RELATO REGISTRADO: PROTOCOLO

REANÁLISE DE **VOGAIS MÉDIAS PRETÔNICAS** COM O ALINHADOR FORÇADO MFA

Livia OUSHIRO
Departamento de Linguística - Universidade de São Paulo (USP)
São Paulo, SP, Brasil

RESUMO

A Sociofonética combina a análise acústica com o estudo da fala natural de pessoas com diferentes perfis sociais. Um desafio do campo é lidar com grandes volumes de dados mantendo o rigor metodológico da Fonética e da Sociolinguística. A ferramenta de alinhamento forçado Montreal Forced Aligner pode contribuir nesse sentido, pois segmenta automaticamente os sons da fala; contudo, o alinhador ainda não foi amplamente testado para o português brasileiro. Este trabalho replica análises prévias que investigaram a pronúncia das vogais /e, o/ (como em "legal", "morada") na fala de migrantes nordestinos em São Paulo, observando a possível aproximação à variedade paulistana na situação de contato dialetal. Os estudos originais utilizaram a ferramenta EasyAlign com correção manual das vogais. Neste estudo, aplica-se o Montreal Forced Aligner aos mesmos dados para avaliar se os resultados obtidos são equivalentes, contribuindo para validar seu uso na pesquisa sociofonética com dados do português.

PALAVRAS-CHAVE

Sociofonética; Montreal Forced Aligner; Vogais Médias Pretônicas; Aquisição de Segundo Dialeto; Migrantes Nordestinos em São Paulo.

EDITORES
- Miguel Oliveira, Jr. (UFAL)
- Mailce Mota (UFSC)

AVALIADORES
- Pablo Arantes (UFSCar)
- Elisa Battisti (UFRGS)

Recebido: 30/10/2025
Aceito: 27/02/2026
Publicado: 18/05/2026

COMO CITAR

OUSHIRO, L. (2026). Reanálise de vogais médias pretônicas com o alinhador forçado MFA. *Cadernos de Linguística*, v. 7, n. 1, e905.

TITLE

REANALYSIS OF PRETONIC MID VOWELS WITH MFA

ABSTRACT

Sociophonetics combines acoustic analysis with the study of natural speech produced by individuals with different social profiles. A challenge in the field is managing large volumes of data while maintaining the methodological rigor of Phonetics and Sociolinguistics. The forced-alignment tool Montreal Forced Aligner can contribute in this regard, as it automatically segments speech sounds; however, the aligner has not yet been widely tested for Brazilian Portuguese. This study replicates previous analyses which investigated the pronunciation of the vowels /e, o/ (as in "legal," "morada") in the speech of Northeastern migrants in São Paulo, examining potential convergence toward the São Paulo variety in a dialect contact situation. The original studies used EasyAlign with manual correction of the vowels. In the present study, Montreal Forced Aligner is applied to the same data to assess whether the results are equivalent, thereby contributing to the validation of its use in sociophonetic research with Portuguese data.

KEYWORDS

Sociophonetics; Montreal Forced Aligner; Pretonic Mid Vowels; Second Dialect Acquisition; Northeastern Migrants in São Paulo.

INTRODUÇÃO

A Sociofonética é uma área que reúne os princípios, as questões de pesquisa e os métodos da Fonética e da Sociolinguística (Zimman, 2020; Baranowski, 2013; Foulkes; Scobbie; Watt, 2010; Silveira; Oushiro, 2022). Em comum, essas duas áreas se ocupam da análise de dados reais (i.e., não da intuição), do tratamento quantitativo desses dados e, de modo geral, orientam-se por abordagens empíricas (Thomas, 2013).

Por outro lado, também existem diferenças entre Fonética e Sociolinguística. De acordo com Jannedy e Hay (2006, p. 405), da perspectiva do foneticista, muito da prática variacionista pode ser vista como pouco sofisticada quanto ao rigor metodológico na análise do detalhe acústico. Com efeito, o foco dos estudos sociolinguísticos em grandes amostras de fala resulta em análises pouco detalhadas quanto a características acústicas e articulatórias, de modo que tendem a se pautar mais por categorias fonológicas do que propriamente fonéticas. Da perspectiva variacionista, a pesquisa fonética pode ser criticada pela ênfase em amostras de fala controlada ou "de laboratório", raramente de fala espontânea. Nesses estudos, a variação social é frequentemente tratada como fonte de "ruído"; embora muitos foneticistas reconheçam a existência de variação social, a maioria a considera como dado de pouco interesse teórico (Jannedy; Hay, 2006, p. 405).

É inegável, contudo, que as áreas da Fonética e da Sociolinguística têm se aproximado nos últimos anos. Uma forte evidência da expansão da Sociofonética é a presença de verbetes e capítulos em manuais recentes de ambas as áreas (p.ex., Foulkes; Scobbie; Watt, 2010; Thomas, 2013; Thomas, 2019; Zimman, 2020; Fagyal; Davidson, 2022), em volumes temáticos de revistas especializadas (p.ex., Foulkes; Docherty, 2006; Nycz, 2015) e o surgimento de obras dedicadas exclusivamente ao assunto (p.ex., Preston; Niedzielski, 2010; Celata; Calamai, 2014). Para Labov (2006, p. 500), o campo da Sociofonética se volta ao "estudo da sensibilidade de falantes e de ouvintes ao contexto social em que a língua é produzida e percebida", não havendo dúvidas de que os métodos experimentais e estatísticos dessas análises vão além das práticas correntes nos estudos sociolinguísticos. Labov (2006, p. 501) acrescenta, ainda, que embora o termo possa sugerir a emergência de uma nova disciplina, tal visão é uma "distração" do ponto principal: a aproximação de duas linhas de pesquisa até então apartadas.

Um dos desafios para o desenvolvimento da Sociofonética é o tratamento de dados coletados em situações que se aproximam daquelas de conversas naturais e espontâneas, com vistas a superar o Paradoxo do Observador (Labov, 2006 [1966]), uma vez que o interesse do sociolinguista é observar sistematicamente a fala cotidiana. Isso não raro resulta em gravações com ruídos do ambiente, sobreposições de vozes de interlocutores, hesitações ou inícios interrompidos. Além disso, normalmente se analisam dezenas ou centenas de horas de gravação com múltiplos

indivíduos, totalizando milhares de dados, como costuma ocorrer em estudos sobre variáveis fonético-fonológicas.

O tratamento acústico de tal quantidade de dados tem se facilitado, nas últimas décadas, pelo advento de ferramentas de transcrição alinhada e análise acústica, como o Praat (Boersma; Weenink, 2023) e o ELAN (Sloetjes; Wittenburg, 2008), e de segmentação automática, como o EasyAlign (Goldman, 2011), FAVE (Rosenfelder *et al.*, 2014), Prosodylab-Aligner (Gorman *et al.*, 2011) e Montreal Forced Aligner (MFA; McAuliffe *et al.*, 2017) – sendo este último o mais utilizado em pesquisas recentes pelo número de línguas que comporta e pela possibilidade de treinamento do modelo em novas línguas ou variedades linguísticas.

O MFA é um sistema de acesso aberto que realiza a segmentação acústica de dados a partir de arquivos de áudio e de transcrição ortográfica, baseando-se em modelos acústicos trifônicos para capturar a variabilidade contextual e que inclui adaptações por falantes para modelar diferenças individuais (McAuliffe *et al.*, 2017, p. 498). Sua aplicação se dá em linguagem Python por meio de linhas de comando. Dentre os modelos já treinados, há um para o Português Brasileiro – Portuguese (Brazil) MFA G2P model v2.0.0 – junto a um dicionário – Portuguese (Brazil) MFA dictionary v2.0.0 – (McAuliffe; Sonderegger, 2022); esses modelos, entretanto, ainda não foram testados sistematicamente em dados de entrevistas sociolinguísticas nessa língua (mas ver Batista *et al.*, 2022).

Nesse sentido, o objetivo deste estudo é o de testar e avaliar o alinhador automático MFA com dados semi-espontâneos do Português Brasileiro, obtidos em entrevistas sociolinguísticas advindas do Projeto Coesão & Dispersão (Oushiro, 2023; FAPESP 2023/00968-7) com 39 falantes, cada qual com cerca de uma hora de duração.

Para tanto, replicam-se as análises de dois estudos prévios – Oushiro (2019a, 2019b) – sobre a realização de vogais médias pretônicas (como em "legal" e "morada") na fala de migrantes alagoanos e paraibanos residentes na cidade de São Paulo, em comparação com a fala de paulistanos não migrantes, cujo objetivo foi o de avaliar em que medida os migrantes adotam a pronúncia relativamente menos baixa das vogais pretônicas /e, o/. Nesses estudos, as vogais médias pretônicas foram analisadas acusticamente a partir de segmentação automática realizada com o alinhador EasyAlign (Goldman, 2011)[1] e posteriormente revisada manualmente pela pesquisadora. Ainda que o objetivo em Oushiro (2019a, 2019b) não tenha sido o de avaliar a acurácia do EasyAlign, vale notar que seu emprego reduziu substancialmente o tempo despendido na tarefa de segmentação acústica de milhares de vogais, ao mesmo tempo em que, por outro lado, se atestou a necessidade de revisão humana, uma vez que o alinhamento automático com o EasyAlign pareceu

1 O EasyAlign é um *plugin* que se instala no programa Praat e que realiza, a partir de um arquivo de áudio e a respectiva transcrição ortográfica (ou fonética), a segmentação de enunciados, a conversão de grafema para fonemas e a segmentação fonética (Goldman, 2011, p. 1). Para uma descrição de sua aplicação, ver Oushiro (2019a).

ser menos preciso em trechos de fala rápida, com ruídos e em segmentos articulatoriamente reduzidos (ver Seção 1).

Uma vez que os dados dos estudos originais passaram por revisão humana, o presente trabalho assume que a segmentação e as medições deles extraídas podem ser tomadas como parâmetro confiável para a testagem e a avaliação do MFA, objetivo ao qual este estudo se volta. Mais especificamente, busca-se responder as seguintes questões: (i) aplicando-se a segmentação automática do MFA sobre os mesmos dados, como se comparam os resultados de análises multivariadas de regressão linear (incluindo as variáveis previsoras Gênero, Escolaridade, Faixa Etária, Idade de Migração, Tempo de Residência, Motivo de Migração e Falante) e os resultados de análise de covariação entre a altura das vogais /e/ e /o/ com aqueles dos estudos prévios?; (ii) qual é a taxa de acerto do MFA para as fronteiras de vogais pretônicas /i, e, a, o, u/ em dados de fala semiespontânea de entrevistas sociolinguísticas?

A próxima seção descreve os métodos e os resultados dos estudos a serem replicados, assim como o contexto mais amplo da presente pesquisa. Em seguida, a seção "Métodos" apresenta o protocolo de tratamento de dados do presente estudo e o plano da análise que se objetiva desenvolver.

1. ESTUDOS REPLICADOS E CONTEXTO DA PESQUISA

Em situação de contato dialetal, comumente se observam formas linguísticas híbridas ou intermediárias (Kerswill, 1994; Siegel, 2010). Por exemplo, na fala de migrantes nordestinos residentes em São Paulo, é possível que suas realizações de vogais médias pretônicas /e, o/ não sejam tão baixas quanto a de nordestinos não migrantes (como 'l[ɛ]gal, 'm[ɔ]rada'), tampouco tão alta quanto a de paulistanos (como 'l[e]gal, 'm[o]rada'). Com efeito, da perspectiva sociofonética, é de se esperar que haja variação na altura da vogal em todos os grupos de falantes, migrantes ou não. Uma codificação de oitiva, com base apenas na percepção do pesquisador, pode ser enganosa: é possível que realizações intermediárias de /e, o/ sejam percebidas como média baixas por alguns (quiçá, os paulistas) e como média altas por outros (quiçá, os nordestinos).

Considerando a gradiência sociofonética, sobretudo na fala de migrantes, Oushiro (2019a) descreve os procedimentos metodológicos para o tratamento acústico de vogais médias pretônicas /e, o/ em dados de entrevistas sociolinguísticas.[2] O corpus consistiu de 39 gravações: 21 alagoanos e 11 paraibanos residentes na cidade de São Paulo, em contraste com a fala de sete paulistanos não-migrantes como amostra-controle. Os dados haviam sido coletados, respectivamente, para a

2 Dados e scripts disponíveis em: https://zenodo.org/records/17443974.

pesquisa de mestrado de Silva (2014), para o Projeto Casadinho (Hora; Negrão, 2011) e para o Projeto SP2010 (Mendes; Oushiro, 2012), sendo gentilmente cedidos pelos responsáveis ao Projeto "Processos de acomodação dialetal na fala de nordestinos residentes em São Paulo" (Oushiro, 2016; FAPESP 2016/04960-7[3]). As amostras de migrantes e paulistanos foram ambas balanceadas pelo gênero (feminino, masculino), três faixas etárias (20–34, 35–59, 60 ou mais anos) e dois níveis de escolaridade (até Ensino Médio, Nível Universitário). Das gravações, também se extraíram informações sobre idade e motivo da migração, tempo de residência em São Paulo e local de origem (rural ou urbano).[4]

Os dados das 39 gravações passaram pelas seguintes etapas de tratamento:

1) Transcrição alinhada à onda sonora no programa ELAN (Sloetjes; Wittenburg, 2008);
2) Aplicação, sobre as transcrições, do EasyAlign (Goldman, 2011), a partir de uma transcrição ortográfica (ver Figura 1: a partir da trilha de transcrição, "S1", o EasyAlign criou as trilhas phono, phones e words).

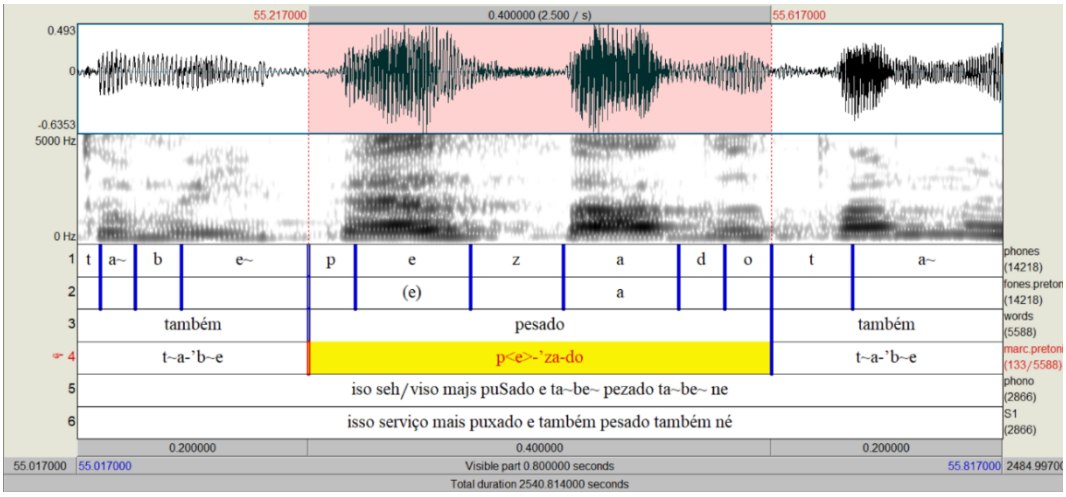

Figura 1. Transcrição e segmentação no Praat após a aplicação do EasyAlign e marcação de vogais pretônicas, antes da revisão humana. Fonte: Oushiro (2019a, p. 164). ALT: Espectrograma, transcrição e segmentação acústica de trecho de fala gravada. A figura destaca a palavra "pesado", sua segmentação nos fones /p/, /e/, /z/, /a/, /d/ e /o/ na trilha "phones", assim como segmentação das vogais-alvo do estudo de Oushiro (2019a) – a vogal /e/ pretônica e a vogal seguinte /a/.

3 Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Universidade Estadual de Campinas sob o número de protocolo 72362317.4.0000.5404 em 05/09/2017. Todos os participantes foram devidamente informados sobre os procedimentos do estudo e assinaram o Termo de Consentimento Livre e Esclarecido.

4 Esta amostra de 32 alagoanos e paraibanos migrantes foi usada como *corpus* exploratório para o levantamento de hipóteses a serem testadas em nova amostra, coletada no âmbito do mesmo projeto. Ver descrição das etapas do Projeto Acomodação em Oushiro (2023).

3) Aplicação, sobre as transcrições, do *script* silac/silacpret (Oushiro, 2018). O silac é um código em linguagem R que transforma uma transcrição ortográfica em transcrição fonêmica. Com base no silac, desenvolveu-se uma sequência adicional de código (denominada "silacpret"[5]) para identificação e marcação de segmentos pretônicos. Assim, a palavra "telefone", por exemplo, é transcrita como "t<e>-l<e>-'fo-ne". A transcrição gerada pelo silacpret foi inserida na trilha "marc.pretonicas" (Figura 1).

4) Sobre o arquivo .TextGrid segmentado pelo EasyAlign, criação de nova trilha, "fones.pretonicas", para revisão humana da segmentação automática e para anotação de ocorrências de vogais médias pretônicas /e, o/ e outras vogais de interesse (/i, a, u/, bem como as vogais das sílabas seguintes às pretônicas). O ajuste de fronteiras de vogais foi feito a partir da oitiva, com fones de ouvido, do trecho em que se encontra o segmento e da inspeção visual do espectrograma; além de pistas com base nos segmentos adjacentes à vogal (como a ausência da barra de vozeamento em oclusivas surdas ou a existência de ruídos acima de 5000 Hz em fricativas), a delimitação das vogais pautou-se pela clara presença das faixas de frequência F1 e F2, tomando-se esta última como principal critério para ajuste de fronteiras, juntamente à amplitude da onda. Para anotação na trilha "fones.pretonicas", adotaram-se os seguintes critérios: (i) busca de ocorrências (p.ex.: "<e") na trilha "marc.pretonicas"; (ii) seleção de ocorrências em contextos favorecedores da aplicação da regra de abaixamento vocálico, segundo Pereira (1997): vogais médias pretônicas /e, o/ cuja sílaba seguinte contém uma vogal média baixa /ɛ, ɔ, ɛ̃, ɔ̃/ ou baixa /a, ã/; (iii) anotação adicional das vogais pretônicas /i, a, u/, para posterior normalização das vogais e plotagem do espaço vocálico; (iv) para cada vogal média pretônica, anotação da vogal da sílaba seguinte, para posterior análise dessa variável previsora; (v) descarte de ocorrências de vogais, mesmo que dentro dos parâmetros acima, caso ocorresse em sobreposição de vozes, com ruído excessivo ao fundo ou com medições de F1/F2 espúrias, de acordo com os parâmetros de Barbosa e Madureira (2015) para vogais do português brasileiro. Em cada gravação, anotaram-se aproximadamente 50 ocorrências para cada vogal média pretônica /e, o/ e 30 ocorrências para cada vogal /i, a, u/.

5) Sobre a trilha "fones.pretonicas", aplicação do script Vowel Analyzer (Riebold, 2013) para extração de medidas de F1, F2 e F3 em 30%, 50% e 70% das vogais, além de sua duração, para uma planilha.

6) Exclusão de outliers das vogais /i, e, a, o, u/ (medições com desvio 1,5 vezes maior do que o intervalo interquartil Q1-Q3).

7) Normalização das vogais pelo método de Lobanov (1971) por meio da linguagem R.

8) Codificação automática de variáveis previsoras linguísticas e sociais,[6] por meio de script em R elaborado pela autora.[7]

5 Disponível em https://github.com/oushiro/CoesaoeDispersao/blob/main/silacpret-adapt8-nov23.R.

6 Linguísticas: Contexto fônico precedente, contexto fônico seguinte, F1 da vogal da sílaba seguinte, F2 da vogal da sílaba seguinte, vogal tônica, distância da vogal tônica (em número de sílabas), estrutura da sílaba pretônica; Sociais: Amostra (ALSP, PBSP, SP2010), Informante (efeito aleatório), Gênero (feminino, masculino), Nível de Escolaridade (até Ensino Médio, Ensino Universitário), Faixa etária (20-34, 35-59, 60+ anos)/Idade (em anos), Idade de migração (9-17, 18-24, 25+ anos), Tempo de residência em SP (menos de 10, 11-29, 30+ anos), Motivo da migração (estudo, família, qualidade de vida, trabalho).

7 Ver arquivo pos-VowelAnalyzer-v4.R, disponível em https://zenodo.org/records/17443974.

Em Oushiro (2019a), esses dados foram analisados em modelos de regressão linear de efeitos mistos (função lmer do pacote lme4), com o objetivo de identificar variáveis sociais correlacionadas com a aquisição de vogais médias pretônicas [-baixa]. Como o conjunto de variáveis sociais não é ortogonal entre si, foram criadas dois modelos: (i) um contendo as variáveis estratificadoras das amostras: Gênero, Nível de Escolaridade e Faixa Etária, além de Falante como efeito aleatório (modelos 9a e 9c); e (ii) um contendo outras variáveis sociais diretamente relacionadas à trajetória do migrante: Idade de Migração, Tempo de Residência e Motivo da Migração, além de Falante como efeito aleatório (modelos 9b e 9d).

9) Modelos de regressão linear:

a. Modelo E1: `F1.NORM ~ GENERO + ESCOLARIDADE + FAIXA.ETARIA + (1|PARTICIPANTE), data = vogal.e`

b. Modelo E2: `F1.NORM ~ IDADE.MIGRACAO + TEMPO.SP + MOTIVO.MIGRACAO + (1|PARTICIPANTE), data = vogal.e`

c. Modelo O1: `F1.NORM ~ GENERO + ESCOLARIDADE + FAIXA.ETARIA + (1|PARTICIPANTE), data = vogal.o`

d. Modelo O2: `F1.NORM ~ IDADE.MIGRACAO + TEMPO.SP + MOTIVO.MIGRACAO + (1|PARTICIPANTE), data = vogal.o`

Os resultados da análise de 1.916 ocorrências de /e/ e 1.645 ocorrências de /o/ mostraram que apenas a Idade de Migração se correlaciona significativamente (nível-α = 0,05) com a altura da vogal média pretônica, para ambas as vogais: quanto mais velho migrou o falante, maior a estimativa de F1, ou seja, maior a tendência a vogais médias relativamente mais baixas. Visto de outro modo, os falantes que migraram mais jovens tendem a realizar vogais médias relativamente mais altas, semelhante ao padrão paulistano, o que é indicativo de aquisição do dialeto da comunidade anfitriã.

Em Oushiro (2019b), os mesmos dados foram utilizados para a análise de covariação (Guy, 2013; Guy; Hinskens, 2016) entre /e/ e /o/. A pergunta principal desse estudo voltou-se à uniformidade linguística: os migrantes tendem a realizar ambas as vogais médias pretônicas em alturas semelhantes, ou a aquisição do traço paulistano pode ocorrer de modo assimétrico? Para gerar a média da altura das vogais /e/ e /o/ por indivíduo, rodaram-se modelos de regressão linear de efeitos fixos (função lm), com os dados de paulistanos (SP2010) como nível de referência e falante como efeito fixo, em subconjuntos de dados separados por gênero e vogal (10a–10d). Em seguida, as estimativas para cada falante foram utilizadas em um teste de correlação de Spearman (10e), que revelou uma correlação significativa, mas fraca, entre as variáveis (ρ = 0,38, p = 0,03).

10) Análise de covariação entre /e/ e /o/:
    a. Modelo EF: `lm(F1.NORM ~ INFORMANTE2, data = subset(pret, SEXO %in% "feminino" & VOGAL %in% "e"))`
    b. Modelo EM: `lm(F1.NORM ~ INFORMANTE2, data = subset(pret, SEXO %in% "masculino" & VOGAL %in% "e"))`
    c. Modelo OF: `lm(F1.NORM ~ INFORMANTE2, data = subset(pret, SEXO %in% "feminino" & VOGAL %in% "o"))`
    d. Modelo OM: `lm(F1.NORM ~ INFORMANTE2, data = subset(pret, SEXO %in% "masculino" & VOGAL %in% "o"))`
    e. Teste de correlação de Spearman: `cor.test(estimativas.e, estimativas.o, method = "spearman")`

Uma análise mais detalhada dos indivíduos indicou que apenas sete das 17 mulheres adquiriram ambas as vogais /e/ e /o/ paulistanas, ao passo que as demais adquiriram apenas uma delas; dentre os homens, seis dos 15 adquiriram a altura [-baixa] de ambas as vogais, enquanto seis se acomodaram a apenas uma delas e três a nenhuma. A aquisição assimétrica explica a fraca correlação observada na comparação entre a altura das vogais.

A presente pesquisa dá continuidade ao estudo da fala de migrantes e se situa no contexto do "Projeto Coesão & Dispersão: Análise sociofonética da variação idioletal em situação de contato entre dialetos" (Oushiro, 2023; FAPESP 2023/00968-7[8]), cujo objetivo principal é o de analisar, na fala de migrantes nordestinos que residem no estado de São Paulo, o papel da variação individual face a outros fatores macrossociais (como gênero do falante, idade e motivo de migração, tempo de residência na nova localidade). A pesquisa parte da observação de que, ao lado de claros padrões de mudança na fala de migrantes em situação de contato – p.ex., a migração em idade mais jovem favorece a aquisição de traços fonéticos da comunidade anfitriã, mas não de traços morfológicos (Oushiro *et al.*, 2023) –, também se constata uma grande variação interfalantes, sendo a variável Indivíduo aquela que explica a maior parte da variância (Silveira, 2022; Oushiro *et al.*, 2023).

O corpus do Projeto Coesão & Dispersão compreende 165 gravações de entrevistas sociolinguísticas: (i) 32 migrantes alagoanos e paraibanos residentes na cidade de São Paulo (Amostra 1, analisada em Oushiro 2019a, 2019b); (ii) 40 migrantes alagoanos e paraibanos residentes na cidade de Campinas (Amostra 2); (iii) 48 gravações da amostra PORTAL, com alagoanos não

8 Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Universidade Estadual de Campinas sob o número de protocolo 71009123.0.0000.8142 em 03/09/2023. Todos os participantes foram devidamente informados sobre os procedimentos do estudo e assinaram o Termo de Consentimento Livre e Esclarecido.

migrantes (Oliveira, 2017); (iv) 24 gravações da amostra VALPB, com paraibanos não migrantes (Hora, 2005); (v) 12 gravações da amostra SP2010 (Mendes; Oushiro, 2012), com paulistanos não migrantes; e (vi) 9 gravações da amostra Mourão (Mourão, 2024), com campineiros não migrantes. Para processamento de tal quantidade de dados, com vistas à análise de quatro variáveis sociolinguísticas (/e/ e /o/ pretônicos, /r/ e /s/ em coda silábica), a equipe do projeto[9] tem se voltado às tarefas de desenvolvimento de protocolos, *scripts* e ferramentas, dentre as quais se encontra a testagem do alinhador forçado MFA.

2. MÉTODOS

Embora o corpus de análise do Projeto Coesão & Dispersão seja mais amplo do que aquele analisado em Oushiro (2019a, 2019b), com 165 gravações, este estudo analisará os mesmos dados dos trabalhos originais (ou seja, as 32 gravações da Amostra 1 e sete gravações da Amostra SP2010) para comparabilidade de resultados.

O Quadro 1 apresenta o protocolo da presente análise (na coluna à direita), que buscará seguir, tão proximamente quanto possível, as etapas das análises de Oushiro (2019a, 2019b), delineadas na Seção 2 e reproduzidas novamente abaixo (na coluna do meio). Nesse quadro, novas etapas são indicadas por →, execuções idênticas são indicadas por = e execuções adaptadas por *.

9 A equipe do Projeto Coesão & Dispersão é formada por alunos de graduação (Isabela Gemma, Clara Moreira, Luis Henrique Alvarenga, Samuel Sprogis), de pós-graduação (Gustavo Silveira, Leonardo Ferraz, Sarah Poli Barbosa, Maylle Freitas, Iris Massuci Dutra) e pelos pesquisadores colaboradores Plinio Barbosa (UNICAMP) e Jennifer Nycz (Georgetown University).

Etapas	**Oushiro (2019a, 2019b)**	**Este estudo**
Protocolo pré-análise		
(1)	Transcrição alinhada à onda sonora no programa ELAN	* Transcrição alinhada à onda sonora no programa ELAN e revisão da transcrição
(2)	Aplicação do EasyAlign e correção manual	→ Inserção de vocábulos faltantes no dicionário do MFA e revisão de todas as entradas → Codificação da qualidade da gravação (péssima, ruim, boa, muito boa, ótima) a depender da clareza das vozes dos interlocutores e do ruído de fundo * Aplicação do MFA
(3)	Aplicação do *script* silacpret	=
(4)	Criação de nova trilha, "fones.pretonicas", para anotação de ocorrências de interesse	→ Criação de *script* em linguagem R para identificação, nos dados segmentados pelo MFA, das mesmas ocorrências analisadas em Oushiro (2019a, 2019b)
(5)	Aplicação do *script* Vowel Analyzer sobre a trilha "fones.pretonicas"	* Aplicação do *script* Vowel Analyzer sobre a trilha "phones", gerada pelo MFA
(6)	Exclusão de *outliers* das vogais /i, e, a, o, u/	=
(7)	Normalização das vogais pelo método de Lobanov	=
(8)	Codificação automática de variáveis previsoras linguísticas e sociais	=
Análises		
→		Criação de duas planilhas: (i) MFA.19, contendo os mesmos tokens das análises de Oushiro (2019a, 2019b); e MFA.26, contendo todas as ocorrências de vogais pretônicas que se encaixem no envelope de variação definido em Oushiro (2019a, 2019b; ver item (4-ii na seção 2).
(9)	Modelos de regressão linear de efeitos mistos	= Análise de reprodutibilidade nos conjuntos de dados MFA.19 e MFA.26
(10)	Modelos de regressão linear de efeitos fixos e teste de correlação de Spearman	= Análise de reprodutibilidade nos conjuntos de dados MFA.19 e MFA.26
→		Análise de acurácia do MFA

Quadro 1. Comparação entre protocolos de análise de Oushiro (2019a, 2019b) e do presente estudo. ALT: Quadro com descrição dos passos da análise de Oushiro (2019a, 2019b) e do presente estudo. Novas etapas são indicadas por uma flecha à direita (→), execuções idênticas são indicadas por sinal de igual (=) e execuções adaptadas por asterisco (*).

As primeiras aplicações do MFA aos dados indicaram a necessidade de adicionar itens lexicais faltantes ao dicionário Portuguese (Brazil) MFA dictionary v2.0.0 (McAuliffe; Sonderegger, 2022) e de revisar todas as suas entradas e respectivas transcrições fonêmicas.[10] Além disso, notou-se que o MFA tende a errar a segmentação (i) na presença de ruídos de fundo, sobreposição de vozes, alongamentos segmentais, trechos ininteligíveis; (ii) em anotações longas; e (iii) em itens lexicais com pronúncia frequentemente reduzida na fala espontânea, como conjugações do verbo "estar" (*tá*, *tô*, *tamos*) e o pronome "você" (cê). Desse modo, todas as transcrições do projeto foram revisadas de modo a separar trechos com ruídos, sobreposições etc. em anotações isoladas; quebrar anotações com mais de 8 segundos; e mudar a transcrição de ocorrências do verbo

10 A versão final do dicionário revisado encontra-se em https://github.com/oushiro/CoesaoeDispersao/blob/main/Dicionario_ProjCoesao_v6.txt.

"estar" e do pronome "você" para *tá*, *tô*, *cê* etc. quando fosse o caso de pronúncias reduzidas. Tendo em vista que a presença de ruídos interfere no funcionamento do MFA, a equipe do projeto também classificou, impressionisticamente, todas as gravações quanto à sua qualidade acústica: péssima (presença excessiva de ruídos); ruim (presença de ruídos mas que permitem uma boa análise de variáveis morfossintáticas ou lexicais, mas talvez não fonéticas); boa (vozes de interlocutores bem audíveis, apesar de ruídos de fundo); muito boa (vozes de interlocutores bem audíveis e ruídos de fundo ocasionais); ótima (vozes de interlocutores bem audíveis e virtual inexistência de ruídos de fundo).

A principal diferença entre o protocolo de Oushiro (2019a, 2019b) e o presente estudo é a aplicação do alinhador automático forçado MFA, em lugar do EasyAlign com correção manual (Etapa 2 do Quadro 1). Sobre o arquivo segmentado gerado pelo MFA, serão executados os mesmos passos (3), (6), (7) e (8), a partir dos mesmos *scripts*. Em lugar da Etapa (4) (anotação de ocorrências na trilha "fones.pretonicas"), será criado um *script* em linguagem R para identificação automática das ocorrências analisadas nos estudos prévios; desse modo, a aplicação do script Vowel Analyzer, para extração de medidas de formantes, se dará sobre a própria trilha "phones" gerada automaticamente pelo MFA (Etapa 5).

Para as análises comparativas, serão criadas duas planilhas: uma que contenha as mesmas ocorrências das análises de Oushiro (2019a, 2019b) – a ser chamada MFA.19 – e outra – MFA.26 – que contenha todas as ocorrências de vogais pretônicas que se conformam ao envelope de variação (vogais médias pretônicas /e, o/ cuja sílaba seguinte contém uma vogal média baixa /ɛ, ɔ, ɛ̃, ɔ̃/ ou baixa /a, ã/), definido a partir da descrição de Pereira (1997) como contexto favorecedor de abaixamento vocálico. O primeiro conjunto de dados possibilitará a comparação direta com os de Oushiro (2019a, 2019b), de modo a avaliar quão confiáveis são os resultados sem correções manuais do pesquisador; quanto ao segundo, considerando que Oushiro (2019a, 2019b), na etapa de revisão manual, também descartou certas ocorrências com medições espúrias (mesmo que segmentadas corretamente), esse conjunto de dados permitirá avaliar se a exclusão de *outliers* apenas por critérios estatísticos (Etapa 6) também gera resultados confiáveis.

Serão conduzidos dois tipos de análise sobre esses dados: (i) análise de reprodutibilidade de resultados; e (ii) análise da acurácia do MFA. No primeiro tipo de análise, os mesmos modelos de regressão linear de efeitos mistos realizados em Oushiro (2019a) e os mesmos modelos de regressão linear de efeitos fixos e teste de correlação de Spearman realizados em Oushiro (2019b), discriminados em (9a)–(9d) e (10a)–(10e) mais acima, serão testados sobre os dois conjuntos de dados MFA.19 e MFA.26. Tais resultados serão contrastados com aqueles das publicações originais.

Para avaliação desses resultados, serão aplicados critérios qualitativos e quantitativos. Qualitativamente, se considerará que os resultados são equivalentes caso as mesmas correlações (ou falta de) forem observadas, nas mesmas direções – ou seja, (i) correlação apenas com a variável Idade de Migração, no sentido de que quanto mais jovem migrou, maior tendência à

realização de vogais [-baixa]; (ii) correlação entre a altura das vogais /e/ e /o/ significativa, mas fraca; e (iii) mesmo padrão geral de apenas cerca de metade dos migrantes terem adquirido a altura paulistana de ambas as vogais. Quantitativamente, os resultados dos modelos serão comparados por meio da função compare_performance do pacote performance, através das medidas de R^2, AIC/BIC e erro padrão.

Para o segundo tipo de análise, relativo à acurácia da segmentação do MFA, será utilizado apenas o conjunto de dados MFA.19, contendo os mesmos tokens, em comparação com o conjunto original de dados. Considerando a revisão humana da segmentação automática feito pelo *plugin* EasyAlign, ressalta-se a assunção de que a segmentação de dados dos estudos originais está correta, o que servirá de parâmetro para avaliação da acurácia do alinhamento automático a ser feito pelo MFA. Será computada uma medida de desvio da fronteira inicial e da fronteira final entre as medições dos estudos originais e aquelas do MFA, em milisegundos: begin_time_orig – begin_time_MFA e end_time_orig – end_time_MFA. Sobre os valores de desvios, serão computadas estatísticas descritivas (média, mediana, desvio padrão) e serão criados modelos de regressão para avaliar se o desvio aumenta ou diminui a depender (i) da qualidade da gravação (péssima, ruim, boa, muito boa, ótima); (ii) da vogal pretônica (i, e, a, o, u); e (iii) contexto fonológico (oclusivas, fricativas, africadas, líquidas, nasais). Seguindo Mackenzie e Turton (2020), será considerado um desvio aceitável o limite de 20ms e uma taxa de concordância de 80% ou mais.

3. HIPÓTESES INICIAIS

Levantam-se as seguintes hipóteses quanto à performance do MFA:[11]

11) Análise de reprodutibilidade de resultados:
- **a.** A análise sobre os dados segmentados pelo MFA produzirá os mesmos resultados das análises originais.

12) Análise da acurácia do MFA:
- **a.** O desvio entre as fronteiras iniciais e finais da segmentação automática do MFA estará dentro do limite aceitável de 20 ms.

11 É importante esclarecer que as formulações em (11) e (12) se referem, efetivamente, a *hipóteses*, e não a *expectativas* de resultados. Considerando a escassez de trabalhos que avaliam o desempenho de alinhadores automáticos em dados de português brasileiro falado e semiespontâneo, a avaliação dessas hipóteses à luz dos resultados permitirá descrever a acurácia do MFA não só de modo global, mas também quanto a condicionadores específicos.

b. A segmentação do MFA será mais precisa quanto melhor for a qualidade acústica da gravação.

c. A segmentação do MFA será mais precisa em vogais médias e baixas (e, a, o).

d. A segmentação do MFA será mais precisa em contextos de consoantes oclusivas.

AGRADECIMENTOS

A autora agradece aos valiosos comentários e sugestões dos avaliadores do artigo. Qualquer equívoco que permaneça, evidentemente, é de minha responsabilidade.

INFORMAÇÕES COMPLEMENTARES

CONFLITO DE INTERESSE

A autora declara que não possui interesses financeiros ou relações pessoais que possam ter influenciado o trabalho relatado neste artigo.

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS

A autora declara que os dados analisados e os códigos de análise se encontram em repositórios abertos e de livre acesso (Licença Creative Commons Attribution 4.0 International): https://zenodo.org/records/17443974 e https://github.com/oushiro/CoesaoeDispersao.

PRÉ-REGISTRO DE PESQUISA

O pré-registro desta pesquisa se encontra na plataforma OSF (https://doi.org/10.17605/OSF.IO/T9NYS). Não houve alterações no plano analítico descrito no pré-registro.

DECLARAÇÃO DE USO DE IA

A ferramenta ChatGPT foi usada para consulta de como comparar resultados de diferentes modelos de regressão.

CONSENTIMENTO E ÉTICA

Esta pesquisa foi aprovada pelo Comitê de Ética em Pesquisa da Universidade Estadual de Campinas sob o número de protocolo 71009123.0.0000.8142 em 03/09/2023. Todos os participantes foram devidamente informados sobre os procedimentos do estudo e assinaram o Termo de Consentimento Livre e Esclarecido.

FONTES DE FINANCIAMENTO

O presente trabalho foi realizado com o apoio de Auxílio à Pesquisa FAPESP (Processo 2023/00968-7) e Bolsa Produtividade em Pesquisa CNPq-Nível C (Processo 301759/2025-1).

AVALIAÇÃO E RESPOSTA DOS AUTORES

Avaliação: https://doi.org/10.25189/2675-4916.2026.V7.N1.ID905.R
Resposta dos Autores: https://doi.org/10.25189/2675-4916.2026.V7.N1.ID905.A

REFERÊNCIAS

BARANOWSKI, M. Sociophonetics. *In*: BAYLEY, R.; CAMERON, R.; LUCAS, C. (Ed.). *The Oxford Handbook of Sociolinguistics*. Oxford: Oxford University Press, 2013. p. 403–424.

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

BATISTA, C.; DIAS, A. L.; NETO, N. Free resources for forced phonetic alignment in Brazilian Portuguese based on Kaldi toolkit. *EURASIP Journal on Advances in Signal Processing*, v. 2022, n. 11, p. 1–32, 2022. Disponível em: https://doi.org/10.1186/s13634-022-00844-9. Acesso em: 15 abr 2026.

BOERSMA, P.; WEENINK, D. Praat: doing phonetics by computer. 2023. Disponível em: http://www.fon.hum.uva.nl/praat/. Acesso em: 04 abr 2026.

CELATA, C.; CALAMAI, S. (Ed.). *Advances in Sociophonetics*. Amsterdam/Philadelphia: John Benjamins, 2014.

FAGYAL, Z.; DAVIDSON, J. Sociophonetics. *In*: GABRIEL, C.; GESS, R.; MEISENBURG, T. (Ed.). *Manual of Romance Phonetics and Phonology*. Berlin/Boston: Walter de Gruyter, 2022. p. 342–374.

FOULKES, P.; DOCHERTY, G. The social life of phonetics and phonology. *Journal of Phonetics*, v. 34, p. 409–438, 2006.

FOULKES, P.; SCOBBIE, J. M.; WATT, D. Sociophonetics. *In*: HARDCASTLE, W. J.; LAVER, J.; GIBBON, F. E. (Ed.). *The Handbook of Phonetic Sciences*. Oxford: Wiley Blackwell, 2010. p. 703–754.

GOLDMAN, J.-P. Easyalign: an automatic phonetic alignment tool under Praat. *In*: *Interspeech, 2011, Florence, Italy. 12th Annual Conference of the International Speech Communication Association*, 2011, p. 3233–3236. Disponível em: https://archive-ouverte.unige.ch/unige:18188. Acesso em: 04 abr 2026.

GORMAN, K.; HOWELL, J.; WAGNER, M. Prosodylab-aligner: A tool for forced alignment of laboratory speech. *Canadian Acoustics*, v. 39, n. 3, p. 192–193, 2011.

GUY, G. R. The cognitive coherence of sociolects: how do speakers handle multiple sociolinguistic variables? *Journal of Pragmatics*, v. 52, p. 63–71, 2013.

GUY, G. R.; HINSKENS, F. Linguistic coherence: Systems, repertoires and speech communities. *Lingua*, v. 172–173, p. 1–9, 2016.

HORA, D. *Projeto Variação Linguística no Estado da Paraíba*. João Pessoa: Ideia, 2005.

HORA, D.; NEGRÃO, E. V. (Ed.). *Estudos da Linguagem*: casamento entre temas e perspectivas. João Pessoa: Ideia, 2011.

JANNEDY, S.; HAY, J. Modelling sociophonetic variation. *Journal of Phonetics*, v. 34, p. 405–408, 2006.

KERSWILL, P. *Dialects converging*. Rural speech in urban Norway. Oxford: Oxford University Press, 1994.

LABOV, W. A sociolinguistic perspective on sociophonetic research. *Journal of Phonetics*, v. 34, p. 500–515, 2006.

LOBANOV, B. M. Classification of Russian vowels spoken by different speakers. *The Journal of the Acoustical Society of America*, v. 49, n. 2, p. 606–608, 1971.

MACKENZIE, L.; TURTON, D. Assessing the accuracy of existing forced alignment software on varieties of British English. *Linguistics Vanguard*, v. 6, n. 1, p. 1–14, 2020.

MCAULIFFE, M.; SOCOLOF, M.; MIHUC, S.; WAGNER, M.; SONDEREGGER, M. Montreal Forced Aligner: Trainable Text-Speech Alignment Using Kaldi. *In*: *Proc. Interspeech 2017*, Stockholm-Sweden, p. 498–502, 2017.

MCAULIFFE, M.; SONDEREGGER, M. Portuguese (Brazil) MFA dictionary v2.0.0. Disponível em: https://mfa-models.readthedocs.io/en/latest/dictionary/Portuguese/Portuguese%20%28Brazil%29%20MFA%20dictionary%20v2_0_0a.html, 2022. Acesso em: 04 abr 2026.

MENDES, R. B.; OUSHIRO, L. O paulistano no mapa sociolinguístico brasileiro. *Alfa*, vol. 56, n. 3, p. 973–1001, 2012.

MOURÃO, N. R. *A realização de (t/d) e (-r) em coda em Campinas e Jundiaí: uma proposta sobre mobilidade*. Dissertação (Mestrado em Linguística). Instituto de Estudos da Linguagem, Unicamp, Campinas, 2024.

NYCZ, J. Second dialect acquisition: A sociophonetic perspective. *Language and Linguistics Compass*, v. 9, p. 469–482, 2015.

OLIVEIRA, A. J. Projeto PORTAL: variação linguística no português alagoano. 2017. Disponível em: http://www.portuguesalagoano.com.br/. Acesso em: 04 abr 2026.

OUSHIRO, L. Projeto Processos de Acomodação Dialetal na Fala de Nordestinos Residentes em São Paulo (FAPESP 2016/04960-7). 2016. Ms.

OUSHIRO, L. silac. 2018. Disponível em < https://oushiro.shinyapps.io/silac>. Online app. Acesso em: 04 abr 2026.

OUSHIRO, L. Questões e métodos: vogais médias pretônicas na fala de migrantes nordestinos em situação de contato dialetal. *In*: VIEIRA, M. S. M.; WIEDEMER, M. L. (Ed.). *Dimensões e experiências em Sociolinguística*. São Paulo: Blucher, 2019a. p. 157–187.

OUSHIRO, L. Linguistic uniformity in the speech of Brazilian internal migrants in a dialect contact situation. *In*: CALHOUN, S.; ESCUDERO, P.; TABAIN, M.; WARREN, P. (Org.) *Proceedings of the 19th International Congress of Phonetic Sciences*, Melbourne, Austrália, p. 686–690, 2019b.

OUSHIRO, L. O estudo da fala de migrantes internos: desafios, procedimentos e resultados do Projeto Acomodação. *In*: BRANDÃO, S. F. B.; VIEIRA, S. R. (Ed.). *Para o estudo comparativo de variedades do Português*: questões teórico-metodológicas e análises de dados. Berlin/Boston: Mouton de Gruyter, 2023. p. 171–196.

OUSHIRO, L. Projeto Coesão e Dispersão: Análise sociofonética da variação idioletal em situação de contato entre dialetos (FAPESP 2023/00968-7). 2023. Ms.

OUSHIRO, L.; SILVEIRA, G. C. P.; SOUZA, E. S.; FERRAZ, L.; MASSUCI, I.; ZHU, R.; BARBOSA, S. P.; OLIVEIRA, A. A.; FIGUEREIDO, J. G. S.. Estudos sociolinguísticos sobre contato dialetal: contribuições do VARIEM e agenda de pesquisa. *Cadernos de Estudos Linguísticos*, v. 65, e023021. p. 1–18, 2023.

PEREIRA, R. C. M. *As vogais médias pretônicas na fala pessoense urbana*. Dissertação (Mestrado em Linguística). Centro de Ciências Humanas, Letras e Artes, Universidade Federal da Paraíba, João Pessoa, 1997.

PRESTON, D. R.; NIEDZIELSKI, N. (Ed.). *A Reader in Sociophonetics*. New York: Walter de Gruyter, 2010.

RIEBOLD, J. Vowel Analyzer, 2013. Script do Praat, versão original de 2013. Disponível em: https://github.com/jmriebold/Praat-Tools/blob/master/Vowel-Analyzer.praat. Acesso em: 04 abr 2026.

ROSENFELDER, I.; FRUEHWALD, J; EVANINI, K.; SEYFARTH, S.; GORMAN, K.; PRICHARD, H.; YUAN, J. FAVE (Forced Alignment and Vowel Extraction) 1.1.3. Zenodo, 2014. Disponível em: https://doi.org/10.5281/zenodo.9846.

SIEGEL, J. *Second dialect acquisition*. Cambridge: Cambridge University Press, 2010.

SILVA, F. G. da. *Alagoanos em São Paulo e a concordância nominal de número*. Dissertação (Mestrado em Linguística) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2014.

SILVEIRA, G. C. P. *The prosody of speech in a dialect contact situation: a sociophonetic study of the speech of Alagoan migrants in São Paulo*. Dissertação (Mestrado em Linguística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 2022.

SILVEIRA, G. C. P.; OUSHIRO, L. Sociofonética. *In*: *Speech Science Entries*, 2022. Speech Prosody Studies Group. Disponível em: https://gepf.falar.org/entries/45. Acesso em: 04 abr 2026.

SLOETJES, H.; WITTENBURG, P. Annotation by category: ELAN and ISO DCR. *In*: *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*. Marrakech, Morocco: European Language Resources Association (ELRA), 2008. Disponível em: http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf. Acesso em: 04 abr 2026.

THOMAS, E. R. Sociophonetics. *In*: CHAMBERS, J. K.; SCHILLING, N. (Ed.). *The Handbook of Language Variation and Change*. Malden, MA: Wiley-Blackwell, 2013. p. 108–127.

THOMAS, E. R. Innovations in sociophonetics. *In*: KATZ, W. F.; ASSMANN, P. F. (Ed.). *The Routledge Handbook of Phonetics*. New York/London: Routledge, 2019. p. 448–472.

ZIMMAN, L. Sociophonetics. *In*: STANLAW, J. (Ed.). *The International Encyclopedia of Linguistic Anthropology*. Malden, MA: Wiley-Blackwell, 2020. p. 1–5.