de rejeitar-se a H_0 . No experimento de Labov aqui ilustrado, existe uma variação significativa entre as lojas pesquisadas quanto à produção/apagamento do /ɹ/;
- Se o valor de p for maior do que *alfa* ($p > 0,05$). No experimento de Labov aqui ilustrado, mais forte é a evidência de aceitar a H_0 , i.e., não há diferenças estatisticamente significativas entre a produção ou apagamento do /ɹ/ nas três lojas em estudo.

A partir do trabalho de Labov, podemos inferir a respeito desse fenômeno linguístico. Uma dessas inferências é que, através da verificação do apagamento/manutenção do /ɹ/, é possível mapear probabilisticamente uma tendência de consumo pelo perfil socioeconômico dos clientes que frequentam as lojas pesquisadas: *Saks*, *Macy's* e *S. Klein*. Os clientes que frequentam a *Saks* e *Macy's*, têm perfis que não destoam estatisticamente. Já na loja mais popular (*S. Klein*) o apagamento do /ɹ/ é significativamente maior do que nas demais lojas. Na loja mais popular, há uma preocupação consideravelmente menor quanto a utilização do fenômeno de prestígio (com base o apagamento do /ɹ/). Confirma-se assim a hipótese de Labov, de que a realização do /ɹ/ seria determinada pelo ambiente socioeconômico em que o falante se encontra.

Podemos, dessa forma, observar como as análises estatísticas podem fazer com que tomemos decisões de situações cotidianas a partir de fenômenos linguísticos e este fato aponta para uma relação positiva entre a matemática, por meio da estatística, e estudos da linguagem, por meio de estudos como: a produção/apagamento do /ɹ/ pós-vocálico, ou mesmo, produções fonéticas de L2, etc.

No que se refere à variação fonético-fonológica de L2 e levando-se em conta o processo de variação apontado por Guy e Zilles (2007) anteriormente, Edwards e Zampini (2008) entendem que, a variação linguística que ocorre quando um falante estrangeiro produz eventos de fala na L2-alvo também configura-se como variação inter- (e intra-) sujeitos de comunidades de fala distintas. Para os autores, é debruçando-se sobre a variação linguística que a teoria sociolinguística teve maior impacto quando se é levada em conta aspectos fonético-fonológicos de L2.

Edwards e Zampini acrescentam que a variação é uma questão crítica para a área de Aquisição de L2, uma vez que qualquer modelo válido de aquisição de L2 deve abordar questões de como a variação na produção fonética se explica e, até que ponto, as transferências em função de processos interlinguísticos devem ser levadas em conta. Na próxima seção, abordaremos aspectos como transferências e formação de categorias fonéticas por parte dos falantes estrangeiros, bem como, alguns dos modelos de aquisição de L2 que procuram explicar o processo de variação e mudanças de tais fenômenos.

2.3 FORMAÇÃO DE CATEGORIAS FONÉTICAS EM L2

Desde o início dos anos 1980, a literatura sobre aquisição de L2 propõe que falantes de L2 normalmente produzem e percebem os sons de uma língua não-nativa diferentemente dos falantes nativos (Flege, 1980, dentre outros). Esses desvios fonéticos da pronúncia-alvo em função da L1 do falante podem ser atribuídos ao sistema fonológico da L1. Munro e Derwing (2008), por exemplo, apontam que os erros de produção de vogais- e consoantes-alvo são os



fatores que mais dominam a produção fonética da L2 e agem como uma das causas de maior impacto para não-inteligibilidade do falante estrangeiro. Jenkins (2000, 2008) assevera que a transferência de elementos fonéticos da L1 para a L2-alvo gera um impacto à percepção e à produção da fala não-nativas podendo ter consequências sociais (por exemplo, estereotipagem) e consequências linguístico-comunicativas com o comprometimento da inteligibilidade.

Na busca por uma sistematização da produção e percepção da fala de L2, os modelos de aprendizado estatístico (*L2 statistical learning models*) ou de percepção de L2 (*L2 perceptual models*) representam um papel importante no cenário fonético propondo modelagens de padrões estatísticos em que os aspectos orais são percebidos a partir de um sistema dinâmico e complexo de forma gradual. Em outras palavras, à medida que o falante/ouvinte recebem um *input* sonoro da L2 (estes *inputs* devendo apresentar boa qualidade das características fonéticas da língua-alvo), sua produção vai se aperfeiçoando gradualmente (cf. BATES e MacWHINNEY, 1989; BEST, 1995; FLEGE, 1995; KUHL e IVERSON, 1995; ESCUDERO, 2005; BEST e TYLER, 2007; GROSJEAN, 2008, 2013 para detalhamento dos modelos perceptuais de aprendizagem de L2). Grosso modo, é como se a cada vez que ouvimos e praticamos a pronúncia da L2, somos atraídos em direção a uma pronúncia mais fina.

Brown (2000) argumenta em favor de um modelo à luz da Fonologia gerativa com base na estrutura interna do fonema, pode oferecer alguns *insights*. A autora propõe que não são as propriedades fonéticas de L1 que fossilizam a habilidade perceptiva, mas a estrutura do sistema fonêmico. Portanto, as propriedades fonêmicas do sistema da L1 determinam como o sistema sonoro L2 será percebido. De acordo com a Brown (2000), durante a aquisição/aprendizado de sua L1, as crianças adquirem conhecimento das representações fonêmicas, bem como dos recursos que compõem essas representações. Usando a geometria de traços e hierarquias do sistema fonológico da L1, estes infantes podem explicar quais traços de um dado fonema da L2 serão percebidos e posteriormente relacionados a um fone específico da L1.

Como exemplo, a autora examina o caso de falantes ingleses aprendendo hindi. Ela explica que o inglês não subdivide o traço de coronalidade em um espaço de articulação mais refinado, como o falante nativo de hindi o faz, que distingue sons retroflexos de não-retroflexos dentro do espaço coronal. Portanto, a autora aponta que os falantes ingleses perceberão todos os sons coronais produzidos no espaço coronal como sendo o mesmo fonema. Brown acrescenta que essa percepção em função da geometria de traços seria o motivo pelo qual algumas L1s optam por substituir as fricativas interdentais não-vozeada e vozeada do inglês, respectivamente /θ, ð/, por consoantes oclusivas /t, d/ e/ou fricativas /s, z, f, v/. O exemplo das fricativas fornecido pelos estudos de Brown ilustra bem os nossos dados que serão apresentados nas próximas seções.

A formalização dos estudos que originaram boa parte dos modelos de aprendizado e percepção de L2 tem origem na teoria Interlíngua (*interlanguage*) com o trabalho de Selinker (1972). Esta teoria prevê que categorias fonéticas de L2 são percebidas gradualmente em ‘estágios’ distintos que não pertencem mais à L1 do falante, e (ainda) não pertencem à L2-alvo. Segundo o autor, em cada ‘estágio’, o falante produziria uma nova interlíngua onde o erro de pronúncia (erro fonético) seria uma tentativa de o aprendiz realizar um reparo de sua produção em relação à produção-alvo, i.e., surgiria uma nova categoria fonética.

Como exemplo desta tentativa de reparo, Silva Jr. (2009) apresenta aprendizes brasileiros de inglês de uma escola pública que foram expostos à produção oral da vogal alta anterior [i:] em palavras como <eat> (comer). Alguns alunos, em um dos estágios de produção da referida vogal, pronunciou-a, [ie], como sendo duas vogais. A vogal tônica [i:]



do inglês apresenta duração mais longa que a tônica [i] do PB (cf. MACHERPE, 1970), e, dessa forma, é possível que a produção das duas vogais tenha ocorrido para compensar a duração percebida da vogal da L2. A representação da tentativa de ajustes fonéticos que são previstos pelos modelos de aprendizado e percepção de L2 podem ser observados no Figura 2:

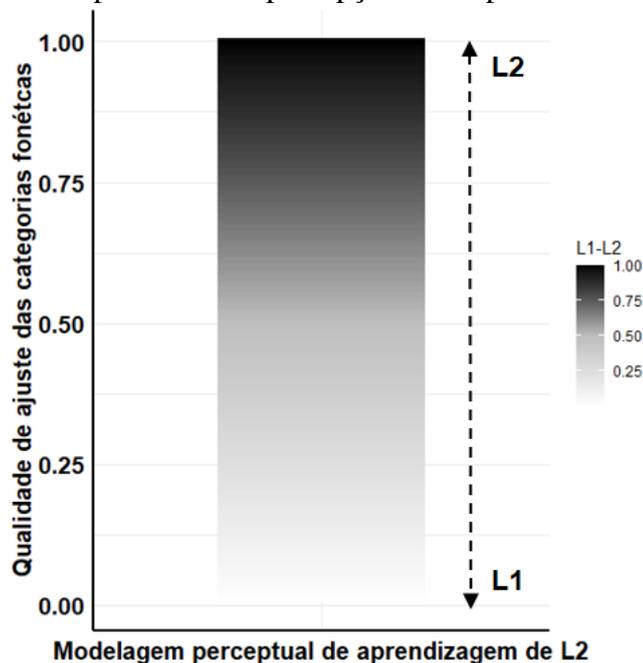


Figura 2: Representação dos modelos estatístico-perceptuais de aprendizado de categorias fonéticas de uma L2-alvo.

Fonte: Os autores.

Observando o Figura 2, perceberemos que a percepção-produção do falante/ouvinte não é estática, nem tampouco, categórica e se movimenta (com maior ou menor velocidade) em direção à L2-alvo, como ilustrado anteriormente na produção da vogal alta anterior [i:] em <eat> produzida como sendo duas vogais, [ie], em função do ajuste duracional da vogal do inglês. Com o *input* adequado, os modelos sugerem que, em algum momento, o falante estrangeiro será capaz de perceber as diferenças de duração vocálica e, dessa forma, realizar o reparo em um espaço cada vez mais próximo da L2-alvo.

Até aqui, vimos como a matemática aplicada (via estatística) está presente nos fenômenos relacionados à linguagem uma vez que, a escolha fonético-fonológica extraída de uma amostra de uma determinada comunidade de fala pode apontar tendências de hábitos de consumo, de comportamento, (dentre tantos outros) de uma determinada população. Ademais, vimos como surgem novas categorias fonéticas de L2 - tomando a L1 como marco - com base em modelos de percepção. Tais categorias são passíveis de mensuração e aplicação de modelos matemático-estatísticos para descrição e caracterização de sotaque estrangeiro, para averiguação de características semelhantes e distintas da L1 do falante, etc. Nas próximas seções, apresentaremos o mesmo procedimento estatístico realizado por Labov ([1972], 2008), todavia, com dados de inglês como L2 produzido por falantes brasileiros e, quais inferências a estatística é capaz de nos fornecer para a interpretação desses dados.

3 METODOLOGIA

A presente pesquisa é de natureza quantitativa e busca avaliar se há correlação entre o grau de escolaridade e a produção da fricativa interdental não-vozeada /θ/ do inglês/L2 por alunos brasileiros. Utilizamos dados extraídos a partir da leitura de um pequeno texto (cf.



SILVA JR., 2009) realizada por alunos brasileiros falantes de inglês como L2 estratificados pelo grau de escolaridade: (ensino fundamental (EF), ensino médio (EM) e ensino superior (ES)). A coleta de dados foi realizada em uma escola e universidade públicas na cidade de Recife-PE.

Extraímos para análise o seguinte excerto³: “(...) *and I was about to finish. It was the fourth and last part of the treatment and the therapy was no longer so bad at that point (...).*”

Realizamos uma análise perceptivo-auditiva para identificar se houve a produção (ou não) do segmento-alvo.

3.1 SUJEITOS, CORPUS E ANÁLISE ESTATÍSTICA

Escolhemos apenas a variável grau de escolaridade para manter a condição de igualdade de variáveis com o experimento de Labov apresentado na seção anterior. Em nossa pesquisa, verificamos se houve produção ou não-produção (que pode se caracterizar por apagamento, troca por outra categoria fonética e/ou qualquer outro fenômeno que não seja a produção-alvo) da fricativa interdental /θ/.

Como mencionado anteriormente, os dados foram obtidos a partir da leitura de um pequeno texto. Para nossa análise, extraímos a produção das palavras “*fourTH*” (quarta) e “*THerapy*” (terapia) com o intuito de verificar produção/apagamento da interdental não-vozeada /θ/. Este som do inglês é percebido e pronunciado com certa dificuldade por falantes brasileiros uma vez que ele não é presente no inventário fonológico do PB (cf. MASCHERPE, 1970; MOUREIRA, 2017; dentre outros para detalhamento da literatura que aborda este segmento). Um total de 120 ocorrências produzidas por 60 sujeitos foram contabilizadas (da leitura de cada sujeito foram extraídas duas ocorrências: <*fourTH*> e <*THerapy*>) conforme apresentação no Quadro 2.

PARTICIPANTES	GRAU DE ESCOLARIDADE	FAIXA ETÁRIA	PRODUÇÕES
20	Ensino Fundamental	11 a 14 anos	40
20	Ensino Médio	15 a 18 anos	40
20	Ensino Superior	> 18 anos	40

Quadro 2: Distribuição dos sujeitos pesquisados por faixa etária e grau de escolaridade.

Fonte: Os autores.

Uma vez contabilizadas as produções, realizamos os mesmos procedimentos do experimento de Labov. O tratamento estatístico de nossos dados, bem como, as figuras apresentadas ao longo do trabalho foram realizados em ambiente R (R CORE TEAM, 2020), disponível em: <<https://cran.r-project.org/>>.

Como apontado em seções anteriores, utilizamos o Teste de Qui-quadrado para análise estatística dos nossos dados a fim de verificar se há uma associação entre as variáveis de linha e coluna (EVERITT e HOTHORN, 2006; LOWIE e SETON, 2013; TRIOLA, 2014) que compõem uma tabela de contingência construída a partir dos dados da amostra.

Vejam na equação (Eq.) 1 a fórmula de Qui-quadrado, bem como, o cálculo dos valores esperados das ocorrências (produção e não-produção de /θ/) por grau de escolaridade.

Eq. 1

³ **Tradução do excerto extraído para análise:** “(...) e eu estava prestes a terminar. Foi a quarta e última parte do tratamento e a terapia já não era tão ruim àquela altura (...)”.



$$\chi^2 = \sum_{k=1}^N \frac{(O_k - E_k)^2}{E_k}, \rightarrow E = \frac{(\text{SOMA de linhas}) * (\text{SOMA de colunas})}{\text{Total da amostra}}$$

Onde:

- O_k = Frequência observada para cada amostra da tabela;
- E_k = Frequência esperada para cada amostra da tabela;
- L= Linhas;
- C = Colunas;
- k = cada uma das amostras;
- N = Total das amostras.

Um valor de 5% foi estabelecido para *alfa*. Se *alfa* < 5% ($p < 0,05$), o grau de escolaridade influencia na (não) produção do segmento-alvo.

- Se o valor de p for menor que *alfa*, ou seja, $p < 0,05$, mais forte é a evidência de rejeitar-se a H_0 , ou seja, o grau de escolaridade influencia na (não) produção da fricativa interdental não-vozeada /θ/;
- Se o valor de p for maior do que *alfa* ($p > 0,05$), mais forte é a evidência de se aceitar a H_0 , ou seja, o grau de escolaridade não influencia na (não) produção da fricativa interdental não-vozeada /θ/;

4 RESULTADOS E DISCUSSÃO

Vejamos os resultados apresentam-se distribuídos quanto à produção (linhas) e o grau de escolaridade dos falantes (colunas) em seus valores absolutos e proporcionais (Tabela 4 e 5 respectivamente)

Ocorrências de /θ/	Grau de escolaridade			TOTAL
	EF	EM	SUP	
<i>Não-produção</i>	30	34	11	75
<i>Produção</i>	10	6	29	45
TOTAL	40	40	40	120

Tabela 4: Tabela de contingência 2x3 de distribuição das produções/e não-produções do /θ/ por Grau de escolaridade.

Fonte: Os autores.

Proporção de /θ/	Grau de escolaridade		
	EF (%)	EM (%)	ES (%)
<i>Não-produção</i>	75	85	27.5
<i>Produção</i>	25	15	72.5

Tabela 5: Tabela de proporção para produção e não-produção do /θ/ por Grau de escolaridade.

Fonte: Os autores.

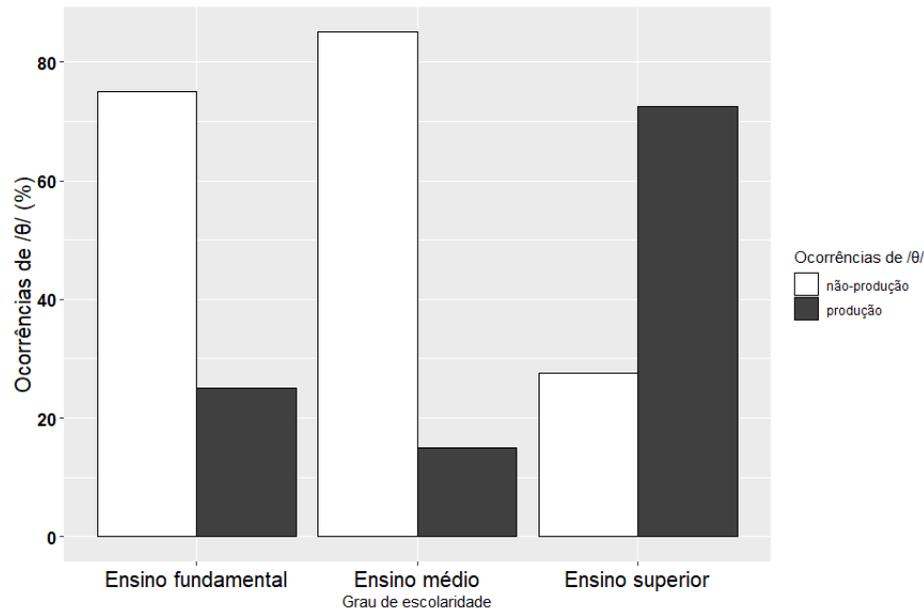


Figura 3: Proporção de ocorrências do /θ/ (%) por grau de escolaridade.

Fonte: Os autores.

Vejamos na Tabela 6, a descrição dos resultados estatísticos extraídos da técnica de Qui-quadrado aplicada sobre dos dados da Tabela 5:

GRAU DE ESCOLARIDADE	VALOR χ^2	P-VALOR
EF; EM; ES	32,21	<0,01
EF; EM	1,25	>0,26
EF; ES	18,06	<0,01
EM; ES	26,87	<0,01

Tabela 6: Resultados dos testes de Qui-quadrado (χ^2) e valor de p (P-VALOR) para produção e não-produção da fricativa interdental não-vozeada /θ/ predito pelo grau de escolaridade.

Fonte: Os autores.

A primeira linha alfanumérica da Tabela 6 [$\chi^2 = 32,21$ (2), $p < 0,01$] nos mostra os valores obtidos a partir da distribuição de Qui-quadrado levando-se em conta os três grupos do grau de escolaridade (EF, EM, ES).

Em Eq 2, realizamos a decomposição dos cálculos da primeira linha alfanumérica da Tabela 6. Os dados observados, que foram utilizados na Eq. 2, encontram-se na Tabela 4.

$$\chi^2 = \frac{(30 - 25)^2}{25} + \frac{(34 - 25)^2}{25} + \frac{(11 - 25)^2}{25} + \frac{(10 - 15)^2}{15} + \frac{(6 - 15)^2}{15} + \frac{(29 - 15)^2}{15},$$

$$\chi^2 = \frac{25}{25} + \frac{81}{25} + \frac{196}{25} + \frac{25}{15} + \frac{81}{15} + \frac{196}{15} = \frac{25 + 81 + 196}{25} + \frac{25 + 81 + 196}{15} = \frac{302}{25} + \frac{302}{15}$$

$$\chi^2 = 32,21$$

Eq. 2

Para calcular os demais valores presentes nas linhas alfanuméricas 2, 3 e 4 da Tabela 6, repita o procedimento desenvolvido em Eq. 6 com os valores observados e esperados atualizados.

Conforme apontado na Metodologia desta pesquisa, se p for menor que 0,05, rejeitamos a H_0 . O valor de Qui-quadrado é igual a 32,21. Considerando que $GL = 2$, este



valor torna o valor de p extremamente pequeno (localizado abaixo da curva Qui-quadrada) provindo da H_0 como mostra a Tabela 6 ($p < 0,01$). O valor de p é igual $2,3 * 10^{-8}$, ($p = 2,3 * 10^{-8} = 0,000000023$). Em outras palavras, com o valor de p nesta magnitude, rejeitamos a H_0 e aceitamos a hipótese alternativa (H_1), ou seja, o grau de escolaridade prediz um efeito sobre a (não) produção da fricativa-alvo. No Figura 3, observamos que os alunos do ES produzem a consoante-alvo estatisticamente mais do que os alunos da educação básica (EF e EM).

Tomando como base os resultados estatísticos descritos na Tabela 6 e no Figura 3, podemos realizar procedimentos inferenciais de ordem sociolinguística e de ordem fonético-fonológica quanto à aquisição de L2, tais como:

- Os alunos do ensino superior, produzem significativamente mais a fricativa interdental /θ/ do que os alunos do ensino fundamental [$\chi^2 = 18,06$ (1), $p < 0,01$]; e médio [$\chi^2 = 18,06$ (1), $p < 0,01$];
- Os discentes da educação básica (EF e EM) diferem de modo discreto estatisticamente, i.e., sem variação significativa quanto à produção/não-produção do segmento-alvo [$\chi^2 = 1,25$ (1), $p > 0,26$];
- Tomando este estudo como base (e considerando suas limitações), podemos afirmar que o grau de escolaridade tem expressiva influência sobre a produção do inglês como L2 no que tange ao segmento-alvo em específico. O teste de Qui-quadrado também revela que; entre os grupos de alunos que compõem a educação básica (EF e EM) não há diferenças;
- Quando o teste é aplicado grupo a grupo, a variação é significativa quando comparamos o grupo de universitários (ES) a cada um dos grupos de alunos da educação básica.

4.1 INFERÊNCIAS ACERCA DOS RESULTADOS DA PESQUISA

As dificuldades apresentadas pelos alunos da educação básica podem ser de ordem diversa; desde o não incentivo às práticas de pronúncia durante as aulas de L2 tradicionalmente apontada pela literatura nas escolas públicas de ensino regular (ALMEIDA FILHO, 2007; OLIVEIRA, 2009; RAMOS, 2009) a questões de ordem socioeconômica que possam vir a comprometer o acesso do discente a ferramentas tecnológicas, cursos de idiomas, viagens de intercâmbio, dentre outras situações que estimulam o processo comunicativo.

Salientamos que, no período da coleta dos dados aqui apresentados, o acesso à internet, a dispositivos móveis (*tablets*) além do uso de aplicativos para dispositivos móveis que contemplasse o ensino de pronúncia era bem mais limitado. Além do mais, variáveis como o nível de proficiência do falante (se o falante é básico, (intermediário ou avançado), tempo de exposição à L2-alvo (se o falante assiste programas em inglês ou mesmo se utiliza a internet em sites e/ou programas em inglês), dentre outras, não foram controladas no estudo.

Sobre a adoção de práticas de pronúncia nas escolas com discentes brasileiros falantes de inglês-L2 durante educação básica, os estudos acústico-experimentais de Alves e Silva Jr. (2015) e Alves (2020) apontam uma melhoria significativa na produção da fricativa interdental não-vozeada a partir da instrução explícita. Os autores utilizaram o protocolo de Ladefoged e Johnson (2011), que toma como base o segundo formante (F2) do segmento vocálico seguinte, para mensurar e determinar a distinção acústica entre a interdental /θ/ e a labiodental /f/. Nos estudos mencionados foram priorizados estímulos orais que promovessem a consciência fonológica do segmento-alvo. Esses trabalhos foram realizados com alunos do ensino fundamental de escolas públicas e particulares.



Quanto à percepção-produção da fricativa interdental não-vozeada /θ/, os trabalhos de Lombardi (2003) e Major (2008) apontam que questões ortográficas também funcionam como um gatilho para a troca de segmentos por influência da L1, pois como a ortografia, por si só, é uma representação abstrata, influências ortográficas na fala de L2 tendem a favorecer uma abordagem mais abstrata. Os autores citam que falantes de russo/L1 e inglês/L2 produzem [t] em vez de [θ]; falantes de japonês/L1 e inglês/L2 produzem [s] em vez de [θ]. Major (2008) acrescenta o grupo dos brasileiros à produção da referida fricativa apontando que falantes de PB/L1 e inglês/L2 alternam entre oclusiva [t] e as fricativas [f, s] em vez do alvo [θ].

Os resultados de Major (2008) para dados de falantes de PB/L1-inglês/L2 são consoantes à proposta da Teoria da Otimalidade (OT) em L2 (*L2 OT-based study*) proposta por Lombardi (2003). A autora postula que alguns falantes de uma dada L1 produzem substituições que têm uma gênese na própria L1 (por exemplo, [θ → s] por falantes do PB), mas outros falantes desta dada L1 produzem substituições que são guiadas por um princípio de marcação que descreve a tendência das línguas de favorecer à produção de oclusivas (por exemplo, [θ → t] por falantes do PB).

Aportada nos estudos de Lombardi (2003), Hancin-Bhatt (2008) realiza uma análise das substituições interdentais com base em restrições e hierarquias de formas marcadas/não-marcadas presente nas línguas a partir da OT de L2 e atribui possíveis explicações, com base fonologia de L2, para as trocas de [θ → t, f, s]. As considerações de Hancin-Bhatt apontam possíveis caminhos para os resultados estatísticos aqui descritos e nos traz possíveis causas afeitas à percepção dos participantes em nosso estudo quanto à realização/não-realização da interdental /θ/.

- As gramáticas contêm uma restrição de marcação que indica que as fricativas são mais marcadas do que as oclusivas;
- As restrições de marcação atuam contra a ocorrência de inventários de segmentos interdentais (neste caso, [θ]);
- As línguas variam em sua classificação quanto ao modo de *faithfulness* em relação às restrições de marcação. A restrição mais relevante para a substituição da interdental é o *modo de articulação*, que é definida pelos seguintes traços de *modo*: [oclusão] >> [continuidade] >> [estridência];
- Línguas que substituem /θ/ → [t] refletem tendências da Gramática Universal em direção a segmentos não marcados (oclusivas), enquanto que línguas que substituem /θ/ → [s] refletem a reclassificação de restrições para uma forma marcada, em que desta vez, o traço de continuidade ganha proeminência na organização fonotática.

O estudo fonético-experimental de Moureira (2017) sobre a produção da fricativa interdental não-vozeada /θ/ por falantes do PB/L1-inglês/L2 apresenta resultados consoantes aos postulados de Lombardi (2003) e Hancin-Bhatt (2008) no que tange às restrições e hierarquias. A partir da técnica estatística de Regressão Linear, Moureira (2017) modelou parâmetros acústicos como o *Centro de Gravidade* das consoantes-alvo e concluiu que: falantes que recebem instrução explícita tendem a moldar suas pronúncias gradualmente em direção ao segmento-alvo na sequência: [t → s/f → θ]. O coeficiente de determinação (R^2) explica 68% de variação ($R^2 = 0.68$) na mudança gradual [t → s → f → θ].

Vale a pena ressaltarmos que, mesmo para falantes nativos de inglês que apresentam a fricativa interdental não-vozeada em seu inventário fonológico, sua aquisição é tardia e ocorre por volta dos 4.5 aos 5 anos de idade (OWENS JR., 2012). O autor afirma que na ordem de aquisição e produção fonética de consoantes do inglês por falantes nativos, as nasais são



adquiridas e produzidas primeiramente, seguidas pelas plosivas, aproximantes, aproximantes laterais, fricativas e africadas (p. 304-305).

Owens Jr. (2012) acrescenta que, mesmo para os nativos, processos de substituição de consoantes na fala podem ser descritos de acordo com o *modo de articulação* do som alvo. Por exemplo, na fala do infante em fase de aquisição, as consoantes fricativas (e africadas) são passíveis de serem substituídas por oclusivas pela ordem de aquisição.

Dessa forma, podemos inferir que a aquisição da interdental não-vozeada não está apenas em função de sua presença/ausência do inventário fonológico da L2, mas também, trata-se segmento é suscetível a restrições e hierarquias baseadas na OT como apontam os estudos acima citados de Lombardi (2003) e Hancin-Bhatt (2008).

5 CONSIDERAÇÕES FINAIS

Neste trabalho, procuramos mostrar a importância da aplicabilidade de um procedimento matemático-estatístico-probabilístico (a técnica de Qui-quadrado) em pesquisas na área de fonética e fonologia de L2 a fim de realizar inferências e interpretações acerca dos resultados. Além disso, apresentamos como métodos estatísticos têm permeado de forma primordial e preenchendo lacunas nas áreas de conhecimento envolvendo as ciências humanas no intuito de protocolar, sistematizar e seguir um determinado desenho experimental. Mostramos também que, a partir da interface linguística-matemática é possível traçar perfis de uma dada comunidade de fala, bem como, realizar estimativas acerca do seu comportamento linguístico.

Assim como no experimento de Labov ([1972], 2008), realizamos um teste de Qui-quadrado em dados de produção/não-produção da fricativa interdental não-vozeada /θ/ por falantes brasileiros de inglês como L2 estratificando os dados pelo grau de escolaridade. Procuramos mostrar a utilização do teste de forma prática e ilustrativa com a intenção de contribuir na formação de pesquisadores que realizem análises quantitativas em pesquisas de caráter experimental. Dessa forma, procuramos quebrar alguns paradigmas quanto ao suposto “alto grau de dificuldade” que um determinado teste estatístico venha a nos imputar defendendo uma interdisciplinaridade entre matemática e linguística.

De algumas décadas para nossa atualidade, os *softwares* computacionais como o (ambiente) *R*, *Stata*, *Microsoft Excel*, *SPSS* e tantos outros têm realizado com eficiência a tarefa de calcular testes e outros procedimentos estatísticos. Atualmente, também é possível realiza-los a partir de aplicativos para dispositivos móveis, ou mesmo por manipuladores virtuais online o que torna o acesso e utilização de técnicas estatísticas mais amigável ao pesquisador.

Os resultados de nossa pesquisa, dadas as limitações mencionadas em seções anteriores, apontam uma correlação entre o grau de escolaridade e a (não)produção de /θ/. Essas considerações indicam que o estudante universitário procura buscar mais questões relacionadas à pronúncia da L2 por alguns motivos distintos como, o aumento do interesse pelo idioma em função da empregabilidade, a necessidade por causa da cobrança do meio acadêmico, viagens ao exterior em programas vinculados à sua instituição de ensino, dentre outros.

Reiteramos que as práticas de ensino de pronúncia de L2 na educação básica devem ser revisitadas; desde o investimento na formação continuada de docentes de L2 nas áreas de fonética e fonologia, a condições e materiais compatíveis com a realidade do aluno. Além disso, há a restrições hierárquicas de aquisição do som da L2-alvo baseadas na percepção e na produção do falante estrangeiro e, tais restrições, podem ser modeladas por procedimentos



estatísticos para uma tomada de decisão acertada por parte do docente-pesquisador ao interpretar seus resultados.

5.1 LIMITAÇÕES DA PESQUISA

Não foi possível realizar análises fonético-acústicas neste estudo em função de os dados de áudio elencados para análise apresentarem-se no formato MP3. Nesse formato de gravação, o áudio sofre cortes das altas (e baixas) frequências (acima dos 8000 Hz), justamente as que caracterizam as fricativas-alvo (e oclusivas alveolares (de 8000 a 9000 Hz). Além disso, o áudio apresenta perda de amplitude significativa em função da baixa taxa de quantização (4 bits em vez de 16 bits).

Gravações de áudio realizadas com as configurações supracitadas enviesariam a realização de análise acústica em consoantes fricativas (cf. Kent e Read, 1992; Shadle e Mair, 1996; Ladefoged, 2003; Ladefoged e Disner, 2012; Barbosa e Madureira, 2015, dentre outros para detalhamento sobre procedimentos técnicos de gravação em alta fidelidade, frequência, intensidade e manutenção de integridade dos parâmetros acústicos a serem extraídos das consoantes fricativas). Isto posto, consideramos que esta é uma das limitações da presente pesquisa.

Além das limitações de ordem técnica, fatores como: um número relativamente pequeno de amostras (apenas duas ocorrências por sujeito – cf. TRIOLA, 2014; BARBOSA, 2021) e as dificuldades da análise de oitiva em distinguir produções da consoante-alvo /θ/ (interdental) e da consoante /f/ (labiodental) em função dos espectros gerados por essas duas fricativas, também se constituem como limitações do presente estudo conforme aponta Barbosa, 2021. Barbosa corrobora o estudo de Shadle e Mair (1996) sobre as dificuldades da análise de oitiva. Segundo as autoras, os quatro momentos espectrais (centro de gravidade, desvio-padrão, assimetria e curtose) das consoantes /f, θ/, além da inclinação espectral em frequências acima da amplitude máxima (que varia com o esforço vocal), não mostraram diferenças significativas. Assim, as autoras concluem que, até o momento da pesquisa, não há um parâmetro acústico preciso que distinga este par de consoantes.

Todavia, devemos mencionar que, mesmo com limitações do ponto de vista técnico que comprometem análises acústico-experimentais dos dados, realizamos uma pesquisa com dados primários, ou seja, fontes primárias de informação (dados reais com produções orais de alunos em escola e universidade públicas que estavam em fase de aprendizagem no momento da coleta de dados). Segundo Brown (1988), a vantagem estatística de uma pesquisa com dados primários (alunos aprendendo a falar uma L2, por exemplo) em vez de uma pesquisa com dados secundários (livros ou outras fontes secundárias que debatem sobre alunos aprendendo a falar uma L2) reflete de modo mais realístico a comunidade em estudo mesmo que optemos por outras técnicas (perceptivo-auditiva nesta pesquisa). O autor acrescenta que:

A pesquisa estatística é tangível na medida em que se baseia na coleta e manipulação de dados reais. Um conjunto de dados pode assumir a forma de escores em testes, (...), o número de aprendizes de L2 que possuem certas características e assim por diante (BROWN, 1988, p. 4, [tradução nossa]⁴).

⁴ **Texto original:** *Statistical research is tangible in that it is based on the collection and manipulation of data from the real world. The set of data may take form of test scores, (...), the number of language learners who have certain characteristics and so on.*



REFERÊNCIAS

ALMEIDA FILHO, J. C. **Dimensões comunicativas no ensino de línguas**. 4. ed. Campinas: Pontes, 2007.

ALVES, A. Produção da fricativa interdental não vozeada do inglês /θ/ por aprendizes brasileiros através do desenvolvimento da consciência fonológica. In: HORA, D.; PEDROSA, J.; LUCENA, R. (Ed.). **Cuadernos de la Alfal**: estudos linguísticos: da aquisição ao uso, Santiago, v. 1, n. 12, p. 179-193, mai. 2020.

ALVES, A.; SILVA JR, L. Percepção-produção da fricativa interdental surda do inglês /θ/ tendo como suporte o desenvolvimento da consciência fonológica. **Leia Escola**, v. 17, 2017, p. 90-100. Disponível em: <<http://dx.doi.org/10.35572/rle.v17i1.856>>. Acesso em: 16 mai. 2021.

BARBOSA, P. **Considerações sobre as fricativas /θ, f/** [Mensagem pessoal]. Mensagem recebida por: leonidas.silvajr@gmail.com. 27 mai. 2021.

BARBOSA, P.; MADUREIRA, S. **Manual de Fonética Acústica Experimental**: aplicações a dados do português. São Paulo: Cortez, 2015.

BATES, E.; MacWHINNEY, B. Functionalism and the Competition Model. In: MacWHINNEY, B.; BATES, E. (Ed.). **The crosslinguistic study of sentence processing**. New York: Cambridge University Press, 2989. p. 3-76.

BEST, C. A direct realist view of crosslanguage speech perception. In: STRANGE, W. (Ed.). **Speech perception and linguistic experience**: theoretical and methodological issues. Timonium: York Press, 1995. p. 171-204.

BEST, C.; TYLER, M. Nonnative and second-language speech perception: Commonalities and complementarities. In: MUNRO, J.; BOHN, O-S (Ed). **Second language speech learning: The role of language experience in speech perception and production**. Amsterdam: John Benjamins, 2007. p. 13-34.

BROWN, C. The interrelation between speech perception and phonological acquisition from infant to adult. In: ARCHIBALD, J. (Ed.), **Second language acquisition and linguistic theory**. Oxford: Blackwell, 2000. p. 4-63.

BROWN, J. **Understanding Research in Second Language Learning**. Cambridge: Cambridge University Press, 1988.

CEDERGREN, H.; SANKOFF, D. Variable rules: performance as a statistical reflection of competence. **Language**, v. 50, p. 333-355, 1974. Disponível em: <<https://doi.org/10.2307/412441>>. Acesso em: 02 mai. 2020.

EDWARDS, J.; ZAMPINI, M. **Phonology and Second Language Acquisition**. Amsterdam: John Benjamins Publishing Co., 2008.



- ESCUADERO, P. **Linguistic perception and second language acquisition**: Explaining the attainment of optimal phonological categorization. Tese (Doutorado em Linguística). Netherlands Graduate School of Linguistics, Utrecht University, Utrecht, 2005. Disponível em: <
https://www.researchgate.net/publication/27690636_Linguistic_perception_and_second_language_acquisition_Explaining_the_attainment_of_optimal_phonological_categorization>. Acesso em: 18 set. 2020.
- EVERITT, B.; HOTHORN, T. **A Handbook of Statistical Analyses Using R**. New York: Chapman & Hall/, 2006.
- FAYOL, M. **Numeramento**: aquisição das competências matemáticas. Tradução de BAGNO, M. São Paulo: Parábola, 2012. Título original: L'acquisition du nombre.
- FLEGE, J. Second Language Speech Learning: Theory, findings, and problems. In: STRANGE, W. (Ed.). **Speech perception and linguistic experience**: issues in crosslanguage research. Timonium: York Press, 1995. p. 233-277.
- FLEGE, J. Phonetic approximation in second language acquisition. **Language Learning**, v. 30, p. 117-134, 1980. Disponível em: <<https://doi.org/10.1111/j.1467-1770.1980.tb00154.x>>. Acesso em: 04 abr. 2021.
- GRIES, S. **Statistics for Linguistics with R: A Practical Introduction**. Berlin: Mouton de Gruyter, 2009.
- GROSJEAN, F. Bilingualism: a short introduction. In: GROSJEAN, F.; LI, P. **The Psycholinguistics of Bilingualism**. Oxford: Wiley-Blackwell, 2013. p. 5-26.
- GROSJEAN, F. **Studying bilinguals**. Oxford: Oxford University Press, 2008.
- GUY, G; ZILLES, A. **Sociolinguística quantitativa**: instrumental de análise São Paulo: Parábola, 2007.
- HANCIN-BHATT, B. Second language phonology in optimality theory. In: EDWARDS, J.; ZAMPINI, M. (Ed.). **Phonology and Second Language Acquisition**. Amsterdam: John Benjamins Publishing, 2008. p. 117-146.
- HOCKETT, C. **Language, Mathematics and Linguistics**. The Hague: Mouton, 1967.
- JENKINS, J. **English as a Lingua Franca: Attitude and Identity**. Oxford: Oxford University Press, 2008.
- JENKINS, J. **The Phonology of English as an International Language**. Oxford: Oxford University Press, 2000.
- KENT, R.; READ, C. **The Acoustic Analysis of Speech**. New York: Cengage Learning, 1992.



- KUHL, P.; IVERSON, P. **Linguistic experience and the “perceptual magnet effect”**. In: STRANGE, W. (Ed.). **Speech perception and linguistic experience: issues in crosslanguage research**. Timonium: York Press, 1995. p. 121-154.
- LABOV, W. **Padrões Sociolinguísticos**. Tradução BAGNO, M. São Paulo: Parábola, 2008. Título original: Sociolinguistic Patterns.
- LABOV, W. Contraction, deletion, and inherent variability of the English copula. **Language**, v. 45, p. 715–762, 1969. Disponível em: <<https://doi.org/10.2307/412333>>. Acesso em: 02 mai. 2020.
- LADEFOHED, P. **Phonetic Data Analysis. An introduction to fieldwork and instrumental techniques**. Malden: Blackwell Publishing, 2003.
- LADEFOGED, P.; DISNER, S. **Vowels and Consonants**. 3 ed. West Sussex: Wiley-Blackwell, 2012.
- LADEFOGED, P.; JOHNSON, K. **A Course in Phonetics**. 6 ed. Boston: Wadsworth, 2011.
- LOMBARDI, L. Second language data and constraints on manner: Explaining substitutions for the English interdental. **Second Language Research**, v.19, p. 225–250, 2003. Disponível em <<https://doi.org/10.1177/026765830301900304>>. Acesso em: 08 mai 2021.
- LOWIE, W.; SETON, B. **Essential Statistics for Applied Linguistics**. Ney York: Palgrave Macmillan, 2013.
- MAJOR, R. Transfer in second language phonology: A review. In: EDWARDS, J.; ZAMPINI, M. (Ed.). **Phonology and Second Language Acquisition**. Amsterdam: John Benjamins Publishing, 2008. p. 63-84
- MASCHERPE, M. **Análise comparativa dos sistemas fonológicos do inglês e do português: classificação dos erros prováveis com sugestões para uma aplicação pedagógica**. São Paulo: Empresa Gráfica das Revistas dos Tribunais, 1970.
- MONTEIRO, J. **Para compreender Labov**. Petrópolis: Editora Vozes, 2000.
- MOUREIRA, I. A. **Produção da fricativa interdental surda /θ/ por aprendizes de inglês como segunda língua**. 38f. Monografia de Graduação (Licenciatura em Letras-ínglês) – Universidade Estadual da Paraíba, Guarabira, 2017.
- MUNRO, M.; DEWRING, T. Segmental acquisition in adult ESL learners. A longitudinal study of vowel production. **Language Learning**, v. 58, p. 479-502, 2008. Disponível em: <<https://doi.org/10.1111/j.1467-9922.2008.00448.x>>. Acesso em: 08 mai 2021.
- OLIVEIRA, L. Ensino de língua para jovens e adultos na escola pública. In: LIMA, D. (Ed.). **Ensino e aprendizagem de língua inglesa: conversa com especialistas**. São Paulo: Parábola, 2009. p. 21-31.



OLIVEIRA, C.; CORDANI, L. Julgando sob incerteza: heurísticas e vieses e o ensino de probabilidade e estatística. **Educ. Matem. Pesq**, São Paulo, v. 18, p. 1265-1289, 2016.

Disponível em: <

[OUSHIRO, L. **Introdução à Estatística para Linguistas**, v 1.0.1, Zenodo, 2017. Disponível em: <<http://doi.org/10.5281/zenodo.4755739>>. Acesso em: 22 jan. 2019.](https://revistas.pucsp.br/index.php/emp/article/download/31484/21943#:~:text=Julgando%20sob%20Incerteza%3A%20O%20Programa%20Heur%C3%ADsticas%20e%20vieses&text=Os%20mesmos%20autores%2C%20Kahneman%20e,uma%20variedade%20de%20padr%C3%B5es%20mentais.> . Acesso em: 02 dez. 2020.</p></div><div data-bbox=)

OWENS JR., R. **Language development: an introduction**. 8 ed. Boston: Pearson Education, 2012.

PARTEE, B.; TER MEULEN, A.; WALL, R. **Mathematical Methods in Linguistics**. Dordrecht: Kluwer Academic Publishers, 1993.

R CORE TEAM: **R. A Language and Environment for Statistical Computing**. Vienna: R Foundation for Statistical Computing, 2021.

RAMOS, E. Transferência fonológica no ensino de língua inglesa. In: LIMA, D. (Ed.). **Ensino e aprendizagem de língua inglesa: conversa com especialistas**. São Paulo: Parábola, 2009. p. 53-58.

SANKOFF, D. Statistics in Sociolinguistics. In: MESTHRIE, R. (Ed.). **Concise Encyclopedia of Sociolinguistics**. Oxford: Elsevier, 2001, p. 828-834.

SELL, F.; GONÇALVES, A. Sociolinguística. Indaial: Uniasselvi, 2011.

SELINKER, L. Interlanguage. **International Review of Applied Linguistics**, v. 10, p. 209-231, 1972. Disponível em: <<http://dx.doi.org/10.1515/iral.1972.10.1-4.209>>. Acesso em: 09 mai 2021.

SHADLE, C.; MAIR, S. Quantifying Spectral Characteristics of Fricatives. In: **Proceeding of Fourth International Conference on Spoken Language Processing**, Philadelphia, v. 3, p. 1521-1524, 1996. Disponível em: <https://www.isca-speech.org/archive/icslp_1996/index.html>. Acesso em: 09 mai 2021. D.O.I: 10.1109/ICSLP.1996.607906.

SILVA JR. L. **Erro de leitura das vogais do inglês americano como língua estrangeira pelos falantes do português do Brasil**. 2009. 69f. Dissertação (Mestrado em Linguística) – Programa de Pós-graduação em Linguística, Universidade Federal da Paraíba, João Pessoa (PB), 2009. Disponível em: <https://repositorio.ufpb.br/jspui/handle/tede/6496?locale=pt_BR>. Acesso em: 02 mai. 2020.

TRIOLA, M. **Introdução à Estatística: Atualização da Tecnologia**. 11 ed. São Paulo: Ed. Saraiva, 2014.