

CORPUS DE MIGRAÇÃO:
**Um estudo da variação dos segmentos [t, d] e [s] na fala
espontânea de Teixeira de Freitas/BA**

*Crysna Bonjardim da Silva Carmo*¹

*Daiana Chaves Lopes*²

*Maria de Fátima Almeida Gonçalves*³

RESUMO

Este estudo tem como objetivo investigar a variação [t, d] e [s] na fala espontânea do município de Teixeira de Freitas/BA, especificamente na fala de um sujeito migrante. Nesse contexto, busca identificar a variação destacada, descrever as variáveis dependentes e independentes que a contingenciam; e, por fim, discutir os impactos do fenômeno destacado na constituição da identidade da fala espontânea desse sujeito na cidade de Teixeira de Freitas - logo do Território de Identidade Extremo Sul da Bahia. Para realizar este estudo, do ponto de vista teórico, recortamos *A Language into Act Theory* (LAcT - CRESTI, 2000; RASO, 2012) e a Teoria Sociolinguística (ALKMIM, 2001; GÖRSKI. E.M.; COELHO, I; et al, 2010, COSTA, 1960, TARALLO, 1960). Já do ponto de vista metodológico, adotamos os pressupostos da Linguística de *Corpus* (BERBER SARDINHA, 2004). Os resultados demonstram que: (1) as variáveis [t, d e s] sofrem flutuação no *corpus* observado: ora ocorrem palatalizadas, ora ocorrem não palatalizadas. No *corpus*, a percentagem é maior para as variantes [t - tia], [d - dia] e [ʃ: buʃka] em oposição a ocorrência das variantes concorrentes [tʃ: tʃia], [dʒ: dʒia] e [s - 'buska]; e (2) a despeito da distância da comunidade de fala de origem, em tempo e em quilômetros, o sujeito não eliminou traços marcantes de sua língua materna. Por outro lado, os 42 anos de convivência intensa na nova comunidade de fala majoritária - Teixeira de Freitas/BA - afetou em alguma medida o seu idioleto.

Palavras-chave: Variáveis [t, d e s].
Linguística de *Corpus*. Sociolinguística.

¹ Doutora pelo Programa de Pós-Graduação em Estudos Linguísticos Universidade Federal de Minas Gerais (UFMG). Mestre em Letras/Linguística pela Universidade Federal de Juiz de Fora (UFJF). Professora Assistente da Universidade do Estado da Bahia (UNEB). E-mail: <crysnabonjardimsc@gmail.com>

²Graduanda do curso de Licenciatura em Letras - Português e Literaturas de Língua Portuguesa do VIII semestre. Email: <wdaypb87@gmail.com>

³Graduanda do curso de Licenciatura em Letras - Português e Literaturas de Língua Portuguesa do VIII semestre. Email: <fatiminhaasg@hotmail.com>

ABSTRACT

This study aims to investigate the variation [t, d and s] in spontaneous speech in the municipality of Teixeira de Freitas / BA, specifically in the speech of a migrant subject. In this context, it seeks to identify the highlighted variation, describe the dependent and independent variables that contend it; and, finally, to discuss the impacts of the phenomenon highlighted in the constitution of this subject's spontaneous speech identity in the city of Teixeira de Freitas - soon of the *Southernmost Identity Territory of Bahia*. To carry out this study, from a theoretical point of view, we cut out: the *Language into Act Theory* (LAcT - CRESTI, 2000; RASO, 2012) and the Sociolinguistic Theory (ALKMIM., 2001; GÖRSKI. EM; COELHO, I; et al, 2010, COSTA, 1960, TARALLO, 1960). From the methodological point of view, we have adopted the assumptions of the *Corpus Linguistics* (BERBER SARDINHA, 2004). The results demonstrate that: (1) the variables [t, d and s] fluctuate in the observed corpus: sometimes they occur palatalized, sometimes they occur non-palatalized. In the *corpus*, the percentage is higher for the variants [t - tia], [d - dia] and [ʃ: bu]ka] as opposed to the occurrence of the competing variants [tʃ: tʃia], [dʒ: dʒia] and [s - 'buska]; and (2) despite the distance of the speech community of origin, in time and in kilometers, the subject did not eliminate the striking features of his mother tongue. On the other hand, the 42 years of intense coexistence in the new majority-speaking community - Teixeira de Freitas/BA - affected to some extent its idiolet.

Keywords: Variables [t, d and s].
Corpus Linguistics. Sociolinguistics.

1. Introdução

Neste estudo, buscamos responder ao seguinte questionamento: até que ponto as marcas dialetais da comunidade de origem influenciam o comportamento linguístico do sujeito na comunidade de fala atual - logo, comunidade majoritária? Nesse contexto, objetivamos investigar a variação dos segmentos [t, d] e [s] na fala espontânea do município de Teixeira de Freitas/BA, especificamente na fala de um sujeito migrante. Vale ressaltar que este estudo inscreve-se dentro de um projeto maior, qual seja: o “Mapa Linguístico da Fala Espontânea do Extremo Sul da Bahia”, que visa identificar os traços linguísticos que configuram a identidade linguística do Território de Identidade Extremo Sul da Bahia, que envolve os municípios de Alcobaça, Caravelas, Ibirapuã, Itamaraju, Itanhém, Jucuruçu, Lajedão, Medeiros Neto, Mucuri, Nova Viçosa, Prado, Teixeira de Freitas (CARMO, 2020). Para realizar este estudo, teoricamente, recortamos a *Language into Act Theory* (LAcT - CRESTI, 2000; RASO, 2012) e a Teoria Sociolinguística (ALKMIM., 2001; GÖRSKI. E.M.; COELHO, I; et al, 2010, COSTA, 1960, TARALLO, 1960). E, metodologicamente, adotamos: a (i) Linguística de *Corpus* (BERBER SARDINHA, 2004) - para auxiliar nessa tarefa, a ferramenta selecionada é o *AntConc* (LAWRENCE ANTHONY, 2017); e (ii) a

Sociolinguística (TARALLO, 1960), que oferece o instrumental para a análise do fenômeno em foco. Tais recortes teóricos e metodológicos, apresentamos a seguir, bem como os resultados deste estudo.

2. Arcabouço Teórico

2.1 *A Language into Theory: uma abordagem sobre a fala espontânea*

A *Language into Act Theory* (LAcT - CRESTI,2000; RASO, 2012) é uma teoria pragmática, pois entende a fala humana como resultado da atividade comunicativa do falante, ou seja, como uma forma de ação. Desenvolvida a partir dos anos oitenta por Emanuela Cresti (com a contribuição de Massimo Moneglia), no Laboratorio di Linguistica da Universidade de Florença (LABLITA), a LAcT parte de pesquisas experimentais baseadas em *corpora* de fala espontânea. Nesse contexto, assume-se como extensão da **Teoria dos Atos de Fala** (AUSTIN, 1962).

Todavia, diferente da teoria proposta por John L. Austin, cujo objeto são os **atos de fala**, os quais foram observados a partir de dados da modalidade escrita da língua, a LAcT toma como campo de investigação a **fala espontânea** - aquela que é executada ao mesmo tempo em que é planejada (NENCIONI, 1983); e recorta como objeto o enunciado – a menor unidade linguística que possui autonomia pragmática e interpretabilidade em isolamento (CRESTI, 2000). Além disso, há uma outra diferença fundamental entre essas duas visões dentro da perspectiva pragmática, a qual diz respeito às dimensões que compõem o ato de fala, ou seja, toda ação que é realizada por meio do dizer: prometer, pedir, asseverar, discordar, etc. Antes cabe lembrar que os atos de fala possuem três dimensões distintas a um só tempo:

- (1) o **ato locucionário** – a produção linguística em si, a qual envolve:
 - (i) os sons da fala através de ondas sonoras; (ii) os símbolos gráficos codificados em algum suporte; e (iii) os sinais visuais acordados escritos/digitados num espaço determinado;
- (2) o **ato ilocucionário** – a ação produzida ao dizer - um pedido, uma promessa, uma ordem; e, por fim,
- (3) o **ato perlocucionário** – o efeito da locução sobre o interlocutor

Assim, a primeira diferença entre a L-AcT e teoria clássica dos Atos de Fala diz respeito à compreensão do ato perlocucionário. Para a L-AcT, este ato não pode ser definido nos termos “do efeito do ato de fala sobre o interlocutor”, pois esta resposta pode ou não ocorrer. Contudo, o

impulso afetivo (*affective*) em direção ao interlocutor tem sua origem objetiva no locutor - haja vista a sua intenção que ao realizar o ato, não pode ser negada, uma vez que ela dispara a própria locução – “dar parabéns” é um ato de fala. Nesses termos, podemos dizer que a L-AcT é uma teoria sobre a produção do ato de fala, centrada única e exclusivamente no locutor. Já a segunda diferença diz respeito ao fato de que a L-AcT assume que é a **prosódia** o componente linguístico responsável pela constituição, identificação e os tipos ilocucionários dos atos de fala. Para além disso, é a prosódia o componente linguístico responsável pela própria estruturação informacional da fala. Nesse contexto, a função da prosódia é maior do que simplesmente ocupar-se da correta emissão de palavras quanto à posição da sílaba tônica na norma culta da língua ou explicar a **silabada**⁴ (BECHARA, 2009). Na realidade, a prosódia é o componente linguístico que “encapsula” os outros componentes da língua, ou seja, o fonético, o morfológico, o sintático, o semântico-pragmático e discursivo.

2.1.1 O Enunciado: a unidade de referência da fala

Para a L-AcT (CRESTI, 2000), o enunciado é a unidade de referência da fala espontânea, o qual é compreendido como a menor unidade linguística que possui autonomia pragmática e interpretabilidade em isolamento e que, por sua vez, é delimitado por quebras prosódicas, percebidas no momento da fala. Nesse contexto, o enunciado pode apresentar um conteúdo locutivo muito diverso. Todavia, o que vai delimitar sua extensão é a percepção da quebra prosódica que apresenta valor conclusivo - a qual, na transcrição, é identificada pela barra dupla (// = quebra terminal) no fim do enunciado. No entanto, o enunciado pode apresentar quebras prosódicas internas, as quais são percebidas como não conclusivas, mas que estruturam internamente o enunciado. Esse tipo de quebra é identificada na transcrição pela barra simples (/ = quebra não terminal) no interior do enunciado. Dessa forma, o enunciado que apresenta mais de uma quebra prosódica (/) é definido como enunciado complexo. Ao passo que aquele que só apresenta uma quebra prosódica, obrigatoriamente a quebra terminal (//), é chamado de enunciado simples.

⁴ Vocábulos cujo acento prosódico é deslocado, ocasionando um erro prosódico.

Para efeito de demonstração, tomemos a sequência em (1), na qual temos uma declaração na modalidade escrita. Essa declaração na fala espontânea poderia apresentar-se como um enunciado simples (1a), pois apresenta apenas uma quebra terminal (/) no final do enunciado, na qual, pragmaticamente, temos uma declaração. Já em (1b), temos um enunciado complexo, no qual temos uma estrutura de Tópido-Comentário, cuja estrutura interna apresenta duas quebras internas não terminais (/) e uma quebra terminal (/).

(1) Galileu pode ter criado suas próprias teorias sobre óptica no passado.

(1a) Galileu pode ter criado suas próprias teorias sobre óptica no passado //

(2b) Galileu / pode ter criado suas próprias teorias sobre óptica / no passado //

Vale ressaltar que essas quebras internas no enunciado constituem a sua estrutura informacional. Cabe ressaltar que esta estrutura informacional da fala configura uma “gramática” própria nos termos das chamadas **unidades informacionais** (UI). Criadas pelas quebras prosódicas, as UIs apresentam ordem e funções próprias. Dentre essas, destaca-se a UI de **Comentário** (COM). Por ser a unidade raiz, o COM carrega a **força ilocucionária** do enunciado, i.e., a ação da fala. Nesses termos, o enunciado realiza o ato de fala, uma vez que relaciona o domínio das ações à unidade linguística.

2.2 A Teoria Sociolinguística: uma abordagem social da língua

Em linhas gerais, podemos afirmar que o objeto da Sociolinguística é a língua falada, observada, descrita e analisada em seu contexto social, isto é, em situações reais de uso. Dessa forma, o seu ponto de partida será sempre uma comunidade linguística, a qual pode ser definida, grosso modo, como um conjunto de normas que envolve os usos linguísticos de um dado grupo de indivíduos. Contudo, uma comunidade de fala pode ser constituída a partir de uma cidade como Salvador; ou um povo, como os Pataxó, que vive em Barra Velha (Porto Seguro/BA), os estudantes de Letras da UNEB; as mulheres que integram o Coletivo Feminista Diva Guimarães ou, ainda, os migrantes do Território de Identidade Semiárido Nordeste II. Esses últimos grupos dentro de Teixeira de Freitas, município localizado no Território de Identidade Extremo Sul da Bahia. Em outras palavras, a comunidade de fala a ser

Religião, Língua e Literatura

delimitada e descrita dependerá dos objetivos e da natureza da pesquisa a que se quer implementar.

Nesse sentido, cabe lembrar que a língua falada não é homogênea, mesmo dentro das comunidades de fala. De acordo com Costa (1960), cada uma dessas comunidades é marcada por um conjunto de **variações linguísticas**, as quais cobrem categorias factuais como: idade, sexo, escolaridade, classe econômica e pelas experiências históricas, culturais, políticas e sociais que cada falante viveu. Isto é, a heterogeneidade que caracteriza a língua reflete a complexidade que marca as sociedades humanas nos termos da estratificação que marca as relações entre os falantes e as práticas sociais nas quais esses se encontram.

Dentro desse conjunto, a Sociolinguística identifica pelos menos três grandes categorias (ALKMIM, 2007; GÖRSKI; COELHO; Et al, 2010): **a variação diatópica** (geográfica ou regional): trata-se das diferenças linguísticas relacionadas ao espaço físico, como zona rural e urbana, estados, regiões, países, etc; **a variação diastrática** (social): trata-se dos traços de identidade dos falantes e da organização sociocultural da comunidade de fala na qual esses se encontram. Nesse contexto, tais traços são atravessados por fatores como escolaridade, nível socioeconômico, sexo/gênero e faixa etária; e, por fim, **a variação diafásica** (estilística ou situacional): trata-se das variações que ocorrem conforme os papéis sociais desempenhados pelos falantes, os quais vão se alterando de acordo com as situações comunicativas, envolvendo o uso formal e informal da língua.

Todavia, há ainda dois tipos de variação que atravessam os estudos sociolinguísticos: **a variação diamésica** (meio ou código): trata-se das diferenças entre a língua falada e a língua escrita; e a **variação histórica** (diacrônica): trata-se daquelas que ocorrem de acordo com as diferentes épocas vividas pelos falantes, a exemplo da variação que separa o português arcaico do português moderno, bem como a identificação das diversas palavras que caíram em desuso ou daquelas palavras novas que surgiram, haja vista as novas necessidades sociais. Para fazer esse tipo de investigação é preciso considerar o período histórico, algumas só poderão ser realizadas via diamesia escrita. Nesse sentido, a pesquisa se afasta do objeto primordial da Sociolinguística que é a língua falada espontânea - foco deste estudo.

3. Procedimentos metodológicos

Este estudo busca responder ao seguinte questionamento: até que ponto, as marcas dialetais de sua comunidade de origem influenciam o comportamento linguístico do falante na comunidade de fala atual - logo, majoritária? Nesse recorte, objetiva investigar a variação do [t, d] e [s] na fala espontânea do município de Teixeira de Freitas/BA, especificamente na fala de um sujeito migrante. Para tanto, adota uma abordagem empírica de pesquisa. Nesse quadro, adota as premissas da Linguística de *Corpus* (BERBER SARDINHA, 2004) e alguns parâmetros da Sociolinguística (TARALLO, 1960).

3.1 As premissas da Linguística de *Corpus*

Conforme Sardinha (2004), a Linguística de *Corpus* (LC) ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Nesse contexto, utiliza o computador para explorar e fazer a extração desses dados. A LC é marcada pela empiria (os dados são provenientes da observação da experiência objetiva) e pela premissa de que a linguagem é um sistema probabilístico. Em outros termos, embora traços linguísticos sejam possíveis teoricamente, eles não ocorrem com a mesma frequência e não são aleatórias: há uma correlação entre os traços linguísticos e o contexto de uso. Nesse quadro, *corpus* (ou *corpora*) é definido pela existência de uma coletânea de dados linguísticos naturais, legíveis por computador, diferente de outras coletâneas que não são manipuladas eletronicamente - segundo o autor.

Contudo, há diversos tipos de *corpus*. Em todos alguns elementos são obrigatórios: os textos do *corpus* precisam ser autênticos, isto é, produzidos por humanos em linguagem natural (falado ou escrito); a seleção dos dados precisa obedecer a característica do *corpus* a ser compilado; o *corpus* precisa ser representativo, isto é, no mínimo, registrar a linguagem natural utilizada por falantes e escritores da língua em situações reais, permitindo assim a análise de traços identitários do

Religião, Língua e Literatura

contexto pesquisado; o *corpus* precisa ser processado por computador⁵. Posto isso, o *corpus* que serve a este estudo, definido como *Corpus de Migração*, pode ser caracterizado nos seguintes termos, no Quadro 1:

Quadro 1: Critérios de compilação do *corpus*

Critérios definidores do <i>Corpus de Migração</i>		
Modo	falado	porções de fala transcritas
Tempo	contemporâneo	período de tempo corrente
Seleção	de amostragem [<i>sample corpus</i>]	composto por porções de textos ou de variedades textuais, planejado para ser uma amostra finita da linguagem como um todo
Conteúdo	regional ou dialetal	textos provenientes de uma ou mais variedades sociolinguísticas específicas
Finalidade	de estudo	o <i>corpus</i> que se pretende descrever

Como já dito, um *corpus* para a LC precisa ser manipulado por computador obrigatoriamente, assim selecionamos o concordanciador *AntConc*. Desenvolvido por Lawrence Anthony⁶, essa ferramenta está disponível livremente para *download*⁷ e pode ser rodado em programas como *Windows*, *Mac* e *Linux*. O *AntConc* apresenta as seguintes ferramentas: (1) *Word List*: gera uma lista de palavras do *corpus* em ordem alfabética, por tipo e frequência; (2) *Concordance*: gera as linhas de concordância na qual um determinado termo se encontra no *corpus*; (3) *Concordance Plot*: gera um gráfico semelhante a um “código de barras”, que mostra a distribuição do termo pesquisado no *corpus* ou em um de seus arquivos; (4) *File View*: localiza as diferentes ocorrências do termo no *corpus* ou em um de seus arquivos; (5) *Clusters*: gera uma lista do termo destacado em ordem: alfabética, de frequência, de terminações e de probabilidade; (6) *Collocates*: gera uma lista das outras palavras próximas ao termo destacado; e (7) *Keyword List*: gera uma lista de palavras-chave com vistas a comparação entre a frequência das palavras do *corpus* de estudo com a frequência das palavras do *corpus* de referência (LAWRENCE ANTHONY, 2017).

⁵ Para mais detalhes, ver Sardinha (2000).

⁶ <<http://www.laurenceanthony.net>>

⁷ <<http://www.antlab.sci.waseda.ac.jp/software.htm>>

3.2 Parâmetros da Sociolinguística na compilação do corpus

Para selecionar o informante do qual extraímos os dados de fala para a compilação do *corpus* de estudo, guiamo-nos pela obra **A pesquisa sociolinguística** (TARALLO, 1985), na qual o autor estabelece as seguintes etapas para a realização da pesquisa: (1) seleção do informante – considerando a representatividade desse dentro da comunidade linguística, tendo em vista o universo da amostra e o tamanho e estratificação da amostra; (2) metodologia de coleta de dados: usa-se a gravação⁸ de entrevistas individuais com vista a obtenção de narrativas, estimuladas nos termos da entrevista sociolinguística, cuja finalidade é a composição de um banco de dados para estudos sociolinguísticos; (3) o envelope de variação: refere-se a descrição detalhada de uma variável, de suas variantes e dos contextos em que elas podem ou não ocorrer, ou seja, de como exatamente um fenômeno de variação está se manifestando na língua; (4) levantamento de questões e hipóteses; e (5) codificação de dados e análise estatística.

3.2.1 Etapas de compilação do corpus de estudo

Para a compilação do *corpus* de estudo, esta pesquisa cumpre as seguintes etapas: (1) seleção do objeto de estudo (comunidade linguística, grupo social, variação linguística), bem como definição e estudo do referencial teórico orientador da pesquisa; (2) entrevista com o informante cujos dados servem para a constituição do *corpus*. Antes da realização da entrevista é importante que o informante assine um termo de consentimento. Vale salientar que essa deve ser em contexto natural informal; (3) transcrição dos dados oriundos da entrevista, a qual obedece ao critério semiortográfico para a transcrição dos textos orais adotados pela LAcT (CRESTI, 2000; RASO, 2012); (4) compilação do *corpus* e o processamento deste no concordanciador com vista a geração de lista de palavras, linhas de concordância nas quais estão localizadas o fenômeno linguístico em observação e, por fim, o contexto original do fenômeno; e, por fim, (5) produção de relatório dos dados obtidos, tendo em vista a sua análise.

⁸ O gravador selecionado foi *Sony Digital Voice Recorder 4GB - ICD-PX240* (gravador e reproduzidor de voz).

4. Corpus de migração: resultados e discussão

4.1 Perfil sociolinguístico do sujeito da pesquisa

Para realizar este estudo, delimitamos um *grupo social* formado por sujeitos migrantes que também faz parte da comunidade linguística majoritária de Teixeira de Freitas⁹, localizada no Território de Identidade Extremo Sul do estado. Selecionamos, especificamente um sujeito que migrou de *Ribeira do Pombal*¹⁰, localizado no Território de Identidade Semiárido Nordeste II. Na Tabela 2, podemos observar seus dados sociolinguísticos traduzida em metadados:

Quadro 2: Metadados do Informante

Origem	Ribeira do Pombal – BA
Idade	65 anos
Escolaridade	2º grau completo
Ocupação	comerciante/empresário
Outras informações relevantes	- casado - pai de 3 filhos - mora em Teixeira de Freitas há 42 anos - convive com outros familiares da região de origem que também migraram para Teixeira de Freitas

Como podemos constatar no quadro acima, o sujeito da pesquisa migrou de sua região de origem, pelo menos, há 42 anos. Abaixo, na Figura 1, podemos ver o deslocamento desse sujeito através da linha vermelha no *Mapa dos territórios de identidade da Bahia*¹¹, a qual liga o ponto de origem (o lugar onde nasceu - letra A) ao seu destino final (lugar onde mora - letra B). Já na Figura 2¹², destacada em azul, temos essa distância medida em quilômetros (942 km):

⁹ Para mais detalhes: <<https://cidades.ibge.gov.br/brasil/ba/teixeira-de-freitas/panorama>> e <https://pt.wikipedia.org/wiki/Teixeira_de_Freitas>

¹⁰ Para mais detalhes: <https://pt.wikipedia.org/wiki/Ribeira_do_Pombal>

¹¹ Para mais detalhes: <<http://www.seplan.ba.gov.br/modules/conteudo/conteudo.php?conteudo=17>>

¹² Calculado via Google Maps.

Religião, Língua e Literatura



Figura 1: Mapa dos territórios de identidade da Bahia



Figura 2: Distância em quilômetros (km) entre os Ribeira do Pombal e Teixeira de Freitas

Em outras palavras, podemos ver a distância entre as comunidades linguísticas que influenciam comportamento linguístico do sujeito, fonte para a construção do *corpus* da pesquisa. Na extremidade nordeste do estado, temos uma região na qual os falantes utilizam as variantes [t, d e ʃ], pois encontram-se próximos aos estados de Sergipe, Alagoas e Pernambuco, marcados por essa performance linguística. Ao passo que no extremo sul da Bahia, temos uma região que preferencialmente usa as variantes [tʃ, dʒ e s]¹³, assim como os estados do Espírito Santo e de Minas Gerais que lhe fazem fronteira. Vale ressaltar que o motivo da migração desse sujeito foi o fator econômico, ou seja, a busca por melhores condições de vida que, objetivamente, passa pela qualidade do trabalho, como podemos atestar no excerto (2)-(2a) abaixo, extraído do *Corpus* de Migração:

(2) DAI [3] e qual foi o motivo / que trouxe você / para morar / aqui // [...]

(2a) GEF [10] foi em busca / de / conseguir um meio de sobrevivência porque nossa região era muito difícil / nós passamos dificuldade / o Norte sempre foi sofrido né / e aqui / eu vi facilidade //

Uma vez apresentado o perfil sociolinguístico da fonte, passemos às proporções do *corpus*, nomeado como *Corpus* de Migração.

¹³ Nas cidades praianas, como Prado e Caravelas, é mais comum ouvir [ʃ] no lugar de [s].

4.2 Caracterização do Corpus de Migração

O *Corpus* de Migração foi compilado com a finalidade de estudo, ou seja, para servir a investigação de uma das variações linguísticas que constitui a identidade da comunidade linguística majoritária de Teixeira de Freitas - aqui, especificamente, a variação dos segmentos [t, d e s] no conjunto da fala espontânea do município. Nesse contexto, caracteriza-se como um *corpus* de amostragem, uma vez que compõe-se com uma porção de linguagem falada, resultante de uma entrevista informal, de tamanho finito, na qual uma variedade sociolinguística é destacada. Sendo assim, quanto ao seu conteúdo, pode ser compreendido como um *corpus* dialetal, marcado temporalmente: o ano de 2019.

Quanto aos números brutos, o *Corpus* de Migração resulta de uma entrevista de 1h30m (uma hora e trinta minutos). Contudo, o recorte para a sua compilação foi de 15m39s (quinze minutos e trinta e nove segundos), dado que o fenômeno a ser investigado é de natureza fonológica, logo mais simples para rastrear em uma porção menor de língua falada. Assim, temos esses números na Tabela 4:

Tabela 1: Números do Corpus de Migração

Total	
128	Enunciados
2339	Número de palavras (<i>tokens</i>)
565	Tipos de palavras (<i>types</i>)

O *Corpus* de Migração é composto por 128 enunciados, distribuídos 2.339 palavras que, conforme as necessidades comunicativas, distribuem-se entre os 565 tipos. Por exemplo, a palavra *Teixeira*, uma das 565 palavras do *corpus*, do ponto de vista da frequência ocorre 31 vezes, ocupando a 13^o posição no *rank* de ocorrências. Cabe destacar que o *Corpus* de Migração foi transcrito respeitando o comportamento da prosódia tal como advoga a *Language into Act Theory* (Cf. seção 2.1) que apresenta um conceito radical de fala espontânea no qual o enunciado é a unidade mínima, opondo-se, portanto, à sentença, unidade da língua escrita, definida pela presença de um núcleo verbal e cuja divisão é realizada por sinais de pontuação que simulam as características da língua falada. Aqui, observamos as quebras prosódicas,

independente do conteúdo dentro do enunciado. A seguir, discutimos os dados do *corpus* quanto a variação em estudo.

4.3 *Corpus* de Migração: análise comparativa das variáveis [t, d] e [s]

Nesta seção, identificamos a variação dos segmentos [t, d] e [s] no *Corpus* de Migração, bem como descrevemos as variáveis dependentes e independentes nessa variação. Para tanto, explicitamos os dados quantitativamente no Gráfico 1.

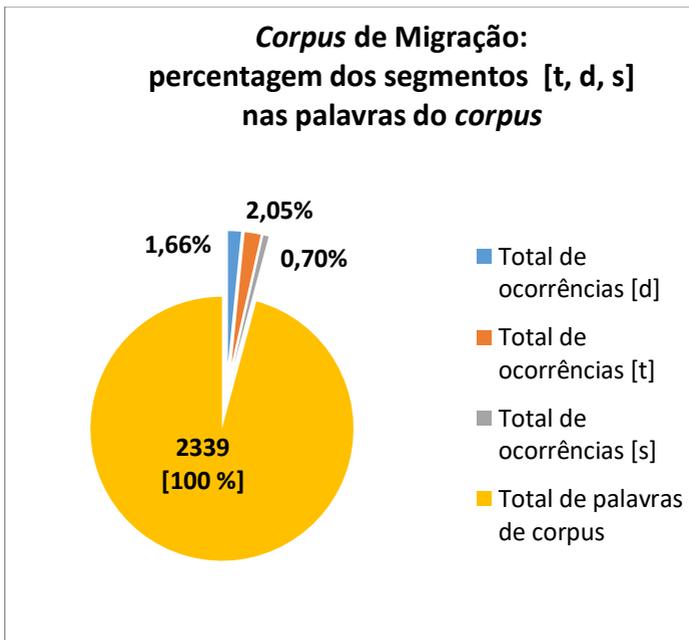


Gráfico 1: Percentagem das ocorrências dos segmentos [t], [d] e [s]

Do montante de 2339 [100%] palavras do *corpus*: 1,66% equivale ao total de ocorrências com o segmento [d]; 2,05% com o segmento [t]; e, apenas, 0,70% com a segmento [s] no final de sílabas dentro do contexto fonológico das palavras. Tais números brutos foram conseguidos via contagem na ferramenta computacional utilizada, o

Religião, Língua e Literatura

AntConc. Posteriormente, transformadas em porcentagem. No entanto, esses números dizem respeito a totalidade das palavras do *corpus*, logo envolve também as intervenções das entrevistadoras. Este estudo, todavia, objetiva analisar apenas o número de ocorrências de [t], [d] e [s] na fala do sujeito da pesquisa. Para resolver esta questão, o áudio do *corpus* foi ouvido, acompanhado pela transcrição. Cada ocorrência presente na performance do sujeito da pesquisa era anotada. Uma vez identificadas as palavras, foi criada tabelas eletrônicas para cada uma. O áudio do *corpus* foi ouvido novamente, com vistas a identificação do enunciado, bem como a variação do segmento correspondente e o seu número de ocorrências. Tal como pode ser visto no Quadro 3 abaixo:

Quadro 3: Exemplificação

Palavra	participei	
Enunciado	[par'ticipei]	[par'tj'icipei]
GEF [50]	-	x
GEF [121]	x	-

De outro modo: a palavra *participei* destacada no Quadro 3 ocorreu duas vezes no *corpus*: a primeira no enunciado GEF[50]: “fomos/inclusive **participei** da emancipação de Teixeira/transformando Teixeira em uma cidade//”; e a segunda no enunciado GEF[121]: “nós fizemo a primeira vaquejada/eu **participei** junto ele/que eu corria vaquejada também//”. Considerando a porcentagem das variações estudadas [t] - 2,05%, [d] -1,66%, [s] -0,70%, analisamos os resultados a nas próximas seções.

4.3.1 Corpus de Migração: porcentagem das variantes [t] e [tʃ]

Conforme a seção 4,3, a variável [t] ocorre em 2,05% das palavras do *corpus*. Entretanto, identificamos uma variação na sua produção. Para efeitos de exemplificação, tomemos a palavra “habitantes” que no *corpus* ocorre duas vezes, respectivamente em GEF[11] e [12]:

GEF[11]: quando aqui cheguei tinha seis mil e poucos **habitantes** não chegava a sete//; e

GEF[12]: hoje Teixeira de Freitas deve tá próximo a cento e oitenta mil **habitantes**//.

A partir do oitavo do *corpus*, constatamos que em GEF[11], temos a ocorrência da variantes oclusiva alveolar desvozeada [t], ao passo que em GEF[12], temos a variante africada alveopalatal desvozeada [tʃ], como apresentadas no Quadro 4 a seguir:

Quadro 4: Variantes [t] e [tʃ]

Enunciado	habitantes	Variante
GEF [11]	[abl'tâtis]	[t] oclusiva alveolar desvozeada
GEF [12]	[abl'tâtʃis]	[tʃ] africada alveopalatal desvozeada

Uma vez identificada essa variação, uma varredura foi realizada no corpus de amostragem. Dessa forma, chegamos a seguinte percentagem: em 64,58% dos casos, temos a variante oclusiva alveolar desvozeada [t] enquanto em 35,42%, a variante africada alveopalatal desvozeada [tʃ]. O peso desse montante pode ser visto no Gráfico 2:

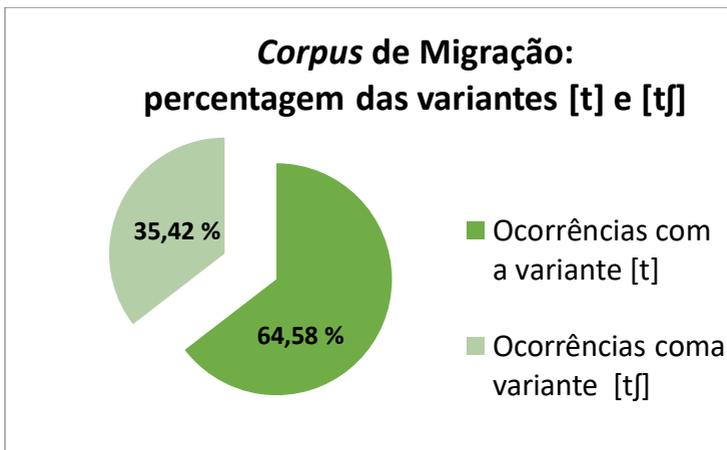


Gráfico 2

4.3.2 Corpus de Migração: percentagem das variantes [d] e [dʒ]

Religião, Língua e Literatura

A variável [d] ocorre em 1,66% das palavras no *corpus* de amostragem (Cf. seção 4.3). Entretanto, há uma variação na sua produção. Para efeitos de exemplificação, tomemos as palavras “diria” e “distrito” no *corpus* que ocorrem, respectivamente, nos enunciados GEF [4] e [47]:

GEF[4]: o motivo foi/eu **diria** que é o segundo Pedro Alvares Cabral / né //,
GEF[47]: Teixeira / era um **distrito** de Alcobaca / impressionante //

Em GEF[4], temos a ocorrência da variante oclusiva alveolar vozeada [d], enquanto que em GEF[47], temos a variante africada alveopalatal vozeada [dʒ], como expostas na Tabela 7 abaixo:

Quadro 5: Variantes [d], e [dʒ]

Enunciado	Habitants	Variante
GEF [4]	[ˈdiria]	[d] oclusiva alveolar vozeada
GEF [47]	[dʒisˈtrito]	[dʒ] africada alveopalatal vozeada [dʒ]

Considerando o *corpus*, chegamos a seguinte percentagem: em 79,48 % dos casos, temos a variante oclusiva alveolar vozeada [d] enquanto em 20,51%, a variante africada alveopalatal vozeada [dʒ]. O peso dessa distribuição pode ser visto no Gráfico 3 seguir:

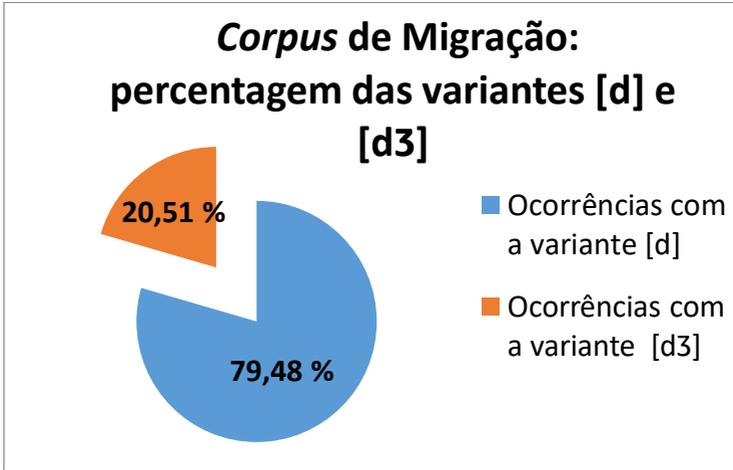


Gráfico 3

4.3.3 Corpus de Migração: percentagem das variantes [s] e [ʃ]

Considerando os dados apresentados na seção 4,3, a variável [s] no final de sílabas dentro do contexto fonológico das palavras ocorre em 0,70% das palavras no *corpus*. No entanto, identificamos variação na sua produção. Para exemplificar, destacamos as palavras “gostei” no enunciado GEF[8] no qual temos a realização da variante [ʃ] fricativa alveopalatal desvozeada; e “busca” em GEF[10] no qual temos a variante [s] fricativa alveolar desvozeada, recortadas no Quadro 6:

GEF[8]: e / resultado eu acabei ficando aqui / *gostei* daqui / me identifiquei com essa cidade aqui / fiz muita amizade e tô aqui até hoje //
 GEF[10]: foi *busca* / de / conseguir / um meio de sobrevivência nossa região era muito difícil / nós passamos dificuldade / o Norte sempre foi sofrido né / e aqui / eu vi facilidade//

Quadro 6: Variantes [s] e [ʃ]

Enunciado	habitantes	Variante
GEF [8]	[ˈgoʃtei]	[ʃ] fricativa alveopalatal desvozeada
GEF [10]	[ˈbuska]	[s] fricativa alveolar desvozeada

Após varredura no *corpus* de amostragem, chegamos a seguinte percentagem: em 75% dos casos, temos a variante fricativa alveopalatal desvozeada [ʃ], ao passo que em 25%, a fricativa alveolar desvozeada [s]. O peso dessa distribuição pode ser visto no Gráfico 4 abaixo:

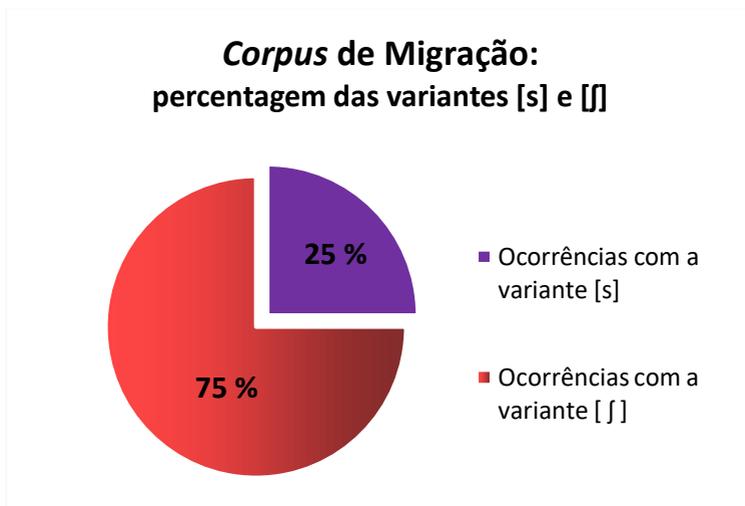


Gráfico 4

4.4 Discussão dos impactos da variáveis [t, d] e [s] no Corpus de Migração

Considerado os dados discriminados nas seções anteriores, podemos constatar que há uma flutuação clara na produção das variáveis [t, d e s] no *Corpus* de Migração - que apresenta um recorte da fala espontânea de um sujeito migrante de Teixeira de Freitas/BA. Essa flutuação pode ser observada na Tabela 9 na qual temos um quadro comparativo das variantes encontradas, as quais correspondem às variáveis destacadas:

Quadro 7: Quadro comparativo das variantes presentes no *Corpus* de Migração

Religião, Língua e Literatura

V ariável	Varia nte 1: %	Varia nte 2: %
t/ /	[t]: 64,58%	[tʃ]: 35,42%
d/ /	[d]: 79,48%	[dʒ]: 20,51%
s/ /	[ʃ]: 75%	[s]: 25%

Como podemos ver, os segmentos que apresentam a maior percentagem de ocorrências no *corpus* diz respeito a três variantes que não são originárias do Extremo Sul da Bahia: [t] com 64,58%, [d] com 79,48% e [ʃ] no final das sílabas dentro do contexto fonológico das palavras com 75%. Conforme Silva (2003), tais variantes são provenientes - ou se concentram - na região Nordeste do Brasil. No *corpus* de estudo, elas disputam a expressão das variáveis [t], [d] e [s] com as variantes [tʃ], [dʒ] e [s] que ocorrem cada com o percentual de 35,42 %, 20,51% e 25% respectivamente. Contudo, vale ressaltar que a variante [s] ocorre nos Estados de Minas Gerais e Espírito Santo (v. Yacovenco e Scherre, 2012), os quais fazem fronteira com o Extremo Sul da Bahia, ao passo que a variante [ʃ] - que ocorre tanto no final de sílabas dentro das palavras quanto no final destas, é usada no estado do Rio de Janeiro¹⁴-capital, baixada fluminense e região serrana. Diante desse contexto, como explicar a flutuação entre as variantes destacadas no *corpus*, tendo em vista a realização das variáveis [t], [d] e [s]?

A primeira coisa que precisamos considerar é a **variável dependente**, isto é aquela que é de natureza linguístico-estrutural. De modo geral, todos os segmentos aqui investigados são influenciadas pelo ambiente ou contexto fonológico¹⁵ no qual se encontram. Como destaca Silva (2003, p. 119), uma das quatro premissas da Fonêmica diz respeito ao fato de que os sons tendem a ser modificados pelo ambiente em que se encontram. Nesse recorte, entram: (a) sons vizinhos – precedentes ou seguintes; (b) fronteiras de sílabas, morfemas,

¹⁴Para mais detalhes: LIMA, O. M.G. O linguajar carioca: Fatores de diferenciação. Linguagem em (Re)vista, Ano 02, N° 02. Niterói, jan./jun. 2005 (pdf)

¹⁵ Considerando o mecanismo básico ou restrições fisiológicas da fala, o **contexto fonológico** diz respeito a interação entre os agrupamentos de segmentos fonéticos que formam palavras, os quais se influenciam mutuamente.

Religião, Língua e Literatura

palavras e sentenças; e (c) a posição do som em relação ao acento; ou seja, são ambientes ou contextos propícios à modificação de segmentos.

No caso das variantes [tʃ] e [dʒ], que se opõem as variantes [t] e [d], os sons vizinhos seguintes as afetam, pois passam pelo processo de **palatalização**. Fisiologicamente, este

consiste no levantamento da língua em direção a parte posterior do palato duro, ou seja, a língua direciona-se para uma posição anterior (mais para a frente da cavidade bucal) do que normalmente ocorre quando se articula um determinado segmento consonantal (SILVA, 2003, p.35).

Nesse contexto, a consoante apresenta um efeito auditivo de sequência de consoante seguida por vogais anteriores [i] e [e], sejam essas orais ou nasais. De acordo com Cristóvão Silva (2003, p. 57), “as oclusivas t/d manifestam-se como africadas alveopalatais tʃ/dʒ quando seguidas da vogal **i** (oral ou nasal). Nestes dialetos temos [tʃi'tfia] para “titia” e [dʒika] para “dica” (SILVA, 1999c). Em outras palavras, a palatalização condiciona a produção das variantes [tʃ] e [dʒ].

Por outro lado, a variante [ʃ] no final das sílabas dentro do contexto fonológico das palavras ocorre por influência de segmentos consonantais desvozeados, ou seja, aqueles produzidos sem a vibração das cordas vocais. No *corpus* de estudo, encontramos ocorrências como festa > [ˈfɛʃta], gostei > [ˈɡoʃtei], história > [hiʃˈtória] e distrito > [diʃˈtrito]. De acordo com Silva (2003, p. 121), “as sibilantes [s,z,ʃ,ʒ] em português quando em posição final de sílaba tendem a ser modificadas pelos contextos”, neste caso, pelo sons vizinhos seguintes.

Todavia, na perspectiva deste estudo, o que influencia a flutuação dessas variáveis no *corpus* de estudo, é a **variável independente**, ou seja, aquela que envolve fatores externos à língua que podem exercer pressão sobre os usos, tais como escolaridade, faixa etária, gênero e origem. Assim sendo, a hipótese deste estudo é a de que essa pressão é exercida principalmente por dois fatores: origem e faixa etária. Para comprovar essa hipótese, observemos os excertos extraídos do *corpus*:

DAI. [1] então xx¹⁶ é / há quanto tempo / você mora aqui em Teixeira//
GEF [2] **quarenta- e- dois anos**//

¹⁶ Substitui o nome do sujeito da pesquisa.

Religião, Língua e Literatura

[...]

DAI [23] e quando você veio pra cá//?

DAI [24] você veio/ sozinho//?

[...]

GEF [26] inicialmente/foi só

[...]

GEF [30] aí depois/conseguir aos poucos/trazer toda a família//

GEF [31] é pra morar aqui//

DAI [32] mas/ dessa época/quando você veio/ quem você deixou/ lá/prá tras assim//

GEF [33] toda minha família/lá ficou/ mãe/pai/irmãos/todos nós/aqui só tinha eu//

GEF [34] eu acho de Teixeira de Freitas/de **Ribeira de Pombal** / aliás / eu tenho a impressão que só tinha eu/aqui na época //

[...]

GEF [68] que nós falamos diferente/eu diria que não mudei nada/ sou Pombalense em voz/ né/ continuo com a mesma/é/ palavreado/do Nortista/né/eu/você e outros mais/né//

[...]

GEF [124] meus filho / tenho quatro filho né/ então aqui também / minha / vida é Teixeira/ cheguei aqui com **dezenove anos** //

Conforme o enunciado GEF[34], o sujeito da pesquisa migrou de sua região de origem, Ribeira do Pombal, localizada no Território de Identidade Semiárido Nordeste II para Teixeira de Freitas, município do Território de Identidade Extremo Sul do estado. Em GEF[124], ele afirma que chegou em Teixeira de Freitas aos 19 anos de idade.

Diante desses fatos, podemos concluir que embora tenha passado a maior parte de sua vida na comunidade linguística de Teixeira de Freitas, o sujeito da pesquisa não eliminou traços marcantes de sua língua materna, ou seja, do dialeto do Território de Identidade Semiárido Nordeste II. Isso se expressa no maior montante das variantes [t], [d] e [ʃ] que - em oposição as variantes [tʃ], [dʒ] e [s], traços do dialeto do Extremo Sul da Bahia - realizam as variáveis [t], [d] e [s] em seu idioleto. Além disso, a que se ressaltar que, embora não seja tão marcado, o padrão prosódico que “envelopa” o conteúdo linguístico de sua fala traz alguns traços fonológicos que marcam a sua origem linguística cuja marca é a da migração, identificando dentro de um grupo linguístico minoritário - sujeitos migrantes.

5. Considerações finais

Diante dos resultados, acreditamos que respondemos a questão inicial: até que ponto as marcas dialetais da comunidade de origem influencia o comportamento linguístico do sujeito na comunidade de fala atual – logo, comunidade linguística majoritária? Ou seja, sim o sujeito da pesquisa conserva traços de sua comunidade de origem. Sobretudo aqueles que dizem respeito ao nível fonológico. Para finalizar, parafraseando Marcos Bagno (1999), lembremos que as línguas são heterogêneas. Tais como um rio corrente, estão sempre em movimento, sempre em transformação, haja vista aspectos como tempo, região e a própria evolução que sofrem naturalmente. Aprendemos uma língua (ou várias), mas não podemos esquecer que esta é apenas um recorte rígido, inflexível e muitas vezes inútil dentro do imenso rio que é o idioma. Dentro deste, há outras várias que chamamos de dialetos ou variedades que, ao mesmo tempo em que se constituem dentro de comunidades linguísticas específicas, entrecruzam-se, transformam-se e transforma os sujeitos que as manifestam. Eis o poder de tudo isso.

REFERÊNCIAS

- ANTHONY, L. *Lawrence Anthony Website (AntConc)*, 2017. Disponível em < <http://www.laurenceanthony.net/software/antconc/>> Acesso em abril de 2020.
- ALKMIM, T. M. Sociolinguística In MUSSALIM, F & BENTES A. C, (ORGS), *Introdução à Linguística* Vol.I. Domínio e Fronteiras. 4 Ed. São Paulo. Cortez. 2004.
- SARDINHA, Tony Berber. Linguística de corpus: histórico e problemática. *DELTA* [on line], vol. 16, n.º 2, p. 323-67, 2000. Disponível em: <www.scielo.br/scielo.php?pid=S0102-44502000000200005&script=sci_abstract&tlng=pt>. Acesso: 23.mai.2020.
- BAGNO, M. *Preconceito Linguístico, o que é, como se faz*. Edições Loyola, São Paulo, Brasil, 1999.
- COSTA. V. L. A. *A importância do conhecimento da variação linguística*. Educar. Curitiba, n.12, p51.1960.
- CAMACHO, R. G. Sociolinguística? Parte II. In: MUSSALIM, F; BENTES, A. C. *Introdução à linguística? Domínios e fronteiras*. 7. Ed. São Paulo: Cortez, 2007.

Religião, Língua e Literatura

- GÖRSKI, E.M.; COELHO, I; et al. *Sociolinguística*. Florianópolis : LLV/CCE/ UFSC, 2010.
- LABOV, W. *Padrões sociolinguísticos*. São Paulo: Parábola, 2008.
- LIMA, O. M. G. *O linguajar carioca*: Fatores de diferenciação. *Linguagem em (Re)vista*, Ano 02, Nº 02. Niterói, jan./jun. 2005 (pdf)
- MARCONI, M.A; LAKATOS, E.M. **Fundamentos de metodologia científica** . 5. ed.- São Paulo : Atlas 2003.
- MARCUSCHI, L. A; DIONISIO, A. P. *Fala e escrita* . 1 ed. Belo Horizonte: Autêntica, 2007. 208 p
- RASO, T. *Fala e escrita*: meio, canal, consequências pragmáticas e linguísticas. *Domínios de Lingu@Gem*, Uberlândia, v. 7, n. 2, p. 12-46, jul./dez. 2013. Disponível em: <<http://www.seer.ufu.br/index.php/dominiosdelinguagem/article/view/23730>>.
- TARALLO, F. *A Pesquisa Sociolinguística*. Ática S.A. 3ª edição. São Paulo, 1990
- VANIN, A. A. Considerações relevantes sobre definições de ‘comunidade de fala’. *Acta Scientiarum Language and Culture*, Maringá, v.31. n.2, p. 147-153, 2009.
- YACOVENCO, L.C.; SCHERRE, M.M.P.et al. Projeto PORTVIX: a fala de Vitória/ES em cena. *Alfa*, São Paulo, vol. 56, n. 3, p. 771-806, 2012. Disponível em:<<https://periodicos.fclar.unesp.br/alfa/article/view/4946/4361>>.