



UFPE

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE ARTES E COMUNICAÇÃO
DEPARTAMENTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS

Meiriany Cristinaide Nascimento Souza Alcântara

**UM ESTUDO SOCIOLINGUÍSTICO DOS REQUISITOS DE APOIO DISCURSIVO
UTILIZADOS POR PESSOAS DO SERTÃO DO PAJEÚ – PERNAMBUCO**

Recife
2018

MEIRIANY CRISTINAIDE NASCIMENTO SOUZA ALCÂNTARA

**UM ESTUDO SOCIOLINGUÍSTICO DOS REQUISITOS DE APOIO DISCURSIVO
UTILIZADOS POR PESSOAS DO SERTÃO DO PAJEÚ – PERNAMBUCO**

Dissertação apresentada ao Programa de Pós-graduação em Letras, da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Mestre em Letras.

Área de Concentração: Linguística.

Orientador: Prof. Dr. Marcelo Amorim Sibaldo

Recife
2018

Catálogo na fonte
Bibliotecária Delane Mendonça de Oliveira Diu CRB4-849/86

A347e Alcântara, Meiriany Cristinaide Nascimento Souza
Um estudo sociolinguístico dos requisitos de apoio discursivo utilizados por pessoas do sertão do Pajeú-Pernambuco / Meiriany Cristinaide Nascimento Souza Alcântara. – Recife, 2018.
176 f.: il.

Orientador: Marcelo Amorim Sibaldo.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. Centro de Artes e Comunicação. Programa de Pós-Graduação em Letras, 2018.
Inclui referências e apêndices.

1. Marcadores conversacionais . 2. Requisitos de apoio discursivo. 3. Sociolinguística. I. Sibaldo, Marcelo Amorim (Orientador). II.Titulo.

410 CDD (22.ed.) UFPE (CAC 2018-222)

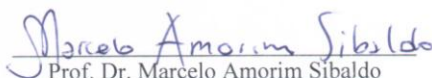
MEIRIANY CRISTINAIDE NASCIMENTO SOUZA ALCÂNTARA

**UM ESTUDO SOCIOLINGUÍSTICO DOS REQUISITOS DE APOIO
DISCURSIVO UTILIZADOS POR PESSOAS DO SERTÃO DO PAJEÚ –
PERNAMBUCO**

Dissertação apresentada ao Programa de Pós-Graduação em Letras da Universidade Federal de Pernambuco como requisito para a obtenção do Grau de Mestre em LINGUÍSTICA.

APROVADA EM 1/10/2018.

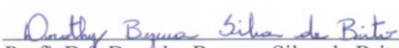
BANCA EXAMINADORA



Prof. Dr. Marcelo Amorim Sibaldo
Orientador – LETRAS - UFPE



Prof. Dr. Adeilson Pinheiro Sedrins
UAST - UFRPE



Profª. Drª. Dorothy Bezerra Silva de Brito
UAST - UFRPE

Esse trabalho é dedicado a Asafe Alcântara. Ele ajudou e atrapalhou ao mesmo tempo. Por ele eu parei e prossegui muitas vezes a escrita deste texto.

AGRADECIMENTOS

Agradeço em primeiro lugar a Deus, acredito que Ele conduziu-me em todo o processo de escrita deste texto, muitas intempéries aconteceram, mas Ele resolveu todas elas. Além disso, trouxe como um presente a vida e a participação de cada uma das pessoas que citarei aqui. Valeu mesmo, Deus!

Para a realização deste estudo uma pessoa foi essencial: meu orientador, prof. Dr. Marcelo Sibaldo. Pensava nele durante o processo de escrita, sabia que tinha que ‘caprichar’ pois ele iria ler o texto, ao mesmo tempo, sabia que se viesse a cometer erros despercebidos por mim, não passariam despercebidos por ele, como é bom ter alguém assim do seu lado. Sibaldo, você acreditou em mim muito mais do que eu mesma, serei eternamente grata, espero um dia honrá-lo!

Sou muita grata pelo *insight* para a realização deste estudo, ele veio a mim por intermédio de uma banca avaliadora em 22 de janeiro de 2015 composta pelos professores Adeilson Pinheiro Sedrins e Dorothy Brito. Lembro das palavras ditas por ele: “não podes parar por aqui”, lembro das palavras escritas por ela no texto: “o seu trabalho é uma árvore frondosa e frutífera dessa imensa floresta dos marcadores”. Talvez eles não saibam, mas estas palavras motivaram a realização deste estudo. Dizer muito obrigada é bem pouco para externar minha gratidão.

Agradeço imensamente a professora Renata Lívia, suas orientações iniciais acerca da Sociolinguística acompanham-me até hoje e parte do que foi dito neste estudo aprendi com você. Muito obrigada, Renata.

Preciso agradecer aos meus professores das disciplinas do curso de mestrado, Medianeira, Evandra, Xavier, Marcelo, Vírginia e Alberto Poza, aprendi bastante com vocês e parte deste aprendizado tentei trazer para este estudo, espero que eu tenha conseguido. Os demais membros do Programa de Pós-graduação do PPGL – Letras – UFPE e os discentes do programa também participaram desta trajetória, direta ou indiretamente vocês me assistiram melhor do que eu mesma esperava, super obrigada!

Externo minha gratidão a CAPES/DS pelo incentivo para a realização deste estudo, sem ele teria sido ainda mais difícil, para não dizer impossível!

Da mesma forma, agradeço aos avaliadores deste estudo, membros titulares e suplentes, ser avaliada por vocês além de imensa responsabilidade é um grande privilégio, obrigada!

Por trás deste time acadêmico, contei com uma equipe que não sabia muito acerca do que era esta pesquisa, mas foram fundamentais para que ela se tornasse realidade. Meu esposo Luiz Carlos, quanto incentivo tive de você, muito obrigada, você é incrível! Obrigada por sonhar este sonho junto comigo, muito obrigada!! Minha mainha Carmelita, que não entende nada disso, mas sempre dizia: “deixa que eu faço isso, você vai fazer o trabalho! Falta quando folhas para terminar?”, Mainha esse trabalho também é seu, viu? Hilquias, cara, sem sua ajuda, como lidaria com os números? Não sei mesmo, obrigadão! Ana e Rany, obrigada por esperar eu terminar o texto e esperar em amor, adoro vocês! Meus irmãos e minhas cunhadas que sempre me perguntavam acerca da escrita, está aqui, parece que consegui, obrigada por se importar com algo tão distante do universo de vocês. Lelinha e Cidinha, minhas irmãs do coração, não tenho como agradecer toda atenção e cuidado comigo nesse processo de escrita, é tão bom ter vocês. Déreck, não tenho mesmo como te agradecer, espero um dia poder te ajudar o tanto que você ajudou-me aqui, valeu mesmo, amiguinho.

Por fim, mas não menos importante, quero agradecer aos informantes deste estudo, sem a participação de cada um de deles não teria análise dos Requisitos de Apoio Discursivo, variantes identificadas, não teria pesquisa! Só fiz este estudo porque vocês aceitaram participar dele, obrigada mesmo!

RESUMO

Esta pesquisa tem como objetivo identificar e analisar os Marcadores Conversacionais, mais especificamente os Requisitos de Apoio Discursivo, uma das nove categorias dos marcadores propostas por Macedo e Silva (1996), utilizados por pessoas da região do Sertão do Pajeú, em Pernambuco, observando se variáveis sociais como gênero, escolaridade, idade e localidade, assim como, se variáveis internas à língua, como a entonação interrogativa, a massa fônica, a classe de palavra que originou e a classe de palavra que antecede o marcador influenciam no uso das variantes destes elementos. Além disso, objetivamos também, observar se, nos dados coletados, há outras variantes para a variável estudada, além das formas já apresentadas em estudos anteriores com outras comunidades de fala. Para tanto, fazemos uma breve discussão acerca dos Marcadores Conversacionais, objetivando expor e tornar mais conhecido o fenômeno que será analisado. Esta discussão é realizada com base nos trabalhos de Marcuschi (1989), Macedo e Silva (1996), Castilho (2004) e (2016), Valle (2001), Freitag (2008) e Risso, Silva e Urbano (2015), além destes estudos, temos como base a pesquisa realizada por Alcântara (2015), tendo em vista que este estudo foi realizado em uma comunidade de prática pertencente à comunidade de fala aqui estudada e, com base nos resultados alcançados, aponta que é possível que haja formas específicas que desempenham a função de Requisitos de Apoio Discursivo na fala dos sertanejos. O presente estudo está baseado no arcabouço teórico-metodológico da Sociolinguística, mais precisamente, da Sociolinguística Variacionista (LABOV, 2008 [1972]), pois esta teoria parte do princípio de que a língua deve ser analisada em situações reais de uso, uma vez que é nesse contexto em que a variação linguística se manifesta. Sendo assim, baseados nos pressupostos teórico-metodológicos da Sociolinguística Variacionista, realizamos a coleta e a análise dos dados de fala. Os dados foram obtidos por meio de entrevista gravadas com 54 os informantes de três municípios do Sertão do Pajeú: Serra Talhada, São José do Egito e Triunfo. Após a coleta de dados, autorizada pelo Comitê de Ética Institucional, realizamos a seleção, a transcrição e a análise dos dados coletados, por meio desta, chegamos aos seguintes resultados: há formas que parecem ser comuns na região do Sertão do Pajeú e que não ainda foram descritas em estudos realizados em outras comunidades, como, por exemplo, as variantes *hein? visse?, tá ligado?, num sabe?, num é? e certo?*, visto que, em análises feitas em outras regiões, estas variantes não foram mencionadas como presentes na fala dos informantes. Da mesma forma, há variantes que parecem ser típicas não apenas da comunidade de fala, mas de localidades

que integram esta comunidade de fala, como, por exemplo, a forma *visse?* em São José do Egito. Além disso, as análises realizadas a partir das variáveis sociais selecionadas, a saber: gênero, localidade, escolaridade e faixa-etária apontaram que alguns destes fatores como, por exemplo, a faixa-etária, pode condicionar o uso dos Requisitos de Apoio Discursivo, pois as pessoas mais jovens parecem que usam menos estes elementos.

Palavras-chave: Marcadores conversacionais. Requisitos de apoio discursivo. Sociolinguística.

ABSTRACT

This research aims to identify and analyze the Conversational Markers, more specifically the Discursive Support Requirements, one of the nine categories of markers proposed by Macedo and Silva (1996), used by people from the Sertão do Pajeú region, in Pernambuco, observing social variables such as gender, schooling, age and locality, as well as whether variables internal to the language such as interrogative intonation, phonic mass, the word class that originated and the word class that precedes the marker influence the use of the variants of these elements. In addition, we also aim to observe if, in the data collected, there are other variants for the studied variable, in addition to the forms already presented in previous studies with other speech communities. To do so, we make a brief discussion about the Conversational Markers, aiming to expose and make known the phenomenon that will be analyzed. This discussion is based on the works of Marcuschi (1989), Macedo e Silva (1996), Castilho (2004) and (2016), Valle (2001), Freitag (2008) and Risso, Silva and Urbano (2015). of these studies, we are based on the research carried out by Alcântara (2015), considering that this study was carried out in a community of practice belonging to the speech community studied here, and, based on the results achieved, points out that it is possible that there are forms specific functions that play the role of Discursive Support Requirements in the speech of the sertanejos. The present study is based on the theoretical-methodological framework of Sociolinguistics, more precisely, on Variationist Sociolinguistics (LABOV, 2008 [1972]), since this theory starts from the principle that the language should be analyzed in real situations of use, since it is in this context that linguistic variation is manifested. Therefore, based on the theoretical-methodological assumptions of Variationist Sociolinguistics, we performed the collection and analysis of speech data. The data were obtained through interviews recorded with 54 informants from three municipalities in the Sertão do Pajeú: Serra Talhada, São José do Egito and Triunfo. After the data collection, authorized by the Institutional Ethics Committee, we carried out the selection, transcription and analysis of the collected data, through which we arrived at the following results: there are forms that seem to be common in the Sertão do Pajeú region and that do not have been described in studies conducted in other communities, such as variants *hein? visse?*, *tá ligado?*, *num sabe?*, *num é?* and *certo?*, since, in analyzes made in other regions, these variants were not mentioned as present in the informants' speech. Likewise, are there variants that seem to be typical not only of the speech community, but of localities that integrate this speech community, such as the

visse form? in São José do Egito. In addition, analyzes based on selected social variables, namely: gender, locality, schooling and age range, pointed out that some of these factors, such as the age range, may condition the use of the Discursive Support Requirements, because younger people seem to use less of these elements.

Keywords: Conversational markers. Requirements discourse support. Sociolinguistics.

LISTA DE FIGURAS

Figura 1:	Exemplo de esquema de análise combinatória dos fatores sociais.....	58
Figura 2:	Mapa das mesorregiões do estado de Pernambuco.....	77
Figura 3:	Mapa das 12 microrregiões do estado de Pernambuco.....	78
Figura 4:	Mapa das cidades que integram a região do Sertão do Pajeú.....	78
Figura 5:	Museu do cangaço em Serra Talhada.....	81
Figura 6:	Cenário ao ar livre para apresentação do espetáculo.....	81
Figura 7:	Apresentação do espetáculo O massacre do Angico.....	82
Figura 8:	Visão parcial da cidade de Triunfo.....	83
Figura 9:	Pico do Papagaio - ponto mais alto do estado de Pernambuco localizado na cidade de Triunfo.....	83
Figura 10:	Apresentação dos Carretas em frente ao teatro Guarani durante o festival de cinema.....	84
Figura 11:	Portal de entrada da cidade de São José do Egito.....	86
Figura 12:	Escultura posta no meio da cidade que enaltece a viola e o cantador Egipsiense.....	87
Figura 13:	Esquema que expõe as variáveis extralinguísticas para seleção dos Informantes.....	89

LISTA DE GRÁFICOS

Gráfico 1:	Porcentagem de uso das formas finais e iniciais de MCs do estudo realizado por Marcuschi (1989).....	26
Gráfico 2:	O percentual de uso das formas de RADs encontradas no corpus.....	123
Gráfico 3:	O uso dos RADs por cidade.....	125
Gráfico 4:	Os RADs em Serra Talhada.....	128
Gráfico 5:	Percentual de uso da variante hein? estratificado por faixa-etária.....	131
Gráfico 6:	Percentuais de uso estratificados pela classe que antecede a variante hein?.....	133
Gráfico 7:	Os RADs em São José do Egito.....	134
Gráfico 8:	A variante visse? estratificada pela faixa-etária.....	139
Gráfico 9:	A variante visse? estratificada pela escolaridade.....	140
Gráfico 10:	A variante visse? estratificada pela classe de palavra que antecede o RAD.....	142
Gráfico 11:	Os RADs em Triunfo.....	143

LISTA DE QUADROS

Quadro 1:	Emprego dos RADs por gênero discursivos.....	30
Quadro 2:	Marcadores Discursivos: funções e colocação do enunciado.....	33
Quadro 3:	Número de realizações e porcentagem de uso das variantes encontradas em cada contexto.....	41
Quadro 4:	Principais distinções entre comunidade de fala, redes sociais e comunidade de prática.....	52
Quadro 5:	Distribuição de informantes do estudo realizado por Valle (2001).....	59
Quadro 6:	Alguns corpora disponíveis para estudo na área.....	61
Quadro 7:	Modelo de teste de percepção para estudo com RADs.....	69
Quadro 8:	Modelo de teste de produção para estudos com RADs.....	69
Quadro 9:	Distribuição da amostra de informantes da pesquisa.....	90
Quadro 10:	Codificação dos dados coletados.....	105
Quadro 11:	Variantes de RADs já estudadas por diferentes pesquisas.....	111
Quadro 12:	Variantes e ocorrências destas no corpus em análise.....	112
Quadro 13:	Ponderação acerca das formas como sendo variantes de uma mesma Variável.....	113
Quadro 14:	Os RADs encontrados na comunidade de fala em análise.....	115
Quadro 15:	Variantes encontradas no corpus analisado por Alcântara (2015) e no presente.....	119
Quadro 16:	Possíveis relações entre as variantes e RADs.....	121
Quadro 17:	Percentual de uso das variantes dos RADs descritas e não descritas em Língua Portuguesa.....	148
Quadro 18:	As formas descritas e não descritas de RADs estratificadas por gênero.....	149
Quadro 19:	Formas descritas e não descritas estratificadas pela faixa-etária.....	151
Quadro 20:	Formas descritas e não descritas estratificadas pela localidade.....	152
Quadro 21:	Formas descritas e não descritas estratificadas pela escolaridade.....	153
Quadro 22:	Formas descritas e não descritas estratificadas pela classe de palavra originária do RAD.....	154
Quadro 23:	Formas descritas e não descritas estratificadas pela massa fônica.....	155
Quadro 24:	Formas descritas e não descritas estratificadas pela classe de palavra que antecede o RAD.....	156

Quadro 25:	Formas descritas e não descritas estratificadas pela entonação	
	Interrogativa.....	157
Quadro 26:	Peso relativo dado para o grupo 1 (gênero).....	159
Quadro 27:	Peso relativo dado para o grupo 3 (Localidade).....	160

SUMÁRIO

1	INTRODUÇÃO.....	17
2	SOBRE OS MARCADORES.....	22
2.1	ALGUNS ESTUDOS.....	22
2.1.1	Marcuschi (1989).....	24
2.1.2	Macedo e Silva (1996).....	28
2.1.3	Castilho (2004) e (2016).....	32
2.1.4	Risso, Silva e Urbano (2015).....	34
2.1.5	Valle (2001).....	37
2.1.6	Freitag (2008).....	39
2.1.7	Alcântara (2015).....	41
2.2	ACONCEPÇÃO DE MARCADOR QUE ASSUMIMOS NESTE TRABALHO.....	43
2.2.1	Os Requisitos de Apoio Discursivo.....	45
3	SOBRE A SOCIOLINGUÍSTICA.....	47
3.1	A SOCIOLINGUÍSTICA VARIACIONISTA.....	47
3.2	OS PROCEDIMENTOS METODOLÓGICOS NA PESQUISA SOCIOLINGUÍSTICA.....	55
3.2.1	Escolha da comunidade de fala.....	55
3.2.2	Seleção de informantes.....	56
3.2.3	A coleta de dados.....	60
3.2.4	Transcrição, seleção, codificação e quantificação dos dados coletados.....	70
4	METODOLOGIA.....	75
4.1	A ESCOLHA DA COMUNIDADE DE FALA.....	75
4.1.1	O sertão do Pajeú.....	77
4.1.1.1	<i>Serra Talhada</i>	79
4.1.1.2	<i>Triunfo</i>	82
4.1.1.3	<i>São José do Egito</i>	85
4.2	A SELEÇÃO DOS INFORMANTES.....	87
4.3	A COLETA DE DADOS.....	91

4.4	A TRANSCRIÇÃO, SELEÇÃO, CODIFICAÇÃO E QUANTIFICAÇÃO DOS DADOS.....	95
4.5	AS VARIÁVEIS E AS HIPÓTESES LEVANTADAS.....	107
5	ANÁLISE DOS DADOS.....	110
5.1	AS VARIANTES ENCONTRADAS NO SERTÃO DO PAJEÚ.....	110
5.2	OS NÚMEROS DE OCORRÊNCIAS DAS VARIANTES.....	121
5.2.1	Serra Talhada.....	127
5.2.1.1	<i>Hein?</i>	130
5.2.2	São José do Egito.....	134
5.2.2.1	<i>Visse?</i>	137
5.2.3	Triunfo.....	143
5.2.3.1	<i>Certo?</i>	145
5.3	OS RESULTADOS DO VARBRUL.....	151
5.3.1	Análises da rodada mais significativa.....	158
6	CONSIDERAÇÕES FINAIS.....	162
	REFERÊNCIAS.....	168
	APÊNDICE A - Roteiro de entrevista.....	173
	APÊNDICE B - Ficha social do informante.....	175

1 INTRODUÇÃO

Os itens linguísticos estudados nesta pesquisa, nomeados de Requisitos de Apoio Discursivo, possuem um alto índice de recorrência na língua, quer seja falada, quer seja escrita. No entanto, tais itens, não estão no inventário da gramática tradicional como uma classe de palavras ou um elemento gramatical da língua portuguesa, além disso, os linguistas, que por sua vez os aceitam como um fenômeno muito recorrente na língua e dedicam-se a estudá-los, não chegaram a uma definição precisa do que veem a ser, a ponto de nomeá-los de forma unânime, resultando, assim, em nomenclaturas distintas que se alternam, comumente, entre marcadores conversacionais ou marcadores discursivos (Cf. RISSO, SILVA E URBANO, 2015).

Apesar de os marcadores¹ não serem contemplados nas gramáticas tradicionais, nem possuírem uma nomeação precisa, eles têm sido objeto de inúmeras pesquisas, tendo em vista sua importância para os estudos linguísticos, pois são elementos presentes na Língua Portuguesa, com, inclusive, um alto número de realizações, sobretudo, na língua falada. Nestas pesquisas², é notória a oscilação da nomenclatura dos marcadores, pois, alguns estudos, como Freitag (2008), Risso, Silva e Urbano (1996), definem estes itens como sendo marcadores discursivos, pois consideram que são itens que estabelecem a interação entre falante e ouvinte não apenas no texto oral. Já estudos como os de Castilho (2004) e (2010) oscilam na nomenclatura dos marcadores, pois ora nomeiam de marcadores conversacionais, ora de marcadores discursivos. Inclusive em Valle (2001), a oscilação do termo se dá dentro de uma mesma pesquisa, ou seja, a autora não reconhece a necessidade de distinguir uma definição da outra para a realização da pesquisa que objetiva reconhecer as diferentes formas de marcadores como variantes de uma mesma variável. Apesar de não justificar o porquê da escolha, Macedo e Silva (1996) definem estes itens como sendo marcadores conversacionais. Vale destacar aqui, que este é o nome mais antigo dado a estes elementos linguísticos na língua portuguesa, utilizado também pelo principal precursor dos estudos com marcadores em uma das obras mais significativas em que este fenômeno é analisado no

¹Utilizaremos o nome “marcador” até apresentarmos uma discussão acerca destas nomenclaturas que nos seja capaz de adotar determinado nome como sendo pertinente para rotular os itens que nos propomos analisar.

² As pesquisas a serem citadas como exemplos da oscilação das nomenclaturas dos marcadores também nos servirão como base teórica para discussões posteriores acerca destes itens, tendo em vista que as problematizações e os resultados alcançados constituem-se em um importante material para estudos com marcadores.

Português Brasileiro, Marcuschi (1989). Além disso, é o termo mais utilizado pelos linguistas brasileiros ao se referirem ao fenômeno que aqui objetivamos analisar.

Por causa das diferentes nomenclaturas, há entre os linguistas inúmeras definições acerca dos marcadores. No entanto, é consenso entre eles que os marcadores são itens ou expressões lexicais, e até mesmo elementos não verbais, como os prosódicos, bastante recorrentes, principalmente na fala, que possuem inúmeras formas, funções e posições no discurso. Estas constatações, feitas com base nos estudos supracitados, evidenciam que estes itens não são “vícios de linguagem”³, mas constituem-se em um recurso que o falante usa, de maneira sistemática, para suprir algumas necessidades no momento em que vai expressar-se.

Há, nos estudos linguísticos dos marcadores de língua portuguesa, importantes divisões dos marcadores, estas realizadas a partir de suas funções, formas e lugar que ocupam na sentença. Castilho (2016) os dividiu a partir de suas funções em dois grupos: i) marcadores pragmáticos ou interpessoais e marcadores textuais ou ideacionais. Já Macedo e Silva (1996) dividiram os marcadores em nove categorias, nomeadas e exemplificadas abaixo⁴, a partir de dados retirados do *corpus* montado para ser analisado no presente estudo⁵.

- (1) - **Iniciadores:** “*bom*⁶, eu vou dizer, primeiramente, eu atribuo ao criador” (Inf18 – SJE)⁷.
- (2) - **Requisitos de Apoio Discursivo:** “Para mim, é ter mais médico, *né?*” (Inf01-SJE).
- (3) - **Redutores:** “Não arrumar a casa, desobedecer e *por aí vai*” (Inf47-ST).
- (4) - **Esclarecedores:** “*E é assim*, a questão de trabalho é muito boa” (Inf04 – SJE).
- (5) - **Preenchedores de pausa:** “Os problemas são... *ééé... hummm...* não tem área de lazer” (Inf14 – SJE).
- (6) - **Sequenciadores:** “*Aí* não posso esquecer meu professor Ferrari” (Inf36 – TR).
- (7) - **Resumidores:** “Eu queria que meu filho fosse diferente *e tal*, essas coisas” (Inf46 – ST).

³“Vícios de linguagem” ou “cacoetes linguísticos” são nomenclaturas utilizadas, segundo estudo de Freitag (2007), para nomear, de maneira estigmatizada, os marcadores.

⁴ Os termos dado pelas autoras serão definidos posteriormente no capítulo 1 desta dissertação.

⁵ Vale ressaltar que as sentenças que servem como exemplos de cada categoria foram retirados do *corpus* da presente pesquisa, mas as nove categorias elencadas foram divididas, nomeadas e classificadas pelas pesquisadoras Macedo e Silva (1996).

⁶ Os elementos considerados marcadores estão destacados, em itálico, nas sentenças enumeradas de 1 a 9.

⁷ Esta sigla refere-se às informações do informante em nosso *corpus*, sendo composta pelo número do informante no *corpus* (Inf+número) e a localidade onde reside o informante, a saber, ST para Serra Talhada, TR para Triunfo e SJE para São José do Egito.

(8) - **Argumentadores:** “*Olha só, ééé... em todas as cidades existem problemas*”(Inf33 – TR).

(9) - **Finalizadores:** “Um dos grandes problemas é isso, *isso aí*” (Inf23 – TR).

É importante salientar que essa é a divisão mais recorrente nos estudos linguísticos uma vez que contemplam três importantes aspectos dos marcadores, a saber, as formas, as funções e as posições na sentença. Essas divisões foram, segundo as autoras, motivadas por fatores como o grande número de marcadores em língua portuguesa e pelo fato de os marcadores ocuparem diferentes posições na sintaxe da sentença. No que diz respeito à quantidade de marcadores, é pertinente destacar que esse número tem sido cada vez maior quando levamos em consideração que as interações entre falante e ouvinte parecem propiciar não apenas a utilização destes itens, mas o surgimento de novas formas, que desempenham as diferentes funções de marcador exemplificadas acima.

É certo que, devido ao extenso número de marcadores no Português Brasileiro, os pesquisadores têm adotado a divisão abordada acima como meio de delimitar o objeto de análise em seus estudos, tendo em vista que analisar todos estes elementos em suas inúmeras formas e funções seria um trabalho que, de tão exaustivo, poderia ter seus resultados comprometidos. Sendo assim, as pesquisas com marcadores, inclusive as citadas anteriormente, têm tido como foco, por exemplo, a análise dos marcadores que exercem a função de sequenciar o discurso, ou seja, os sequenciadores. No entanto, pesquisas como Valle (2001) e Freitag (2008), aportadas na perspectiva sociolinguística variacionista, defendem que as diferentes funções destes elementos constituem-se em uma variável linguística em que cada forma desta variável vem a ser uma variante, uma vez que possuem equivalência de significados.

Sendo assim, tais estudos, realizados com a variável *requisitos de apoio discursivo* (cf. (2) acima), apresentam, como resultado alcançado e, ao mesmo tempo, como sugestão de ampliações de investigações futuras, o seguinte dado: “uma variedade linguística da região sul do Brasil, difere-se em alguns aspectos de uma variedade da região nordeste” (FREITAG, 2008, p.11). Este importante dado, também foi observado por Alcântara (2015) que, ao analisar o uso dos requisitos de apoio discursivo em dois contextos de diferentes níveis de formalidade, objetivando mapear se os marcadores são utilizados em contextos mais formais, observou que na fala de radialistas do sertão pernambucano há quatro formas, a saber, (i) *tá*

certo?, (ii) *tá bom?*, (iii) *hein?* e (iv) *ok?*, que funcionam como variantes da variável *requisito de apoio discursivo*⁸, em realizações como:

- (10) Vale a pena você conferir, *tá certo?* (Inf17 – ST)⁹
- (11) Essa música é linda, *hein?* (Inf14- TR)
- (12) A festa será a partir das oito horas da noite, *ok?* (Inf06 – FL)
- (13) Liga pra mim, *tá bom?*(Inf25 – SJE)

Partindo destas considerações, a presente pesquisa objetivadesenvolver um estudo preciso dos *requisitos de apoio discursivo* e das variantes utilizadas por falantesdo Sertão do Pajeú, com o intuito de averiguar se há, nesta região, variantes ainda não analisadas e como se dá o uso das formas que já existem, ou seja, que já foram mapeadas e analisadas, contribuindo assim, para o estudo deste fenômeno no Português Brasileiro. Para tanto, recorreremos ao aporte teórico-metodológico da Sociolinguística Variacionista(LABOV, 2008 [1972]), pois objetivamos analisar se: i) estas formas são as mesmas que são utilizadas por diferentes comunidades de fala já estudadas, no que diz respeito ao uso dos marcadores; ii) esses itens são formas que, segundo a Gramática Tradicional, tem outra função na língua portuguesa e que, por isso, deveriam desempenhar outra função e/ou ter outra posição na sintaxe da sentença; iii) as variáveis sociais e linguísticas influenciam no uso dos elementos identificados como *requisitos de apoio discursivo* na fala da comunidade em análise.

Para tanto, serão analisados dados de fala de pessoas do Sertão do Pajeú – Pernambuco, mais especificamente, de três cidades distintas desta região, a saber: São José do Egito, Triunfo e Serra Talhada. É válido mencionar que a escolha desta região, assim como das cidades mencionadas, se dá em virtude de não existir, até o presente momento, nenhum estudo de descrição e análise deste fenômeno nesta região, evidenciando assim a relevância e o pioneirismo deste estudo, pois analisar e descrever a língua de um povo é, dentre outras coisas, uma maneira de divulgar e perpetuar a identidade de um povo. Além disso, estas

⁸ Apesar de considerá-las como variantes de uma mesma variável linguística, estamos cientes que parece haver alguns fatores que condicionam o uso de determinada forma em detrimento de outra a ponto de, em alguns casos, determinadas formas parecerem que não podem ser substituídas por outras consideradas também variantes desta variável. Porém, consideramos que para identificar que fatores são esses é necessário uma descrição inicial das formas que atuam como RADs, que é exatamente o que propomos realizar nesta pesquisa, não descartando, inclusive, a possibilidade de estudos futuros que busquem analisar tais fatores.

⁹Esta sigla refere-se as informações do informante *nocorpus*da referia pesquisa, sendo composta pelo número do informante no *corpus* (Inf+número) e a localidade onde reside o informante, a saber, ST para Serra Talhada, TR para Triunfo, SJE para São José do Egito e FL para Flores.

descrições permitem observar o quanto a língua possui um caráter heterogêneo, tão abordado nos estudos linguísticos. Os dados serão coletados e analisados seguindo o aporte teórico-metodológico da Sociolinguística Variacionista ou Teoria da Variação e Mudança Linguística, Labov (2008 [1972]) e Mollica (2003), pois este pressupõe que os estudos acerca da língua devem ser feitos com dados provenientes de situações reais de comunicação.

Esta dissertação está organizada, a partir desta introdução, em quatro capítulos. No primeiro, apresentamos o fenômeno em análise por meio de alguns estudos realizados com os marcadores. No segundo capítulo, discorremos acerca do aporte teórico-metodológico da Sociolinguística Variacionista, ressaltando a metodologia proposta por esta perspectiva teórica. No terceiro capítulo, enfatizamos como a pesquisa foi realizada, retomando os passos metodológicos da Sociolinguística Variacionista. O último capítulo traz a análise dos dados coletados, elucidando quais as hipóteses que foram, ou não, confirmadas. Após a análise dos dados, apresentamos as algumas considerações acerca dos resultados obtidos na pesquisa e como estes podem ser norteadores para trabalhos futuros com os marcadores, evidenciando, assim, a importância do estudo que foi realizado e apresentado neste texto.

2 SOBRE OS MARCADORES

Neste capítulo, objetivamos discutir acerca dos marcadores, mais especificamente os Requisitos de Apoio Discursivo que é o nosso objeto de análise, elucidando os traços identificadores, os diferentes nomes e as definições dadas a estes itens, assim como, qual será a nomenclatura que seguiremos na presente pesquisa e o porquê desta escolha. Além disso, buscamos também expor alguns estudos realizados com os marcadores que, além de nos oferecer dados importantes acerca dos elementos que nos propomos analisar, nos servirão como ponto de partida para a realização da presente pesquisa.

Apesar de lidarmos com elementos estigmatizados¹⁰ do ponto de vista da Gramática Tradicional da Língua Portuguesa, Rizzo, Silva e Urbano (2015, p.371) afirmam que eles são “um dado de análise sempre presente nas preocupações dos linguistas”, uma vez que possuem alto índice de realização na língua. Essa presença vem, no entanto, passando por modificações com o passar dos anos e com o avançar das pesquisas com marcadores, pois estudos como os de Marcuschi (1986) e (1989)¹¹, considerados precursores na análise deste fenômeno no Português, objetivavam, basicamente, reconhecer as formas, funções e posições dos marcadores, chamados por ele de marcadores conversacionais. Já estudos mais recentes, como o de Freitag (2008), em que a autora nomeia o fenômeno de marcadores discursivos, o objetivo principal do estudo foi analisar duas variedades do Português Brasileiro. Sendo assim, é possível afirmar que os estudos com marcadores passaram por mudanças em seu foco de pesquisa, no entanto, tanto um quanto outro é imprescindível para a sucessão dos estudos destes itens. Portanto, o presente capítulo trará, na seção seguinte, uma exposição dos primeiros estudos com os marcadores, com o intuito de alcançar os objetivos elencados acima e, posteriormente, apresentará as pesquisas que, como dito anteriormente, serão um ponto de partida para o presente estudo.

2.1 ALGUNS ESTUDOS

Como citado anteriormente, há uma quantidade significativa de pesquisas voltadas para a análise dos marcadores. As teorias que fundamentam cada uma delas diferem bastante

¹⁰ Afirmamos serem elementos estigmatizados tendo em vista que as gramáticas não descrevem como itens necessários à comunicação a ponto de não apresentar pontos e estudos acerca da sistematização e uso destes itens na língua.

¹¹ Apesar de informarmos aqui como sendo dois estudos distintos, a versão de 1989 é uma reorganização da de 1986, sendo assim, usaremos apenas o texto de 1989, por ser mais atual e ter passado pelo processo de reescrita pelo próprio autor.

uma das outras, há estudos dos marcadores baseados em aquisição da linguagem, análise do discurso, análise da conversação, sociolinguística variacionista, entre outras.

Castilho (2004) afirma que os primeiros trabalhos com marcadores foram elaborados por Beinhauer (1964) e Keller (1979). Beinhauer nomeou este fenômeno como sendo “muletillas conversacionais” e Keller, estudando os marcadores em inglês, os nomeou de *gambits*. Marcuschi (1989) faz um levantamento de pesquisas também realizadas fora do Brasil sobre marcador discursivo e cita, como exemplo, os estudos de R. Meyer-Hermann (1979), que desenvolveu uma tese sobre os marcadores metacomunicativos¹² e um ensaio sobre as formas de atenuação (1983) a partir de análises feitas com o português falado em Portugal; Em francês, o estudo sobre os marcadores que recebeu grande destaque foi o de E. Gülich (1970), pois investigou o francês falado e os marcadores existentes nesta fala. No italiano, há muitas pesquisas voltadas para o uso dos marcadores na língua falada, dentre elas se destaca a coletânea de G. Holtus/E. Radtke (1985).

No que tange aos estudos com marcadores do Português Brasileiro, de agora em diante PB, Castilho (2016, p. 229) afirma que Urbano (1993) considera que “Manuel Said Ali teria sido o primeiro a classificar os marcadores discursivos no PB”, já em Castilho (2004, p. 48) o autor afirma que, embora Urbano (1993) tenha destacado que foi Manuel Said Ali¹³ o pesquisador a ter uma percepção pioneira dos marcadores, foram os estudos de Marcuschi (1989) e Macedo e Silva (1996)¹⁴ que deram visibilidade ao fenômeno, motivando, até os dias atuais, análises e descrições destes elementos.

Sendo assim, utilizaremos como estudos que servirão para apresentar os passos iniciais das pesquisas realizadas no Brasil sobre os marcadores, os trabalhos de Marcuschi (1989) e Macedo e Silva (1996), porém, mencionaremos também os estudos realizados por Castilho (2004) e (2016) e o texto de Risso, Silva e Urbano (2015), pois, embora tenham sido publicados recentemente, são reorganizações de trabalhos anteriores que se encaixam no grupo dos estudos precursores dos marcadores, já mencionados neste capítulo. Além destes, tomaremos como base também as pesquisas realizadas por Valle (2001), Freitag (2008) e

¹²Embora o autor não defina o que são marcadores metacomunicativos para o pesquisador que desenvolveu esta pesquisa, ele deixa claro que se tratam de expressões que permitem que haja uma interação entre indivíduos que compartilham a mesma língua.

¹³ O texto de Urbano (1993), assim como os de Castilho (2004) e (2016) não trazem referências bibliográficas, como título e ano da publicação de Manuel Said Ali. Além disso, as demais pesquisas como marcadores, das mais antigas até as mais recentes, não apresentam informações relacionadas ao referido estudo, inviabilizando, assim, o nosso acesso ao conteúdo da pesquisa e, conseqüentemente, a utilização deste como sendo o pioneiro nos estudos dos marcadores.

¹⁴ Este importante estudo foi publicado em 1987, como nesta última publicação as autoras não retiraram informações, apenas modificam aspectos estruturais do texto, utilizamos, como referência, no presente trabalho, a obra de 1996.

Alcântara (2015), as subseções seguintes trazem algumas considerações importantes acerca destes estudos.

2.1.1 Marcuschi (1989)

Marcuschi (1989), no estudo intitulado como *Marcadores Conversacionais do Português Brasileiro: formas, posições e funções*, divide os marcadores em dois grupos, a saber, os *marcadores pragmáticos* e os *marcadores textuais*, no entanto, ele nomeia-os como marcador conversacional, pois considera-os

tanto em suas propriedades interacionais (na condução dos atos ilocutórios e das relações interpessoais) bem como em suas propriedades intratextuais (na estruturação da cadeia linguística). A rigor, isto supõe que o uso da língua na interação verbal ocorre com a aplicação de princípios pragmáticos e de regras linguísticas. Assim, os marcadores conversacionais operam simultaneamente, como organizadores da interação, articuladores do texto e indicadores de força ilocutória, sendo, pois, multifuncionais (MARCUSCHI, 1989, p. 282).

Esta definição dada aos marcadores é exemplar no que tange ao conteúdo do estudo (Cf. MARCUSCHI, 1989), tendo em vista que, muito embora o trabalho seja realizado de acordo com uma perspectiva funcionalista, observando as funções dos marcadores em contextos e condições de produção, a partir de três inquéritos¹⁵ produzidos em atividades de interação, a saber, conversa telefônica e conversação sobre tema sugerido, o autor recorre à teoria de noção de face (*Face Threatening Act*), a estruturação conversacional e a estruturação da cadeia discursiva, com o intuito de mostrar a alta frequência dos marcadores conversacionais em contextos que envolvem a fala e o momento de sua produção, o trabalho não apresenta quantificações e análises extensas dos marcadores, pois segundo o próprio autor, o objetivo principal do estudo é ser uma “introdução programática ao estudo dos marcadores conversacionais no Português Brasileiro” (MARCUSCHI, 1989, p. 283).

Marcuschi afirma que os marcadores conversacionais são bastante utilizados na fala e “têm alta frequência no Português Brasileiro, o que motiva e justifica uma análise sistemática de suas formas, posições e funções.” (MARCUSCHI, 1989, p. 281). Entretanto, para este autor, os marcadores conversacionais, doravante MCs, não constituem uma classe gramatical, pois elementos de todas as classes gramaticais podem funcionar como MCs, por isso, ele dividi-os em quatro grupos com o objetivo de sistematizar as formas em classes. Os grupos foram assim estabelecidos:

¹⁵ Estes textos foram retirados do banco de dados do Projeto de Estudo da Norma Linguística Urbana Culta – NURC. Este projeto possui um *corpus* de análise com dados coletados em Porto Alegre, São Paulo, Rio de Janeiro, Salvador e Recife, porém os dados utilizados por Marcuschi (1989) são apenas os de São Paulo e Recife.

I- MC simples: todo aquele que se realiza com um só lexema ou para-lexema, tais como as interjeições, os advérbios, os verbos, os adjetivos, as conjunções, os pronomes etc. II- MC composto: de caráter sintagmático, com grande tendência à estereotipia e com pouca variação morfológica no tipo produzido. III- MC oracional: que se realiza como pequenas orações, podendo vir em todos os tempos e formas verbais ou modos oracionais (assertivo, indagativo, exclamativo). Aqui entram MCs de caráter estritamente semântico e pragmático como as paráfrases, os resumos, as repetições de frases curtas etc. IV- MC prosódico: trata-se do MC realizado com recursos prosódicos e geralmente produzido com algum MC verbal. Aqui entra a entonação, a pausa, a hesitação, o tom de voz e outros (MARCUSCHI, 1989, p.290-291).

Após o autor definir, através destes grupos, possíveis formas, ele busca estabelecer posições para os MCs. Para isso, ele rompe com algumas ideias de autores que desenvolveram estudos com este fenômeno em outras línguas, entre eles estão J. Rehbein (1979) e D. Schiffrin (1987), pois, para estes autores, as posições iniciais e finais dos MCs são canônicas, mas para Marcuschi não, pois compreende que “o locutor, na interação, organiza o seu texto o tempo todo, o que lhe dá oportunidade de inserir MCs em qualquer ponto” (1989, p.291), sendo assim, o autor insere mais duas posições: medial e ouvinte. A posição medial, como o próprio nome sugere, é quando os MCs aparecem no meio do turno e a posição do ouvinte é quando há sobreposição de vozes, ou seja, quando as falas das pessoas envolvidas na interação se ‘misturam’. No que tange às funções dos MCs, na pesquisa, são apresentadas algumas das funções intratextuais. Marcuschi (1989, p.297) afirma que:

Os MCs exercem funções estruturadoras relevantes, coincidindo distribucional e funcionalmente com operações de organização sintática. Constituem um fator importante na articulação textual, evitando que a conversação se torne uma simples sucessão de monólogos paralelos. Mas, na medida em que encadeiam coesivamente um texto, os MCs também segmentam, agindo como fatores de segmentação e suprimindo em boa medida o papel de pontuação da fala (MARCUSCHI, 1989, p. 297).

Sendo assim, estas funções são analisadas através da observação do texto e da maneira como os MCs atuam nele, dando coesão e segmentando as partes dele¹⁶. Na tentativa de simplificar a análise e se reportar mais afincadamente à teoria, o autor cria um esquema para expor os processos interacionais, partindo do princípio de que “o formato do esquema vincula-se à natureza da ação praticada. Isto quer dizer que o *para quê* exige que se observe *o quê* os interactantes fazem ou podem fazer numa interação verbal” (MARCUSCHI, 1989, p. 305). Sendo assim, o esquema elaborado para explicitar as funções interacionais dos MCs é:

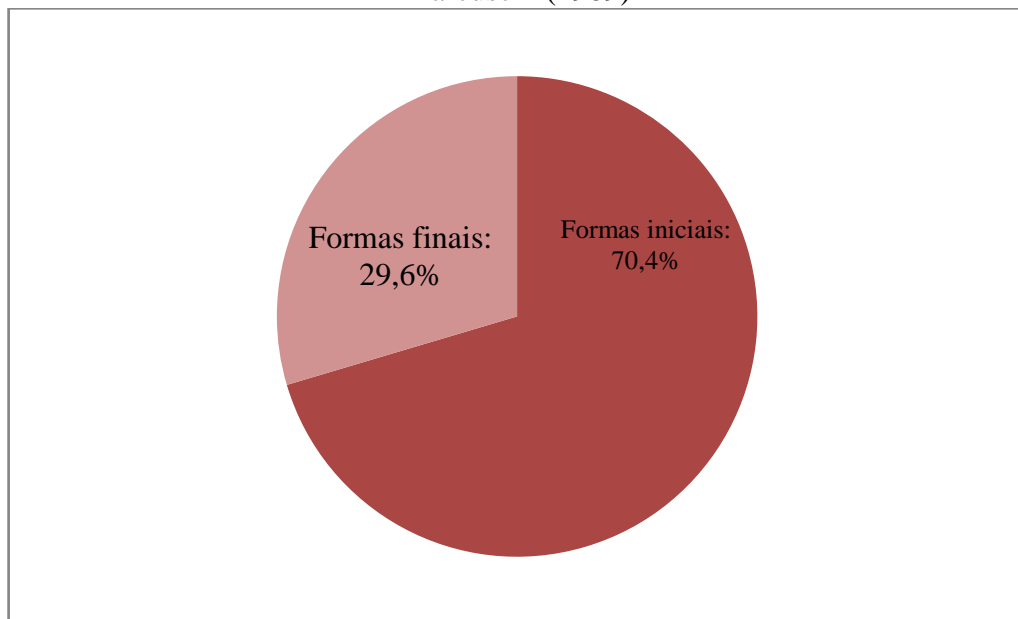
¹⁶ É neste ponto do estudo que o autor justifica a escolha dos textos utilizados para análise, afirmando que a seleção de tais textos se deu como maneira de mostrar ao leitor que os MCs são elementos bem recorrentes na língua, assim como, mostrar como eles, em diferentes contextos e condições sintáticas, são produzidos.

num quadro de ação K_j, K_{jj}, \dots, K_n , um locutor A_1, A_2, \dots, A_n emprega o esquema interacional $EMC_1, EMC_2, \dots, EMC_n$ com a finalidade de obter do seu interlocutor H_1, H_2, \dots, H_n o efeito desejado F_1, F_2, \dots, F_n . Nesta formulação o quadro de ação de K define-se pela natureza do ato produzido, contextualizado e hierarquizado no evento comunicativo em curso; o locutor A e o interlocutor H são os interactantes no evento em curso; o esquema interacional EMC é o recurso ou estratégia utilizado(a); o efeito desejado F representa a função pretendida para aquele esquema (MARCUSCHI, 1989, p. 305).

A partir deste esquema, a pesquisa mostra as funções interacionais dos MCs, uma vez que estas são utilizadas, pelos interlocutores, para comandar e controlar as estratégias usadas no momento da interação.

Partindo destas assertivas, a pesquisa chegou aos seguintes resultados¹⁷: no que tange às posições dos MCs, o número de formas e variações de MCs é bem maior em posições iniciais do que em posições finais¹⁸, conforme podemos ver no gráfico abaixo¹⁹.

Gráfico 1: Porcentagem de uso das formas finais e iniciais de MCs do estudo realizado por Marcuschi (1989)



¹⁷ Alguns dos resultados aqui apresentados não são acompanhados de números e de outras considerações que julgamos serem pertinentes, isso se dá porque a pesquisa não trouxe tais resultados. Sendo assim, justificamos a afirmação que fizemos quando iniciamos a discussão desta pesquisadizendo que este estudo não traz quantificações e análises extensas.

¹⁸ É importante destacar que o autor traz esta afirmação acerca dos marcadores iniciais e finais no texto, (Cf. Marcuschi, 1989, p.292-293), mas não apresenta este número em porcentagem, apenas em números reais de realizações, porém para facilitar a compreensão do nosso leitor, com base nos números apresentados pelo autor no quadro II (p.293), fizemos o cálculo em porcentagem do uso dos marcadores iniciais e finais e expomos por meio do gráfico apresentado abaixo.

¹⁹ O gráfico abaixo não foi retirado do texto de Marcuschi (1989), criamos o gráfico, a partir dos dados apresentados pelo autor, com o intuito de expor de maneira mais clara os resultados por ele alcançados.

No que tange aos MCs mediais e do ouvinte, a pesquisa chega aos seguintes resultados: “Os MCs mediais de natureza verbal são mais frequentes em turnos mais longos...Os MCs do ouvinte têm seu lugar preferido em sobreposição de vozes, ou em momentos de micropausas ou com um outro MC verbal qualquer²⁰” (MARCUSCHI, 1989, p.296).

Além destas considerações, o estudo aponta que as formas verbais de MCs são mais utilizadas que as formas nominais e isso se dá tanto no início quanto no fim do turno, pois 80% dos turnos são iniciados por formas verbais e 20% por formas nominais e na finalização dos turnos, em que se tem um número menor de formas, os MCs verbais finalizam 75% deles, ou seja, os MCs finalizadores são basicamente verbais, tendo em vista que apenas 35% deles são MCs nominais²¹. Já no que diz respeito aos contextos, há diferenças entre os MCs utilizados no telefonema e os MCs utilizados na conversação com o tema sugerido²², sendo que o telefonema apresentou variedade de MCs em todas as posições²³.

Em se tratando das funções, embora o autor tenha desenvolvido um esquema para analisá-las, ele não fez uma análise detalhada sobre elas, não chegando, por isso, a resultados finais. O autor justifica tal prática afirmando que “o esquema aqui apresentado constitui uma parcela mínima do que se poderia analisar e serve apenas como exemplificação” (MARCUSCHI, 1989, p.317). No entanto, ele deixa claro que há a necessidade de se fazer, a partir deste esquema ou de outros a serem desenvolvidos, estudos detalhados sobre as funções interacionais e intratextuais dos MCs. Acreditamos que um estudo mais detalhado das funções dos MCs foi realizado por Macedo e Silva (1996), por isso, apresentamos na subseção seguinte que as autoras apresentam acerca dos marcadores, quais as funções que elas reconhecem que os MCs possuem, qual teoria serviu como base para tal estudo e quais os resultados a que chegaram ao finalizar a pesquisa.

²⁰ Para os MCs mediais e do ouvinte não foi possível o cálculo em porcentagem tendo em vista que nos quadros IV E V apresentados por Marcuschi (1989, p.296), não há o número geral de realizações, algo que nos seria necessário para o cálculo da porcentagem.

²¹ O estudo não traz esses dados relacionados às realizações de MCs mediais e do ouvinte.

²² O pesquisador afirma que os informantes têm formação universitária, são professores ou profissionais autônomos e que a coleta se deu da seguinte forma: o texto I foi coletado pela equipe do projeto NURC/Recife, o texto II pela equipe do projeto NURC/São Paulo e o texto III é uma gravação de uma conversa telefônica de duas grandes amigas acerca de um desfile, sendo que a conversa foi gravada sem que as interlocutoras soubessem. No entanto, o autor não menciona como esta gravação foi realizada.

²³ Com base nisso, o autor afirma que “justifica que se faça uma análise mais aprofundada para verificar a variação de formas de acordo com os tipos de textos” (MARCUSCHI, 1989, p. 295).

2.1.2 Macedo e Silva (1996)

Macedo e Silva em sua pesquisa intitulada *Análisesociolinguística de alguns marcadores conversacionais* (1996)²⁴ apresentam algumas das nomenclaturas que os marcadores possuem. Elas utilizam, assim como Marcuschi (1989), o conceito de Marcador Conversacional, ou simplesmente MC. No entanto, reconhecem que a indefinição do que vem a ser um marcador é tão grande que, de acordo com muitas concepções, qualquer palavra que ajuda a estruturar o que está sendo dito é um marcador, por isso, na tentativa de não seguir este equívoco, as autoras chamam expressões mais longas de ‘estratégias discursivas’ e têm como MCs as estruturas mais pontuantes, dito de outra forma, as estruturas menores. Uma vez que

se chamarmos de ‘marcador’ qualquer partícula ou expressão ‘que ajuda a arrumar o que se quer dizer’, fica difícil saber onde parar (...) tudo no discurso poderá vir a ser chamado ‘marcador’ pois tudo marca ou organiza ou sinaliza alguma informação (MACEDO E SILVA, 1996, p.13).

Feitas estas delimitações, é necessário abordar o aporte teórico-metodológico utilizado na pesquisa, uma vez que este estudo, em muito, se assemelha ao trabalho que pretendemos desenvolver aqui. A pesquisa realizada por Macedo e Silva (1996) tem como base teórica os pressupostos da Sociolinguística Variacionista e utiliza um *corpus*, retirado da Amostra Censo, formado por 64 entrevistas. Os dados desse *corpus* foram estratificados por sexo, idade e escolaridade.

Além disso, foi levado em consideração o gênero do discurso, uma vez que, para as autoras, “não seria possível estudar quantitativamente essas formas sem uma primeira grande separação dos vários gêneros do discurso” (MACEDO E SILVA, 1996, p. 14). Sendo assim, foram selecionados trechos representativos de alguns gêneros, tais como: narrativa, descrição de vida, argumentação, diálogo, citação, descrição e receita, para analisar o uso dos MCs. Todas estas estratificações/seleções tiveram como base as seguintes hipóteses: quanto menor a faixa etária e o nível de escolaridade, maior seria a utilização dos MCs; As mulheres usariam mais MCs, uma vez que parece que elas “usariam mais rodeios para falar do que os homens”

²⁴ Estas autoras além de terem desenvolvido várias pesquisas com os MCs, orientaram muitas dissertações de mestrado nestas áreas. Os trabalhos de Silva e Assafim (1994), Macedo e Carmo (1994) e Riso, Silva e Urbano (1994) são exemplos disso.

(*sic*)²⁵; e, com relação aos gêneros do discurso, há gênero que favorece o uso de determinado tipo de MC e nenhum MC se restringe a ser usado em determinado gênero.

Falando em uso dos MCs, uma das maiores contribuições desta pesquisa, como afirmamos anteriormente, é a classificação das funções, assim como, as posições e sentido que eles podem exercer no discurso. Macedo e Silva (1996) dividem os MCs em nove subgrupos²⁶. A classificação foi assim estabelecida:

Iniciadores: iniciam turnos, podendo também introduzir o discurso direto. Ex: Ah, bom, bem, não, olha, ih, espera. *Requisito de apoio discursivo*: usados para se certificar da atenção do interlocutor, ocorrendo, na grande maioria das vezes, no final de enunciado. Ex: né? tá? Sabe? entendeu? Viu? não é mesmo?. *Redutores*: evitam uma postura assertiva ou autoritária do locutor. Ex: eu acho, pô, sei lá. *Esclarecedores*: tentam resumir ou retomar com maior clareza parte do discurso. Ex: quer dizer, isto é, deixa eu ver, (xovê). *Preenchedores de pausa*: evitam o silêncio enquanto um novo trecho de fala está sendo preparado. Ex: assim, bem, é, hãa. *Sequenciadores*: marcam sequência no discurso. Ex. Aí, então, depois. *Resumidores*: encerram uma lista e resumem o que se considera ser de conhecimento do interlocutor. Ex: essas coisas, e tal, coisa e tal, e tudo, papapá, tatatá. *Argumentadores*: iniciam argumentação, geralmente contrária ao discurso precedente. Ex: agora, não, não mas, é mas, sim mas, eu pra mim. *Finalizadores*: dão um fecho ao turno de um falante. Ex: então tá, é isso aí, tudo bem (MACEDO E SILVA, 1996, p.11 e 12).

A partir desta classificação, as pesquisadoras selecionaram, para a análise, quatro tipos de MCs²⁷, são eles: os iniciadores *bom, olha ah*, os Requisitos de Apoio Discursivo *né? sabe? entendeu?*, a forma “*assim*” e o sequenciador “*aí*”. A partir desta seleção, as autoras fazem uma discussão, um pouco mais extensa, das funções que cada tipo destes MCs selecionados possui, fazendo um resumo das funções destes, Macedo e Silva (1996, p.16) defendem que:

Os requisitos de apoio discursivo ligam-se à interação entre os falantes. O marcador *aí* diz respeito à articulação de pequenos segmentos. A forma *assim* é um anunciador de complementos, podendo estar ligado ao processamento da fala. Os iniciadores podem tanto organizar segmentos maiores do discurso, como tópicos ou assuntos da entrevista, ou ainda exercer função interacional na marcação das tomadas de turno.

Até aqui, apontamos algumas definições e metodologias utilizadas na pesquisa de Macedo e Silva (1996). Fizemos estes apontamentos acreditando que essa pesquisa traz um

²⁵ As autoras utilizam ainda mais dois argumentos para defender a hipótese de que as mulheres usam mais os marcadores, são elas: “o estilo feminino seria menos assertivo por influência da educação e socialização dos dois sexos. Acreditamos que, ainda em nossa sociedade patriarcal, há submissão e insegurança da mulher que, ao se comunicar, busca amenizar o seu discurso apoiando-se no maior emprego dos marcadores em geral” (Macedo e Silva, 1996, p.16).

²⁶ Esta divisão é feita levando em consideração o sentido, a função e a posição dos MCs no discurso.

²⁷ Segundo as autoras, esta seleção se deu, a princípio, devido à grande quantidade de realização durante as entrevistas, porém, no desenvolvimento da pesquisa, elas perceberam que estes tipos de MCs são bastante representativos das funções que, geralmente, os MCs recebem.

excelente material que pode ser consultado para o desenvolvimento de futuros estudos dos MCs. Porém, além disso, os resultados alcançados nesse estudo também são de grande valia para futuras pesquisas, assim como para o presente estudo, daí julgarmos ser necessário apresentá-los aqui.

Os resultados alcançados foram apresentados em duas partes: na primeira, foram expostos os que foram obtidos para cada tipo de MCs, sendo assim, os Requisitos de Apoio Discursivo, os Iniciadores, o item “assim” e o sequenciador “aí” foram analisados separadamente, levando em consideração as variáveis sociais e o gênero do discurso, assim como algumas variáveis linguísticas específicas de cada tipo de MC. Na segunda parte, as autoras apresentam, de maneira geral, os resultados a que chegaram²⁸.

As pesquisadoras, ao analisarem os dados, chegaram aos seguintes resultados: o uso dos RADs parecem não ser influenciado por variáveis sociais, pois a probabilidade de uso foi de,50²⁹ para as variáveis sociais selecionadas na pesquisa, a saber, sexo, escolaridade e faixa-etária. No entanto, as autoras afirmam que no que diz respeito ao gênero há uma frequência maior de uso dos RADs nas argumentações. Os números de realizações, porcentagem e probabilidade de uso são apresentados pelas autoras num quadro intitulado *Emprego dos RADs por gênero discursivo*. Para facilitar a compreensão, expomos os dados de probabilidade de uso, em peso relativo no quadro abaixo.

Quadro 1: Emprego dos RADs por gênero discursivo³⁰

Argumentação	,72
Descrição	,59
Descrição de vida	,58
Receita	,45
Narração	,48
Citação	,18

Fonte: Macedo e Silva (1996, p.16)

²⁸ As análises feitas por Macedo e Silva (1996) foram muito detalhadas e extensas e, conseqüentemente os resultados também foram. Em virtude disso, optamos por apresentar aqui, os resultados gerais destas análises, mostrando as conclusões a que as autoras chegaram a partir das análises das variáveis sexo, idade, escolaridade e gênero do discursivo, desprezando, portanto, os resultados obtidos para cada tipo de MC.

²⁹ Os números de peso relativo são geralmente apresentados após um ponto (.), no entanto, no texto, estes dados são apresentados após uma vírgula (,), não sabemos a razão, no entanto, hipotetizamos que o uso da vírgula seja comum ao programa utilizado pelas autoras para obtenção do peso relativo, a saber, o programa computacional SWAVA, citado pelas autoras. Sendo assim, grafamos o peso relativo assim como foi apresentado pelas autoras, ou seja, precedido por uma vírgula.

³⁰ É importante destacar que não consideramos a nomenclatura de gênero discursivo pertinente para nomear os gêneros apresentados pelas autoras. No entanto, seguimos a nomenclatura dada pelas autoras em sua pesquisa.

No que tange aos demais marcadores analisados, o estudo mostra que, diferentemente dos RADs, o “aí”, o “assim” e os iniciadores mostraram ser influenciados por algumas destas variáveis sociais, o “aí”, por exemplo, mostrou ser influenciado pela variável escolaridade, pois a probabilidade de uso, apontou ,55 para pessoas com ensino fundamental I, evidenciando que “quanto menor for o nível de escolaridade, maior é o uso do conectivo “aí” (Macedo e Silva, 1996, p.29). Para a variável faixa etária, a conclusão foi a seguinte: “quanto mais jovem o falante, mais frequente será o uso da forma ‘aí’” (Macedo e Silva, 1996, p. 28), as autoras afirmam isso baseadas nos números de probabilidade obtidos, pois para a faixa etária de 7-14 anos, a probabilidade de uso foi de ,70, enquanto que para as pessoas com + de 50 anos, a probabilidade de uso é de apenas ,30. O “assim” mostrou ter influência da faixa etária e do sexo, com probabilidade de uso ,60 para as mulheres e ,40 para os homens; e diminuição do uso nos falantes com mais idade ,70 para pessoas entre 7 e 14 anos e ,30 para pessoas com mais de 50 anos. A variável sexo mostrou influenciar o uso dos iniciadores, pois a probabilidade é de ,60 para as mulheres e ,40 dos homens. Diferentemente da pouca influência que os MCs sofrem das variáveis sociais, os gêneros do discurso influenciam o uso dos MCs e cada gênero favorece determinada função dos MCs, uma vez que, por exemplo, a função interativa é mais constante no gênero da argumentação, a conexão ocorre com mais frequência em narrativas e receitas; já na citação, o “ah” é mais usado, pois ele atua como iniciador de turnos anteriormente citados.

Com base em tudo isso, fica muito claro que pesquisas como estas podem servir como uma base para próximos estudos. A pesquisa de Macedo e Silva (1996) é bastante significativa quando utilizamos os resultados para compararmos com análises posteriores, mas é ainda mais importante quando consultamos e analisamos as funções, posições e sentidos que as autoras defendem que os MCs têm³¹.

³¹ Além das pesquisas que estas autoras realizaram com os marcadores discursivos e que já foram citadas anteriormente, a autora Alzira Tavares de Macedo desenvolveu um estudo voltado para a aquisição dos marcadores conversacionais (MACEDO, 1997). O trabalho, intitulado de “Aquisição de marcadores em primeira e em segunda língua”, percebeu que as pessoas que estão adquirindo a L2 usam mais os MCs do que as crianças que estão adquirindo a língua materna. Segundo a autora, isso ocorre por causa do amadurecimento, tanto em idade, quanto em fluência discursiva em geral, pois os textos das crianças são mais curtos e menos elaborados, necessitando de menos marcas coesivas, de menos reformulações. No entanto, as crianças usam mais os marcadores *aí, ah, oh, e ih* do que adultos e não utilizam as formas *olha e bom*.

2.1.3 Castilho (2004) e (2016)

Na obra intitulada *A língua falada no ensino de português*, Castilho (2004) faz uma breve discussão sobre a análise da conversação. Neste capítulo, o autor apresenta algumas considerações acerca dos marcadores. Ele os nomeia de marcadores conversacionais, ou simplesmente MCs, e os define como recursos prosódicos que ocorrem com frequência na fala e que funcionam como articuladores da conversação. Esses recursos são “pausas, articulação enfática, alongamentos, certos itens lexicais e pré-lexicais, ou mesmo expressões mais complexas” (CASTILHO, 2004, p. 46). Além desta definição, o autor expõe algumas das nomenclaturas encontradas nas Gramáticas Tradicionais e na Análise da Conversação a respeito dos marcadores, sem fazer referências a nomes, o autor afirma que nestas gramáticas eles são “descritos como palavras expletivas ou denotativas, expressão de realce e palavras de difícil classificação e na Análise da Conversação como organizadores globais” (CASTILHO, 2004, p. 47).

A despeito destas e de outras afirmações que seguem a mesma linha, Castilho (2004, p.47) resume suas ideias a respeito dos marcadores mostrando a maneira como os compreende, afirmando que eles são segmentos

(i) sintaticamente independentes do verbo (ii) constantes de um ou de mais de um item lexical, ou mesmo de expressões não lexicais, (iii) funcionando no monitoramento da conversação e na organização do texto, (iv) distribuídos no início, no meio ou no final da unidade de análise. Os MCs, em suma, verbalizam o monitoramento da fala, sendo frequentemente vazios de conteúdo semântico, portanto irrelevantes para o processamento do assunto, porém altamente relevantes para manter a interação (CASTILHO, 2004, p. 47).

Além destas assertivas, Castilho, especificamente neste texto, expõe uma opinião própria acerca dos marcadores, pois, para ele, eles são divididos em três grupos funcionais: os marcadores pragmáticos, os marcadores textuais e os modalizadores de tópico. Afirmamos ser uma opinião própria deste autor e específica deste texto, pois outros autores, dentre eles Marcuschi (1989), consideram que existem apenas dois grupos funcionais, a saber, os marcadores pragmáticos e os marcadores textuais e em outros textos, como o que apresentaremos a seguir, o próprio Castilho menciona apenas os dois primeiros grupos.

Já no quinto capítulo da Nova Gramática do Português Brasileiro (2016), intitulado *A conversação e o texto*, este mesmo autor traz algumas das funções e a colocação dos marcadores no enunciado. Desta vez, Castilho usa a nomenclatura de *marcador discursivo* e

defende que “essa classe é polifuncional, operando o mesmo item em mais de uma função” (p. 229). Para apresentá-los, ele faz uma distinção dos marcadores, ou seja, afirma que eles podem ser linguísticos (verbais e prosódicos) e não linguísticos (olhar, riso, expressão corporal). Além disso, o autor apresenta os marcadores discursivos que fazem parte dos grupos³²: marcadores pragmáticos, estes orientados para o interlocutor, e marcadores textuais, orientados para o texto. Neste texto, o autor parece ter abandonado a ideia de três grupos, como citado no texto de Castilho (2004), pois lida apenas com esses dois. Neles, o autor subdivide os marcadores discursivos a partir de suas funções, no caso dos marcadores textuais são os que: *iniciam o tópico, recusam o tópico, aceitam o tópico, organizam o tópico, operam a mudança de tópico, modalizam o tópico e finalizam o tópico*, já no caso dos marcadores discursivos pragmáticos, eles são subdivididos a partir de sua colocação no enunciado, ou seja, se são: *iniciais, mediais ou finais*. Para melhor compreensão, selecionamos alguns dos marcadores apresentados por Castilho (2016) e expomos no quadro abaixo para exemplificar a subdivisão feita pelo autor³³.

Quadro 2: Marcadores discursivos: funções e colocação do enunciado

Marcadores pragmáticos ou interpessoais	Marcadores textuais ou ideacionais
Iniciais: <i>Ah... eh... ahn...</i> <i>E aí, tudo bem?</i> <i>Mas...</i>	Iniciam o tópico: <i>Bom...</i> <i>Bem...</i> <i>Assim...</i>
Mediais: <i>... é claro...</i> <i>... tô entendendo...</i> <i>... é ...</i>	Recusam o tópico: <i>Peraí, sem essa!</i> <i>Essa não!</i> <i>Corta essa!</i>
Finais: <i>... tá?</i> <i>... sabe?</i> <i>... né?</i>	Aceitam o tópico: <i>Tá bom...</i> <i>Vamos lá...</i> <i>Ok...</i>
	Organizam o tópico: <i>E aí...</i> <i>E tem mais...</i>

³² Estes grupos, pragmáticos e textuais, apresentam apenas os marcadores discursivos que fazem parte dos marcadores linguísticos, ou seja, dos marcadores discursivos verbais ou prosódicos. Os marcadores discursivos não linguísticos não são descritos pelo autor, pois, para isso, seria necessário gravar uma série de vídeos com um número enorme de informantes para se ter conhecimento deles.

³³ O quadro citado é muito extenso, por isso, selecionamos apenas alguns exemplos de marcador para que o leitor tenha uma percepção de maneira mais clara das subdivisões feita por Castilho (2016).

	<i>E então...</i>
	Operam a mudança de tópico: <i>Já</i>
	Modalizam o tópico: <i>Sim, mas...</i> <i>Sei lá...</i> <i>Basicamente...</i>
	Finalizam o tópico: <i>Papapa...</i> <i>Falô...</i> <i>E coisa e tal...</i>

Fonte: Castilho (2016, p. 229-230)

A partir destas subdivisões, é possível constatar a multifuncionalidade dos marcadores discursivos, assim como, quais as classes e/ou grupos que cada item pertence. Além disso, como o próprio autor sugere, é possível, por meio do quadro, conhecer/reconhecer algumas formas que atuam como marcador, uma vez que foram listadas apenas algumas formas de marcadores (Cf. Castilho, 2016, p.229-230).

No limiar dos estudos com marcadores, não no sentido de data, mas tendo em vista o conteúdo e o foco do estudo, como mencionamos anteriormente, destacamos, e apresentamos na próxima subseção, o estudo com marcadores realizado por Risso, Silva e Urbano (2015)

2.1.4 Risso, Silva e Urbano (2015)³⁴

A pesquisa intitulada *Traços definidores dos marcadores discursivos*³⁵ parte da premissa que há uma escassez no que diz respeito a uma definição precisa do que vem a ser um marcador. Sendo assim, os pesquisadores objetivam caracterizar os marcadores a partir de um estabelecimento de traços identificadores alcançados a partir das análises de algumas variáveis recorrentes em diferentes pesquisas, daí afirmarmos que este trabalho pode ser inserido na fronteira dos estudos com marcadores, ou seja, ele traz as marcas do fenômeno analisado a partir de estudo de situações reais de comunicação, ou seja, tem a pretensão de

³⁴ Utilizaremos a edição 2015 do estudo de Risso, Silva e Urbano, mas destacamos que este é um trabalho publicado em (1994), sob a orientação das autoras Macedo e Silva. No entanto, na publicação de 2015, há algumas reformulações no que diz respeito à estruturação do texto, nada do conteúdo e da análise feito na pesquisa foi alterado e/ou modificado.

³⁵ Esta pesquisa também foi publicada no livro *Gramática do Português Falado*, vol. VI: Desenvolvimentos, (1996) Organizado por Ingedore G. Villaça Koch.

ser, não apenas um estudo teórico, mas um estudo de caso que seja capaz de consolidar e ampliar as discussões em torno das definições e caracterizações dos marcadores.

A pesquisainvestigou as ocorrências dos marcadores em gravações, de 15 minutos cada inquirido, do projeto *Nurc - Norma Urbana Culta* - do Projeto Gramática do Português Falado (PGPF). Em um total de 1298 ocorrências, os marcadores foram analisados a partir das seguintes variáveis: “1. padrão de recorrência;2. articulação de segmentos do discurso;3. orientação da interação;4. relação com o conteúdo proposicional;5. transparência semântica;6. apresentação formal;7. relação sintática com a estrutura gramatical da oração;8. demarcação prosódica;9. autonomia comunicativa;10. massa fônica;” (RISSO, SILVA E URBANO 2015, p.373).

Os pesquisadores consideram que o grupo de marcadores em língua portuguesa, além de ser extenso, tem “uma agregação contínua de novos exemplares ao conjunto de marcadores já existente” (p.372), provocando, assim, um grupo de elementos cada vez mais “amorfo e heterogêneo” (p. 372). Em virtude disso, destacam a importância de investigar as propriedades definidoras dos marcadores, com o intuito de operacionalizar os estudos futuros, para que estes não venham a considerar elementos da língua como marcadores, quando eles não o forem, assim como deixar de considerar e analisar outros itens que, de fato, são marcadores, pois, para os pesquisadores, se a definição não for precisa, as formas que integram a classe de marcadores também podem não ser, pois, segundo alguns estudos, algumas formas estão entre os marcadores e, para outros estudos, estas mesmas formas não se inserem entre os marcadores, além de ter formas que atuam em diferentes classe gramaticais ao mesmo tempo que também atuam como marcadores.

Para tentar lidar com este impasse da indefinição, o estudo buscou definir os marcadores por meios das variáveis mencionadas acima, afirmando que:

a seleção recaiu sobre essas variáveis, tendo em vista recobrir e reunir aspectos avulsamente considerados, de forma explícita ou implícita, na literatura linguística sobre o assunto e nas nossas experiências de trabalho com os marcadores discursivos. Dentro de cada variável, levamos em conta as diferentes possibilidades pelas quais ela pudesse se definir. Dessa forma, as variáveis foram recortadas em seus respectivos traços potenciais (RISSO, SILVA E URBANO, 2015, p. 373).

Após a análise dos dados, os pesquisadores chegaram aos seguintes traços definidores dos marcadores: a) alta recorrência; b) exterioridade ao conteúdo proposicional; c) transparência semântica parcial; d) invariabilidade formal ou invariabilidade restrita;

e) independência sintática; f) demarcação prosódica; g) não autonomia comunicativa e h) massa fônica reduzida (Cf. RISSO, SILVA E URBANO, 2015, p.381). Apesar de os autores afirmarem que estes parecem ser traços fortes para a definição dos marcadores, é importante destacar que os próprios pesquisadores reconhecem que “a condição de marcador discursivo não é dada necessariamente pela *ocorrência absoluta de todos os traços* considerados fortes para caracterizá-lo” (2015, p. 383 – ênfase nossa), ou seja, trata-se de traços identificadores dos marcadores que podem ocorrer, ou não, em sua totalidade, a depender do marcador ou dos grupos de marcadores analisados.

É inegável o quanto as constatações feitas após a análise de Risso, Silva e Urbano (2015) contribuem para a compreensão e estudo dos marcadores. No entanto, as contribuições desse trabalho vão além da exposição destes traços, pois os autores dedicam boa parte do trabalho à discussão acerca das nomenclaturas dadas aos marcadores, pois para eles:

os desencontros na compreensão do estatuto dos marcadores costumam estender-se à sua denominação. Registra-se, com efeito, uma acentuada pulverização de rótulos usados pelos linguistas para fazer referência a eles. Essa pulverização, se de um lado atesta a especificidade dos enfoques dado à descrição das unidades consideradas como marcadores, de outro revela a oscilação na compreensão do que os marcadores venham a ser, ou seja, a falta de um assentamento comum de suas propriedades básicas (RISSO, SILVA E URBANO, 2015, P.372).

Sendo assim, a falta de definição acerca dos marcadores alcança também a denominação destes itens. Por conta disso, encontra-se, nos estudos com marcadores, como já destacamos, pesquisadores que os nomeiam de “marcadores discursivos”, “marcadores conversacionais”, “articuladores textuais”, “operadores argumentativos” ou “vícios de linguagem”³⁶. Para Risso, Silva e Urbano (2015), a denominação de marcador discursivo é mais adequada, apesar de a nomenclatura “marcador conversacional” ser mais frequente. Para estes autores, nomear este grupo de marcador discursivo³⁷ é ser mais abrangente, uma vez que não se detém apenas à língua falada e à conversação:

³⁶ As nomenclaturas “articuladores textuais” e “operadores comunicativos” são mais recorrentes em estudos baseados na Linguística de Texto, Já o termo “vícios de linguagem” é uma nomeação pouco aceita pelos linguistas, tendo em vista seu caráter pejorativo. Tal termo é mais utilizado, segundo Macedo e Silva (1996), por autores de cursos de oratórias e/ou compêndios que abordam normas gramaticais.

³⁷ É válido ressaltar aqui que, embora tenhamos utilizado neste trecho e em outras partes deste estudo a expressão grupo de marcadores e/ou classe de marcador, o conceito de grupo/classe de marcador é algo ainda inacabado nos estudos linguísticos, pois alguns pesquisadores entendem que os marcadores constituem uma classe de palavras, outros preferem chamar de grupo de palavras com o intuito, possivelmente, que ao utilizar o termo “classe de palavras” não venha a ser confundido com as classes de palavras estabelecidas pela gramática tradicional, uma vez que elementos de várias destas classes estariam inseridos no grupo dos marcadores. No entanto, assumir a concepção de que os marcadores constituem um grupo de palavras é assumir uma

adotamos a denominação de marcadores discursivos que nos parece ser mais adequada e abrangente do que a de marcadores conversacionais. Embora esta outra seja a mais corrente e aceita entre os linguistas brasileiros, reconhecemos nela uma limitação por sugerir, de forma inevitável e inadequada, um comprometimento exclusivo com a língua falada, e, dentro dessa modalidade, com um gênero específico, que é a conversação (RISSO, SILVA E URBANO, 2015, p.372).

Tais afirmações fazem compreender que há peculiaridades entre um termo e outro e são essas particularidades que, possivelmente, motivam os linguistas a usarem um ou outro termo, principalmente se irão analisar dados de língua falada ou não. No entanto, há estudos, como o de Valle (2001), que trabalham com língua falada e, mesmo assim, utilizam-se das duas nomenclaturas para se referir ao objeto em análise e isso sem comprometer o estudo e/ou a caracterização do fenômeno linguístico, evidenciando, assim, uma escolha despretensiosa e sem implicações posteriores.

Além destas pesquisas já mencionadas, outro estudo que pode contribuir significativamente para o trabalho com os marcadores é a pesquisa realizada por Valle (2001), esta será apresentada na próxima subseção.

2.1.5 Valle (2001)

O estudo trata-se de uma dissertação de mestrado apresentada ao curso de Pós-Graduação em Linguística da Universidade Federal de Santa Catarina e que apresenta uma característica distinta da dos estudos mencionados até aqui, pois Valle (2001) volta-se mais para a análise do uso dos marcadores, recorrendo a estudos realizados anteriormente para se apropriar das nomenclaturas e definições, mantendo seu foco maior em analisar o uso sistêmico dos marcadores.

Valle (2001) desenvolveu sua pesquisa baseada em dois pressupostos teóricos, a Teoria do Funcionalismo Linguístico, Givón (1993) e (1995) e Moura Neves (1997), e o aporte teórico-metodológico da Teoria da Variação e Mudança Linguística, Labov (1972) e Weinreich, Labov e Herzog (1968), para analisar os itens *sabe? não tem? E entende?* todos, segundo a autora, pertencentes à categoria de Requisitos de Apoio Discursivos, doravante RADs. Pelo fato da pesquisadora fazer menção aos RADs, um dos nove grupos propostos por

nomenclatura não tão recorrente em estudos de língua(gem), pois não é tão costumeiro expressões como o grupo de preposições, de advérbios e assim por diante. Sendo assim, há alternâncias, neste texto e nas inúmeras pesquisas, entre os termos classe, grupo ou conjunto de marcadores.

Macedo e Silva (1996), julgamos, no início da leitura, que ela assumiria o conceito de MC, uma vez que esta nomenclatura foi proposta por Macedo e Silva, após subdividir os MCs em grupos. No entanto, não foi isso que aconteceu. Durante todo o texto a autora alterna entre marcador discursivo e MC.

Os principais objetivos do estudo realizado por Valle (2001) foram tentar mostrar que estes itens são variantes de uma mesma variável, assim como, descobrir quais os percursos de mudança funcional dos elementos que se propôs analisar, ou seja, *sabe? Não tem? e Entende?*. Além disso, a pesquisa busca constatar se o uso destes itens sofre influência de fatores sociais. Para tanto, a autora analisa uma amostra composta por 36 entrevistas com pessoas da cidade de Florianópolis - SC, retiradas do banco de dados do projeto Variação Linguística da Região Sul – VARSUL – e estratificadas a partir do sexo, escolaridade e idade dos informantes.

Antes de apresentarmos os resultados propriamente ditos, que esta pesquisa alcançou, acreditamos que é necessário mencionarmos algo que foi, para a realização de nosso estudo, especificamente, bastante esclarecedor; e, por isso, de muita importância. Trata-se do reconhecimento dos itens que compõem o grupo de RADs como sendo variantes de uma mesma variável linguística. Sabemos que, para os estudos em Variação Linguística, as condições necessárias para a variabilidade entre duas ou mais formas são a equivalência de significado e o uso em um mesmo contexto³⁸. Sendo assim a autora defende que

Tomando o termo significado de forma alargada, podemos abrigar nela a noção de função e prever que os Requisitos de Apoio Discursivo, apesar de atuarem em inúmeras subfunções, encontram-se unidos sob uma mesma propriedade e, portanto, possuem o mesmo significado. Quanto ao contexto dos itens, apesar de contarmos com certas especializações, não acreditamos, *a priori*, que existam comportamentos categóricos. Sendo assim, creditamos a cada um dos três itens em análise o estatuto de variante de uma mesma variável linguística, já que são intercambiáveis em uma mesma função e em um mesmo contexto (VALLE, 2001, p.49).

A autora faz essa afirmação, para justificar a análise dos três itens, *sabe? não tem? e entende?*, como sendo variantes de uma mesma variável linguística. Neste caso, e no nosso também, a variável linguística são os RADs, e as variantes são cada item que compõe este grupo e que serve para requisitar apoio durante o discurso. A análise destas variantes foi desenvolvida a partir dos métodos qualitativos e quantitativos. A autora recorreu a estes métodos por se basear nas ideias propostas por Schifffrin (1987) e reconhecer que “a análise

³⁸ Discutimos melhor sobre variantes e variáveis linguísticas no capítulo seguinte, onde apresentamos o aporte teórico que utilizamos nesta pesquisa.

quantitativa depende da descrição qualitativa anterior e os resultados numéricos não dizem muito sozinhos, devendo ser interpretados em termos qualitativos” (VALLE, 2001, p. 52).

A partir destas análises, a pesquisa chegou aos seguintes resultados: na comunidade de fala estudada, os RADs parecem não ser condicionados por variáveis sociais. Em termos percentuais, homens e mulheres, usam igualmente os RADs, apesar de apresentarem-se mais frequentes nas falas de pessoas mais jovens e mais escolarizadas, as frequências são bastante semelhantes às de pessoas mais velhas e menos escolarizadas. No que diz respeito aos percursos de mudança funcional dos itens analisados, os resultados foram os seguintes: o item *entende?* é menos utilizado, quando comparado com os outros dois itens analisados; já o item *não tem?* ainda é bastante vinculado ao seu sentido-origem, ou seja, negação do verbo ter. Mas, analisando estes dois itens, não foi possível concluir quais deles estão mais avançados no processo de gramaticalização. No que tange ao item *sabe?*, também analisado nesta pesquisa, este está em um nível mais avançado no processo de gramaticalização, uma vez que encontra-se muito distante do seu sentido original, ou seja, o presente do verbo saber.

A autora conclui a pesquisa listando alguns pontos que merecem ser estudados posteriormente, dentre eles, ela cita a necessidade de analisar outros usos dos RADs e levantar outras formas concorrentes as que foram, por ela, analisadas. É seguindo essa linha que Freitag (2008) desenvolve uma pesquisa comparando o uso dos RADs na fala de pessoas da cidade de Itabaiana-Sergipe com os resultados que foram apontados na pesquisa feita por Valle (2001), uma vez que esta desenvolveu pesquisa semelhante, mas em outra região do país, ou seja, Florianópolis – Santa Catarina, o estudo de Freitag (2008) é apresentado na subseção seguinte.

2.1.6 Freitag (2008)

Levando em consideração o uso e a função destes itens, a pesquisa intitulada “Marcadores discursivos interacionais: análise contrastiva entre duas variedades do português falado no Brasil” foi realizada por Raquel Meister Ko. Freitag (2008) com dados retirados do banco de dados GELINS – Grupo de Estudos da Linguagem, Interação e Sociedade. Freitag (2008) chama os marcadores de marcadores discursivos e os define como “construções que atuam tanto no plano textual, estabelecendo elos coesivos entre partes do texto, como no plano interpessoal, mantendo a interação falante/ouvinte e auxiliando no planejamento da fala” (FREITAG, 2008, p. 2). É interessante destacar que Freitag (2008) segue o mesmo

aporte teórico, a mesma metodologia e analisa a mesma variável linguística que Valle (2001), porém assume uma nomenclatura diferente para os marcadores. Cabe destacar também que o fenômeno analisado foi, durante a pesquisa, chamado de RADs – Requisitos de Apoio Discursivo –, porém, várias vezes a autora utilizou, entre parênteses, a nomenclatura Marcadores Discursivos interacionais, nota-se, assim, que a mesma categoria é conhecida por diferentes nomes.

Baseado nos pressupostos da teoria variacionista laboviana (LABOV, 2008 [1972]), o estudo chegou aos seguintes resultados: o *né?* é a forma mais recorrente utilizada pelos informantes para requisitar apoio discursivo em Itabaiana /SE e o mesmo acontece em Florianópolis/SC, como apontou Valle (2001). No entanto, há em Itabaiana duas formas que não são realizadas em Florianópolis enquanto RAD, são elas: *pronto* e *repare*. De acordo com a pesquisa, a forma *repare* desempenha a função de chamar a atenção do ouvinte e a forma *pronto* não teve sua função analisada, uma vez que esta forma mostrou-se ter um comportamento bastante peculiar, digno de uma análise específica que a autora dessa pesquisa não se propôs a fazer. Sendo assim, a pesquisa traz a seguinte conclusão: “uma variedade linguística da região sul do Brasil, Florianópolis/SC, difere-se em alguns aspectos de uma variedade da região nordeste, Itabaiana/SE” (FREITAG, 2008, p.11).

O trabalho de Freitag (2008) traz, ainda, importantes assertivas acerca dos marcadores, ao analisar as propriedades dos RADs e a maneira como eles atuam no plano textual. Dentre elas, destacamos a seguinte conceituação:

os Requisitos de Apoio Discursivo são construções originárias de verbos ou adjetivos que passam por processo de mudança linguística (...) Os Requisitos de Apoio Discursivo atuam, primariamente, no plano interpessoal, dada a sua origem como pergunta (pergunta plena>pergunta semi-retórica> pergunta retórica). No plano interpessoal e textual, atuam como elementos focalizadores de informação no texto (FREITAG, 2008, p. 2).

Por esta e outras definições apresentadas no texto, essa pesquisa mostra-se como um importante estudo acerca dos marcadores, sendo, também, um trabalho pioneiro voltado para os RADs, uma vez que seus resultados, a despeito de todos estes e outros pontos, constituem-se em um importante material a ser consultado para a realização de estudos comparativos com dados de outras regiões do país, como o estudo realizado por Alcântara (2015).

2.1.7 Alcântara (2015)

A pesquisa intitulada *O uso dos Requisitos de Apoio Discursivo por radialistas do Sertão do Pajeú em dois níveis de formalidades diferentes: um estudo sociolinguístico*, buscou analisar, a partir dos pressupostos teóricos da Sociolinguística Variacionista, o uso dos RADs em dois diferentes níveis de formalidade, a saber, entrevistas informais, consideradas o contexto menos informal, e programas de rádios, considerados o contexto mais formal, tendo como hipótese principal que o uso dos RADs poderia ser mais frequente em contextos menos formal. Para tanto, foram analisados dados de fala de radialistas de cinco cidades do Sertão do Pajeú, a saber, Afogados da Ingazeira, Flores, São José do Egito, Serra Talhada e Triunfo. Após as gravações, transcrições e análises realizadas o estudo chegou a significativos resultados, alguns destes apresentamos³⁹ aqui uma vez que eles são norteadores para o presente estudo.

O estudo realizado por Alcântara (2015) observou o quanto o uso dos RADs é recorrente na fala dos radialistas do Sertão do Pajeú, pois foram contabilizadas 654 realizações de RADs, sendo 342 realizações nas entrevistas informais e 312 nos programas de rádio gravados⁴⁰. Além disso, apontou que parece ser que a variável dos RADs possui outras variantes, além das que foram colocadas por Macedo e Silva (1996) na fala dos radialistas sertanejos, pois houve a realização de variantes como *tá bom?*, *tá certo?* *hein* e *ok?* O quadro abaixo mostra, em valores reais e percentuais, o uso das variantes em cada contexto.

Quadro 3: Número de realizações e porcentagem de uso das variantes encontradas em cada contexto

RADs	Entrevista	Programas de rádio
<i>Né?</i>	300 vezes (87,72%)	112 vezes (35,90%)
<i>Entendeu?</i>	22 vezes (6,44%)	0 (0%)
<i>Sabe?</i>	7 vezes (2,05%)	5 vezes (1,61%)
<i>Tá bom?</i>	1 vez (0,29%)	20 vezes (6,42%)
<i>Viu?</i>	10 vezes (2,93%)	74 vezes (23,73%)
<i>Tá?</i>	2 vezes (0,57%)	45 vezes (14,42%)

³⁹Detemos-nos aqui a apresentar apenas os resultados relacionadas às variantes encontradas na comunidade de fala analisada e nos contextos analisados. Não citamos os resultados acerca das variáveis linguísticas e sociais tendo em vista que estes não são relevantes para o que pretendemos aqui ao mencionarmos este estudo, que é fazer referência a um trabalho que, embora não tenha sido o objetivo principal, observou a utilização de variantes de RADs que parecem ser específicos da região.

⁴⁰ Para a coleta de dados, a autora localizou radialistas de cada cidade onde o trabalho foi realizado, após essa localização, foram gravados os programas de rádio apresentados por tais locutores e em seguida, foram gravadas as entrevistas informais com os mesmos locutores, nesse momento, o participante da pesquisa autorizava ou não a análise dos dados de fala obtidos tanto na entrevista quanto no programa de rádio.

<i>Hein?</i>	0	45 vezes (14,42%)
<i>Tá certo?</i>	0	8 vezes (2,54%)
<i>Ok?</i>	0	3 vezes (0,96%)

Fonte: Alcântara (2015, p. 48)

Apesar de ter tido um número bem próximo de realizações nos dois contextos analisados, a hipótese de que em um deles, o menos formal, os RADs seriam mais utilizados foi confirmada, pois na entrevista o número de RADs foi maior do que nos programas de rádio. Ainda em relação ao contexto, a pesquisa concluiu que parece ser que há contextos em que o uso de determinada variante é mais comum do que em outros, pois *oné?* e *entendeu?* foram mais utilizados na entrevista do que nos programas de rádio, sendo que a variante *entendeu?* só foi utilizada neste contexto e foi a segunda variante mais utilizada na entrevista. Além disso, no contexto mais formal a quantidade de variantes dos RADs é maior, diferenciando-se da entrevista onde o *né?* e o *entendeu?* tiveram um número muito alto de realizações.

Alcântara (2015) pondera que a pesquisa realizada corrobora com o que Macedo e Silva (1996) e Freitag (2008) defendem em suas pesquisas sobre os marcadores, quando dizem que o uso deles não se trata de um vício da linguagem das pessoas sem instrução e/ou inexperientes, mas de uma necessidade da língua, pois os dados da pesquisa foram coletados a partir da fala de homens e mulheres com um tempo de atuação profissional considerado razoável, como é o caso de oito anos de experiência, em dois diferentes níveis de formalidade e, ainda assim, o uso destes itens foi muito recorrentes nos dois contextos.

Apesar de um número aparentemente pequeno das realizações de variantes não mencionadas em estudos anteriores, o uso das formas encontradas por Alcântara (2015) motivou a realização da presente pesquisa, pois nos propomos analisar, agora em uma comunidade de fala maior e que caracteriza de maneira mais próxima o falar do sertanejo, os RADs utilizados no sertão pernambucano, observando quais variantes desta variável e/ou fatores, linguísticos e sociais, que condicionam ou não a realização destes itens.

Para tanto, é necessário que tenhamos uma concepção acerca do objeto de análise que até a presente seção chamamos apenas de marcador. Nas seções seguintes, apresentamos a nomenclatura que iremos utilizar durante todo o nosso trabalho, da mesma forma que elucidamos o porquê desta escolha e a classe/grupo de marcador que nos propomos a analisar nesta dissertação.

2.2 A CONCEPÇÃO DE MARCADOR QUE ASSUMIMOS NESTE TRABALHO

Na seção anterior, apresentamos algumas pesquisas e as concepções de marcadores que cada autor assumiu para analisá-los. As pesquisas resenhadas abordaram os marcadores sob a nomenclatura de marcadores discursivos ou marcadores conversacionais, ou ainda alternaram entre as duas concepções, como fez Valle (2001).

Acreditamos que o pesquisador, ao seguir determinada nomenclatura de marcador, faz isso com base nos objetivos da pesquisa e no aporte teórico utilizado para analisar o fenômeno que será estudado, é certo que isso também nos motivou na escolha do termo marcador conversacional, os MCs. Entretanto, assumimos esta nomenclatura baseados também nos seguintes pontos:

(i) os MCs nomeiam, de maneira mais abrangente, os RADs, ou seja, os itens que pretendemos analisar, uma vez que consideramos-os como sendo elementos multifuncionais;

(ii) os marcadores recebem a nomenclatura de MCs quando são considerados como elementos portadores de propriedades intratextuais na cadeia linguística, ou seja, o seu uso é influenciado por regras linguísticas, algo que nos propomos analisar quando decidimos observar se variáveis linguísticas condicionam o uso dos itens que compõem a classe de RADs;

(iii) Os marcadores quando são subdivididos em grupos, em que cada um deles contém, itens diferentes, recebem, nos diferentes estudos realizados, a nomenclatura de MCs. E essa subdivisão permite que haja uma análise separada de cada item, algo que fazemos neste estudo, já que analisar todos os marcadores seria um trabalho tão extenso que fugiria do formato do presente estudo.

Reconhecemos que aderir esta abordagem traz algumas implicações teóricas e, conseqüentemente, metodológicas. Talvez, a principal delas seja a necessidade de se fazer uma junção de diferentes teorias que sejam capazes de analisar o fenômeno, como fez Valle (2001), quando recorreu ao Funcionalismo Linguístico e à Sociolinguística Variacionista. Outra implicação ao assumir a definição de MCs é a necessidade que o pesquisador tem de recorrer, para realização de suas análises, a diferentes teorias que servirão de base para seu estudo, pois, quando falamos em MCs, logo imaginamos que a Análise da Conversação, por exemplo, teria ferramentas pertinentes à análise do fenômeno em tela, tanto que Marcuschi utilizou esse suporte teórico em seu estudo citado anteriormente. Mas este não é o único

aporte teórico capaz de dar conta das análises feitas com os MCs, as pesquisas citadas na seção anterior mostram isso.

Macedo e Silva (1996) recorrem aos pressupostos da Sociolinguística Variacionista para analisar o uso dos MCs. É certo que, várias vezes, as autoras fizeram, mesmo que indiretamente, menção a alguns conceitos da Análise da Conversação, quando, por exemplo, elas falam de interação, manutenção de turnos etc. No entanto, não foi foco da pesquisa analisar estes itens sob este ponto de vista teórico, mas a partir do modelo teórico-metodológico da Sociolinguística.

Cientes de todas estas implicações, utilizamos a nomenclatura de MCs, tendo como base teórica os pressupostos da Sociolinguística Variacionista⁴¹. Não faremos uma incursão teórica acerca da Análise da Conversação, ou de qualquer outra teoria, porque nosso objetivo aqui é analisar se há na comunidade de fala outras formas atuando como RADs e como se dá a sistematização destas realizações⁴².

Os MCs foram, por Macedo e Silva, divididos, baseados em suas funções e posições sintáticas, em nove grupos ou categorias, sendo que, em cada um(a) deles(as), há itens que são específicos e outros que atuam em vários grupos ao mesmo tempo. Como já mencionado neste capítulo, propomo-nosa analisar apenas os MCs que fazem parte de um destes grupos, trata-se dos RADs – Requisitos de Apoio Discursivo.

Para tanto, apresentamos na seção seguinte o que é um RAD para a concepção de MCs, aderida neste trabalho, assim como o que as pesquisas realizadas com esta categoria, citadas anteriormente, abordam sobre eles e o porquê de escolher esta categoria para análise. No entanto, acreditamos que o pesquisador que se interessa em analisar os marcadores, mesmo aqueles, como é o nosso caso, que se propõem a analisar um dos grupos de marcadores, deve ter, logo no princípio da pesquisa, uma noção bem clara da metáfora da floresta virgem, proposta por Macedo e Silva (1996). Esta metáfora consiste em

compararmos o estudo dos marcadores ao de uma floresta virgem em que o número e tamanho das árvores e o emaranhado do bosque dificultam seu estudo e até mesmo sua penetração. Há que se fazer cortes, traçar caminhos e avançar com cautela, voltar sobre os passos e contornar obstáculos (MACEDO e SILVA, 1996, p. 14).

⁴¹ Apresentaremos os conceitos desta teoria no capítulo seguinte.

⁴² Ainda assim, não desprezamos a possibilidade de analisarmos este mesmo *corpus* sob o ponto de vista de outros aparatos teóricos. No entanto, para esta pesquisa utilizamos apenas a Sociolinguística Variacionista, deixando a Análise da Conversação, por exemplo, para estudos posteriores.

Com isso, fica mais fácil compreender que a pesquisa com marcador⁴³ deve passar por observações contínuas do fenômeno, tentativas de análises e reanálises, para que seja possível abrir um caminho onde se possa ter uma visão mais nítida do que acontece com eles, daí tentarmos dar ‘um passo’ nessa floresta virgem ao adentrarmos na área dos marcadores por intermédio dos RADs.

2.2.1 Os Requisitos de Apoio Discursivo

A categoria de RADs tem sido definida como “elementos de contato entre interlocutores, ocorrendo, quase sempre, no final de enunciado e tendo como principal função a manutenção do fluxo conversacional” (VALLE, 2001, p. 19). É certo que os RADs chamam a atenção dos pesquisadores por causa do importante papel que eles têm na interação entre participantes de uma conversa. Segundo Macedo e Silva, os RADs “ainda não perderam, em português, o seu papel interrogativo (...) e mantêm seu papel original de pedir a aquiescência do interlocutor” (MACEDO E SILVA, 1996, p. 17).

As inúmeras pesquisas desenvolvidas a partir do estudo dos RADs têm mostrado que: eles são realizados, na grande maioria das vezes, em frases interrogativas; o seu uso solicita, ao interlocutor, atenção para o que está sendo dito, por isso quanto maior for o trecho da conversa, mais RADs serão utilizados; há gêneros, como constatou Macedo e Silva (1996) que favorecem o uso dos RADs; eles constituem, como mostrou Valle (2001), uma variável linguística, em que a sua “propriedade interativa de requisitar apoio discursivo é comum a todos os itens, podendo corresponder àquilo que na pesquisa variacionista se entende por significado e permitindo, assim, a análise destes itens como variantes de uma mesma variável linguística” (VALLE, 2001, p.161); por fim, mas não menos importante, o *né?* parece ser um dos RADs mais utilizados, como apresentado nas pesquisas de Valle (2001) e Freitag (2008).

Partindo destas assertivas, escolhemos os RADs para analisar, tendo em vista que é uma categoria que, mesmo com tantos estudos realizados, ainda possuem muita área a ser explorada⁴⁴; porque hipotetizamos (i) que pode haver formas atuando como RADs no sertão pernambucano distintas das formas utilizados, por exemplo, no sul do país, variando, inclusive, das formas mais conhecidas como RADs, a saber, *né?*, *tá?*, *sabe?*, *entendeu?*,

⁴³ Afirmamos que isso se dá com pesquisas feitas com os marcadores, porém o mesmo acontece com toda pesquisa realizada no campo da linguagem. Damos ênfase aos estudos dos marcadores porque é nosso foco de análise.

⁴⁴ Utilizamos estes termos por que retomamos a metáfora da floresta virgem citada no final da seção anterior.

Viu?enão é mesmo?; assim como, (ii) que esta possível variação pode estar sendo motivada por fatores linguísticos e extralinguísticos.

Munidos destas afirmações, damos continuidade à pesquisa expondo, no capítulo seguinte, o aporte teórico-metodológico que nos serviu como base para realização do estudo apresentado nesta dissertação.

3 SOBRE A SOCIOLINGUISTICA

Para analisarmos o uso e as variações existentes entre os RADs, como nos propomos a fazer neste trabalho, faz-se necessário analisarmos a língua em uso, ou seja, em situações reais de comunicação, pois é nesse contexto que a variação linguística se manifesta. Sendo assim, recorreremos aos pressupostos da Sociolinguística, mais precisamente da Sociolinguística Variacionista ou Teoria da Variação e Mudança Linguística, como é rotulada por alguns pesquisadores. Neste capítulo, apresentamos algumas considerações sobre este modelo teórico-metodológico. Para tanto, dividimo-lo em três seções. Na primeira, após fazermos uma breve incursão histórica da Sociolinguística, apresentamos os seus pressupostos teóricos e alguns dos conceitos-chave desta teoria, dando ênfase à concepção de língua e ao objeto de estudo adotado pela teoria. Na segunda seção, abordamos a metodologia proposta pela Sociolinguística Variacionista, elucidando cada passo metodológico desta perspectiva teórica.

3.1 A SOCIOLINGUISTICA VARIACIONISTA

A Sociolinguística surgiu por volta do século XX, apresentando uma nova proposta de análise para os estudos linguísticos: a inserção dos fatores sociais no estudo da língua, algo que até então não era considerado pelos estruturalistas e gerativistas⁴⁵. William Labov é considerado o fundador desta teoria, pois foi ele quem, certamente inspirado por modelos teóricos anteriores, como afirma Tarallo (2007), insistiu na existência de uma relação entre língua e sociedade. Seus primeiros estudos, realizados em 1963, começaram a propagar este modelo teórico-metodológico e a despertar o interesse de outros pesquisadores, fazendo com que a Sociolinguística ganhasse um espaço entre as diversas teorias linguísticas.

William Labov realizou inúmeros e pioneiros estudos acerca da interferência de fatores sociais no uso da língua, assim como buscou mostrar que estes fatores podem ocasionar a mudança linguística. Tais estudos resultaram uma série de obras publicadas e que são, até hoje, referências para a Sociolinguística Variacionista, entre elas, destacamos: Labov

⁴⁵ O social para a Sociolinguística é totalmente diferente do social apresentado na teoria estruturalista. Para os sociolinguistas, os fatores sociais, tais como escolaridade, sexo, idade, podem interferir na língua e devem servir como parâmetros para o estudo da língua, mas, para Saussure e os pesquisadores estruturalistas, os fatores externos à língua não são levados em consideração para uma teoria do estudo da língua, para quem a língua é social, na medida em que é fruto de uma convenção estabelecida entre os integrantes de determinado grupo de uma sociedade. Além disso, nas pesquisas estruturalistas, os fenômenos variáveis são vistos na fala (individual) e não na língua (social). Caso o leitor deseje compreender melhor sobre esse assunto, que não é o foco principal desta pesquisa, recomendamos a leitura de Saussure (1977).

(1994), obra em que Labov apresenta, dentre outras, questões sobre mudanças e os seus mecanismos e a forma como estas mudanças são ou podem ser incorporadas no sistema linguístico; Labov (2001), o autor continua com as discussões acerca da mudança linguística, mas enfatiza como fatores sociais, como gênero, etnia, classe social etc., podem contribuir para a mudança linguística. Apesar de serem importantes obras, ainda não há traduções destes textos para a Língua Portuguesa. No entanto, Labov (1968) e (1972) possuem traduções e publicações em língua portuguesa em 2006 e 2008, respectivamente.

Em Labov (2006 [1968]) juntamente com Weinreich e Herzog, os autores apresentam, não a teoria propriamente dita, mas algumas bases para estudos históricos voltados para a mudança linguística, como o próprio título da obra sugere. Já em Labov (2008 [1972]), Labov discorre acerca de princípios teóricos, metodologia e compreensão da língua dentro desta perspectiva. Além disso, os resultados de alguns estudos realizados voltados para o vernáculo do inglês, como as pesquisas realizadas na ilha de Matha's Vineyard, Massachussets.

A partir destes estudos realizados por Labov, muitos pesquisadores têm mostrado um grande interesse pelos pressupostos da teoria, o que, sem dúvida, tem dado uma contribuição significativa para a Sociolinguística, tanto para a sua fixação no campo da Linguística, quanto para as novas gerações de estudiosos da língua que têm desenvolvido pesquisas com base neste modelo teórico. Falando em modelo teórico, apresentamos a seguir algumas das suas principais características, com o intuito de situar o leitor nos pressupostos básicos da teoria que utilizamos neste trabalho.

Comodito anteriormente, a Sociolinguística defende que há uma relação entre língua e sociedade, sendo a sociedade o lugar onde a língua se materializa. Isso se torna um pouco mais claro quando retomamos a explicação dada por Labov:

Crianças mantidas em isolamento não usam a língua; ela é usada por seres humanos num contexto social, comunicando suas necessidades, ideias e emoções uns aos outros. Os monólogos egocêntricos das crianças revelam ser desenvolvimentos secundários derivados do uso social da língua (Vygotsky 1962:19) e pouquíssimas pessoas passam muito tempo falando consigo mesmas. É inquestionável se frases que nada comunicam a ninguém façam parte da língua (LABOV, 2008 [1972], p. 215).

Percebemos com isso que, para Labov, a realização da língua se dá através do contato das pessoas com a sociedade, pois os seres humanos, integrantes de uma sociedade, fazem o uso da língua na e para a sociedade. Por isso, a língua utilizada pelas pessoas de determinado grupo social sofre influência deste grupo, ao passo que também o influencia: “a língua faz

parte de uma sociedade que a utiliza, a influencia e é influenciada por ela, o funcionamento dessa língua só pode ser entendido por meio da sociedade” (SANTOS e VITÓRIO, 2011,p. 13-14).

Além da relação entre língua e sociedade, no campo teórico da Sociolinguística, há outra discussão que merece destaque: a língua é heterogênea. Esse axioma é feito com base em uma série de implicações, dentre elas, destacamos aqui o fato de, mesmo possuindo regras categóricas ou invariantes⁴⁶, a língua também possui regras variáveis, que, por sinal, Labov(2008 [1972]) as considera um conjunto bem maior do que o conjunto de regras invariantes. Por isso, as regras variáveis precisam ser analisadas e sistematizadas. Com isso, fica evidente o porquê de a variação ser objeto de estudo e análise da Sociolinguística, a ponto de se tornarem conceito-chave desta teoria. Considerar que os fatores sociais interferem na língua, levando em consideração que esta é heterogênea faz com que a teoria Sociolinguística contribua com avanços significativos nos estudos da linguagem. Além disso, é através destes postulados que outros conceitos, também fundamentais, foram assimilados por este arcabouço teórico, estamos nos referindo aqui à variação linguística, à sistematização da língua, assim como, a um dos objetos de estudo do modelo teórico-metodológico: a língua falada⁴⁷, pois esta é heterogênea e diversificada. Tarallo (2007) afirma que:

a língua falada é heterogênea e variável; a variabilidade da fala é passível de sistematização. A língua falada é, portanto, um sistema variável de regras. Obviamente, a esse sistema de variação devem corresponder tentativas de regularização, de normatização(p. 57).

Resumimos os conceitos, até aqui apresentados, da seguinte maneira: a língua é objeto social, heterogêneo e, por isso, variável. Essa variação, no entanto, pode ser sistematizada através da análise do uso da língua em contextos reais de comunicação. Mollica (2003) resume da seguinte forma:

A Sociolinguística considera em especial como objeto de estudo exatamente a variação, entendendo-se como um princípio geral e universal, passível de ser descrita e analisada cientificamente. Ela parte do pressuposto de que as alternâncias de uso são influenciadas por fatores estruturais e sociais. Tais fatores são também referidos como variáveis independentes (...)(MOLLICA, 2003, p. 9-10).

⁴⁶ Entende-se por regras categóricas ou invariantes aquelas que não podem ser infringidas, pois, uma vez que não sejam cumpridas, impedem o entendimento do que foi falado ou escrito.

⁴⁷ A língua falada é objeto de estudo desta pesquisa, por isso mantemos este foco na discussão, mas ela não é o único objeto de estudo da Sociolinguística. Muitos pesquisadores, dentre eles, podemos citar Peixoto (2011), Santos (2011), Silva-Brustolin (2009), Vitório (2008), têm desenvolvido estudos sociolinguísticos a partir da língua escrita.

Até aqui, detemo-nos a falar da concepção de língua para a Sociolinguística e os conceitos básicos desta teoria, destacando a relação existente entre língua e sociedade, uma vez que foi a partir da compreensão dessa relação que a Sociolinguística definiu o seu objeto de estudo.

Mollica (2003) afirma que este modelo teórico “estuda a língua em uso no seio das comunidades de fala, voltando à atenção para um tipo de investigação que correlaciona aspectos linguísticos e sociais” (p. 9). Sendo assim, a Sociolinguística estuda a língua falada em comunidades de fala, entendendo ser esta uma maneira eficaz de se analisar como ela realmente é usada no dia-a-dia das pessoas, ou seja, o vernáculo, que é definido como sendo:

a enunciação e expressão de fatos, proposições, ideias (*o que*) sem a preocupação de *como* enunciar-los. Trata-se, portanto, dos momentos em que o mínimo de atenção é prestado à língua, ao *como* da enunciação. Essas partes do discurso falado, caracterizadas aqui como o vernáculo, constituem o material básico para a análise sociolinguística (TARALLO, 2007, p. 19).

Para a Sociolinguística, o material básico de análise não é o que um falante julga ser possível ou não em sua língua, os chamados dados introspectivos/intuitivos do gerativismo, mas o que de fato é produzido nos grupos sociais, no convívio com as pessoas cujo repertório linguístico possui uma certa regularidade em sua homogeneidade e heterogeneidade. A esse grupo social, a Sociolinguística dá o nome de comunidade de fala. Vale ressaltar, todavia, que definir comunidade de fala é algo bem mais amplo do que afirmar que se trata de um grupo de pessoas que interagem por meio da fala. Guy (2001) aponta duas funções da comunidade de fala para a teoria sociolinguística que nos auxiliam da compreensão acerca desta, uma vez que para ele, a comunidade de fala:

Fornece, em primeiro lugar, uma base fundamentada para explicar a distribuição social de semelhanças e diferenças linguísticas, a razão porque certos grupos de falantes compartilham traços linguísticos que os distinguem de outros grupos de falantes. Em segundo lugar, a noção de comunidade de fala fornece uma justificativa teórica para unir os idioletos de falantes individuais (que são os únicos objetos linguísticos cuja existência se pode realmente observar), em objetos maiores, as línguas (que são, na verdade, construções abstratas) (GUY, 2001 p. 2).

Sendo assim, a comunidade de fala evidencia a relação existente entre língua e sociedade, daí a importância desta para os estudos sociolinguísticos, pois “essa teoria pretende correlacionar

aspectos da língua e da sociedade, identificando os grupos de falantes que possuem características em comum” (VANIN, 2009, p.147). E é esse grupo de falantes, com características em comum, que autores como Guy (2001) chama de comunidade de fala.

Na mesma direção de Guy (2001), Vanin (2009) conceitua *comunidade de fala* como sendo um grupo formado por falantes “que compartilham traços linguísticos que distinguem seu grupo de outros, comunicam-se relativamente mais entre si do que com outros e partilham normas e atitudes diante do uso da linguagem” (p.148). Ou seja, parece ser que a comunidade de fala é um grupo de falantes que possuem suas peculiaridades linguísticas, tanto que Labov afirma que “os membros de uma comunidade compartilham um conjunto comum de padrões normativos” (2008 [1972], p. 225).

No entanto, a concepção de comunidade de fala não é tão unânime entre os linguistas, alguns estudiosos, como Severo (2008), não defendem que haja uma homogeneidade na comunidade de fala, ou seja, tenham tanto em comum assim, pelo contrário, para a autora, a comunidade de fala é heterogênea, pois é formada por pessoas que circulam em diferentes grupos sociais, pois ela até considera que a comunidade é um grupo de pessoas, até porque este conceito nos é passado assim pela área da sociologia, porém as pessoas que integram este grupo usam a língua de diferentes formas, constituindo-se, assim, um conjunto de pessoas que, de maneira heterogênea, compartilham alguns traços em comum.

Parece-nos pertinente tal concepção de a comunidade de fala ser heterogênea, uma vez que inúmeros estudos sociolinguísticos apontam variações dentro de uma mesma comunidade de fala e, sendo homogênea, a variação só seria notada se os estudos fossem comparar uma comunidade de fala a outra, o que nem sempre acontece, pois, dentro da própria comunidade de fala, a variação acontece, evidenciando o quanto ela parece ser heterogênea.

As considerações acerca de comunidade de fala, em total consonância ou não, apontam para o que Vanin (2009) pontua: “os sociolinguistas têm valorizado as relações ainda mais estreitadas da noção de comunidade de fala para níveis menores, como o de ‘rede social’ e de ‘comunidade de prática’” (p.152). Sendo assim, parece ser que as diferentes definições de comunidade de fala são resultados de uma mistura de conceitos de coisas distintas, a saber, comunidade de fala, comunidade de prática e rede social, portanto, cabe ressaltar as diferenças destes conceitos, para que cheguemos ao entendimento do que vem a ser cada um e de qual deles deve ser o foco em estudo Sociolinguístico.

Pereira (2017) elaborou um quadro, que apresentamos abaixo, com as principais distinções entre comunidade de fala, rede social e comunidade de prática e tais distinções nos

permitem compreender melhor cada um dos conceitos apresentados aqui. O pesquisador apresenta estas diferenças a partir da interação, relação de proximidade dos membros e do engajamento comum que cada conceito tem. Vejamos:

Quadro 4: Principais distinções entre comunidade de fala, redes sociais e comunidade de prática

Concepção	Sobre a Interação	Sobre a relação de proximidade dos membros	Sobre o engajamento comum
Comunidade de fala	Não é necessário existir interação entre os indivíduos selecionados.	Os indivíduos podem ser selecionados aleatoriamente, ou seja, não precisam, necessariamente, se conhecerem.	Não é necessário que os indivíduos estejam engajados em uma causa em comum.
Rede Social	É necessário haver laços de interação entre os indivíduos.	Os indivíduos precisam possuir uma relação de proximidade.	Não é necessário analisar indivíduos com causas comuns.
Comunidade de Prática	É necessário haver uma relação de interação constante entre os indivíduos.	Os indivíduos possuem um laço de proximidade	É necessário que haja um interesse comum entre os indivíduos.

Fonte: Pereira (2017, p. 72)

As diferenciações apresentadas no quadro mostram muito mais do que as singularidades de cada conceito, nos mostram também que são grupos de pessoas que podem ser objeto de estudo do pesquisador sociolinguista a depender de seu objetivo na pesquisa. Mais adiante, quando apresentarmos os métodos de coletas de dados, retomaremos que as comunidades de prática e/ou rede social podem ser mais úteis para a coleta de dados por meio de interações livres, do que a própria comunidade de fala.

Munidos das informações referentes ao que vem a ser uma comunidade de fala, é relevante apresentar o conceito de variação linguística, agora de maneira mais clara, visto que já o introduzimos anteriormente quando discutimos sobre língua falada e heterogeneidade. Segundo Mollica (2003), o principal objeto de estudo da sociolinguística é a variação linguística e, de acordo com Tarallo (2007), ela está presente em toda comunidade de fala. Resta-nos saber o que é variação linguística.

A variação linguística pressupõe que há formas linguísticas alternativas, a cada uma dessas formas chamamos de variantes, isto é, diversas maneiras de se dizer a mesma coisa com

o mesmo valor de verdade. Ao conjunto de variantes, chamamos de variável, “a variável linguística é, portanto, um conjunto de duas ou mais variantes. Estas, por sua vez, são diferentes formas linguísticas que veiculam um mesmo sentido” (BELINE, 2003, p.122). É preciso destacarmos que a noção de valor de verdade, apresentada por Labov (1978), chamada por ele de noção de significado representacional, diz respeito a um mesmo estado de coisas, ou seja, a formas alternantes que não contrastam, pelo contrário, elas dizem a mesma coisa. É essa noção, de valor de verdade, que nos permite considerar que os RADs podem ser considerados variantes de uma mesma variável.

É certo que alguns estudos, como o de Valle (2001) e Freitag (2008), mostram que cada forma de RAD possui uma função diferente, ou seja, ao escolher a forma *né?* em detrimento da forma *não é?* o falante o faz possivelmente motivado por inúmeras questões, dentre elas, podemos destacar, por exemplo, o contorno interrogativo, pois parece ser que o *né?* possui um contorno menos interrogativo do que o *não é?* (Cf. MACEDO E SILVA, 1996, p.16). No entanto, consideramos as diferentes formas de RADs como variantes linguísticas tendo em vista que elas são tão intercambiáveis a ponto de não serem realizadas concomitantemente (Cf. VALLE, 2001), ou seja, para finalizar uma sentença pedindo a aquiescência ou interrogando seu interlocutor, o falante pode usar sentenças como as que são apresentadas em (14), (15) ou (16), mas nunca sentenças como a exemplificada em (17).

(14) Eu acho, *né?* (Inf33 – TR)

(15) Tu acha, *hein?* (Inf47 – ST)

(16) Nem todo mundo acha, *sabe?* (Inf44 – ST)

(17) Eu acho, *né?/sabe?/Hein?**⁴⁸

No que diz respeito às variações linguísticas, é importante destacar que é por meio delas que ocorrem as mudanças na língua, pois toda mudança é precedida de variação, no entanto, nem toda variação gera uma mudança linguística (TARALLO, 2007).

Sendo assim, resta aos estudos sociolinguísticos identificar, dentro da comunidade de fala que está sendo analisada, quais são os fenômenos linguísticos que estão sofrendo variação e quais são as variantes destas variações. Feito isso, através do conhecimento da comunidade, o pesquisador deve identificar os tipos de variantes. Na teoria Sociolinguística Variacionista, as variantes recebem as seguintes nomenclaturas: *padrão* e *não padrão*; *prestígio* e *estigmatizada*; *conservadoras* e *inovadoras*. Frequentemente, mas não definitivamente, as

⁴⁸ Criamos esta sentença com o intuito de mostrar o quanto o uso das formas de RAD não é concomitante, uma vez que o uso destas formas ao mesmo tempo tornaria a sentença agramatical.

variantes padrão, prestígio e conservadoras são vistas como formas ‘corretas’ de uso da língua, ao passo que as variantes não padrão, estigmatizadas e inovadoras são tidas como desvios da língua e, por vezes, alvo de comentários jocosos e discriminatórios. Cabe ao pesquisador sociolinguista, analisar as variantes que, conforme a metáfora de Tarallo (2007), estão no campo de batalha, apontando para o que pode estar acontecendo com a língua falada pelas pessoas da comunidade de fala em questão.

Seguindo a metáfora do campo de batalha, proposta por Tarallo (2007), para explicitar os conceitos de variação, variável e variantes linguísticas, poderíamos afirmar que nos falta definir as armas que são usadas pelos combatentes no campo de batalha. Estamos nos referindo, evidentemente por meio de metáforas, à variável dependente e às variáveis independentes, ou aos grupos de fatores ou ainda aos condicionadores, como são costumeiramente denominados.

A variável dependente é o fenômeno variável que está sendo estudado e as variáveis independentes são os grupos de fatores que, podem ser externos ou internos à língua, condicionam a escolha entre uma e outra variante. Segundo Mollica (2003), a variável é tida como dependente, porque “o emprego das variantes não é aleatório, mas influenciado por grupos de fatores (ou variáveis independentes) de natureza social ou estrutural” (p.11).

A variação da língua não é aleatória, mas regulada por um conjunto de regras. Naro (2003, p. 15) ressalta que

[t]al como existem condições ou regras que obrigam o falante a usar certas formas e não outras, também existem condições ou regras mutáveis que funcionam para favorecer ou desfavorecer, variavelmente e com pesos específicos, o uso de uma ou outra das formas em cada contexto.

Trata-se, portanto, de aspectos internos e externos ao sistema ou também denominados de fatores linguísticos e sociais, respectivamente, que interferem no uso de uma variável em detrimento da outra. Entre os fatores linguísticos que podem interferir no uso de determinada variável estão: a categoria da palavra, os aspectos semânticos, a ordem dos sintagmas etc. E entre os fatores sociais, podemos destacar o gênero, escolarização, faixa etária, profissão etc. Para Labov (2008 [1972], p. 21), estes fatores sociais exercem grande influência sobre o uso, pois:

não se pode entender o desenvolvimento de uma mudança linguística sem levar em conta a vida social da comunidade em que ela ocorre. Ou, dizendo de outro modo, as pressões sociais estão operando continuamente sobre a língua, não de algum ponto remoto no passado, mas como uma força social imanente agindo no presente vivo.

Todos estes pressupostos discutidos até aqui fazem parte do aporte teórico da Sociolinguística Variacionista e permitem ao pesquisador chegar a um conhecimento mais preciso de como a língua é usada na comunidade de fala. Para que seja possível chegar a este conhecimento, e partindo da assertiva que a teoria está diretamente ligada à metodologia, esta corrente teórica apresenta um modelo metodológico a ser empregado nas pesquisas sociolinguísticas, pois “toda ciência – a linguística, em nosso caso particular – tem uma teoria própria, um objeto específico de estudo e um método que lhe é característico” (TARALLO, 2007, p.17). Por isso, fazemos, na seção seguinte, uma apresentação da metodologia Sociolinguística, que norteará os métodos a serem expostos no capítulo seguinte.

3.2 OS PROCEDIMENTOS METODOLÓGICOS NA PESQUISA SOCIOLINGUÍSTICA

As considerações apresentadas na seção anterior mostram que, do ponto de vista teórico, a Sociolinguística Variacionista possui alguns conceitos que lhes são peculiares, como os já citados, conceitos de variação, variantes, variáveis etc e outros, como o de língua, que difere dos apresentados por outros aportes teóricos⁴⁹. No entanto, a Sociolinguística possui uma metodologia de pesquisa de campo bastante peculiar e criteriosa para que os estudos acerca de variação linguística sejam realizados. As etapas metodológicas da pesquisa sociolinguística são basicamente: a escolha da comunidade de fala, a seleção dos informantes, a coleta de dados, a transcrição/codificação e análise dos dados coletados. Porém, em cada uma destas etapas, há pontos que, de tão importantes, destacamos nas subseções seguintes.

3.2.1 Escolha da comunidade de fala⁵⁰

A comunidade de fala, como mostramos na seção anterior, é um termo muito recorrente na Sociolinguística Variacionista. Nesta seção, enfatizaremos que a escolha desta é um importante passo metodológico na pesquisa Sociolinguística, ao ponto de podermos

⁴⁹ Afirmamos isso porque temos como base, o exemplo, já citado anteriormente, do conceito de língua como sendo social, tendo em vista que Saussure o tem de uma forma, Labov o tem de outra completamente diferente, conforme explicamos.

⁵⁰ Nomeamos esta etapa como escolha da comunidade de fala por esta ser mais recorrente nos estudos sociolinguísticos. No entanto, como destacamos na seção anterior, é possível realizar pesquisas a partir do aporte teórico-metodológico da Sociolinguística Variacionista com comunidades de prática ou com rede social. Por isso, entenda-se, aqui, escolha da comunidade de fala com uma etapa que corresponde ao mesmo que escolha da comunidade de prática ou rede social, guardadas, é claro as particularidades de cada grupo a ser escolhido e que, por sua vez, estejam de acordo com o estudo a ser realizado.

afirmar que seria este o primeiro e um dos mais decisivos passos na realização de uma pesquisa, pois os resultados alcançados com o estudo serão sempre retomados especificando-se a comunidade de fala escolhida, ou seja, ao afirmar, por exemplo, que determinada variável linguística parece estar em processo de mudança, uma vez que duas ou mais formas estão concorrendo entre si no sistema linguístico, isso é feito sempre especificando em que comunidade de fala isto está ocorrendo, tendo em vista que o estudo analisou a fala de informantes que são representativos da comunidade a que pertencem.

Para escolher a comunidade de fala, o pesquisador, que trabalha com dados linguísticos produzidos em situações naturais de uso, deve, primeiramente, identificá-la. Guy (2001) considera que há alguns critérios que podem nos ajudar a compreender um pouco melhor como identificar, antes de selecionar para o estudo, uma comunidade de fala. Coelho *et al* (2010, p.39) citam alguns deles: “i) os falantes devem compartilhar traços linguísticos que sejam diferentes de outros grupos; ii) devem ter uma frequência de comunicação alta entre si e iii) devem ter as mesmas normas e atitudes em relação ao uso da linguagem”. Sendo assim, ao selecionar a comunidade de fala, o pesquisador deve perceber que esta difere, por exemplo, de grupo de moradores de uma cidade ou bairro, pois, na verdade, a comunidade de fala pode estar, e na maioria das vezes até está, inserida em muitas cidades ao mesmo tempo⁵¹. A seleção da comunidade de fala se dá a partir da identificação de características que estes informantes possuem, características que vão além de simplesmente morar em determinado lugar.

Além destas assertivas, destacamos, com o intuito de defender a afirmação acerca da importância deste passo metodológico, que a escolha da comunidade de fala quase sempre chega a ser decisiva para a pesquisa, uma vez que os fenômenos linguísticos analisados no estudo podem até já terem sido examinados por outros pesquisadores sociolinguistas, mas, ao realizar um estudo em uma comunidade até então não selecionada para análises destes fenômenos específicos, o pesquisador traz para a pesquisa um caráter pioneiro, algo tão necessário quando o assunto é estudar a língua do ponto de vista científico, assim como, em seu processo de variação e mudança linguística.

3.2.2 Seleção de informantes

⁵¹ Se, por exemplo, a comunidade de fala de determinado estudo são pessoas de origem indígena que estão em fase de aquisição da Língua Portuguesa, estas podem estar em diferentes cidades/estados do país.

Após escolher a comunidade que será estudada, o sociolinguista deve selecionar os informantes. Tal passo é de extrema importância tendo em vista que é necessário que estes, para serem entrevistados, sejam, de fato, representativos da comunidade da fala a que pertencem, assim como possuam as características sociais pré-estabelecidas pelo pesquisador, características estas que a Sociolinguística Variacionista vai considerar como sendo variáveis extralinguísticas ou sociais, apresentamos na seção 3.1 desta dissertação a importância destas variáveis sociais para o presente estudo.

Seguindo a linha de que um estudo sociolinguístico variacionista apresentará dados de mudança e variação de determinado fenômeno que, como destacamos na seção anterior, pode até já ter sido analisado em outras comunidades de fala, mas a escolha de uma comunidade ainda não estudada pode tornar a pesquisa precursora, a seleção dos informantes é de igual modo importante, uma vez que, se o informante não representar legitimamente a comunidade a que pertence, os dados coletados e os resultados alcançados terão interferências significativas e resultados não fidedignos da realidade linguística da comunidade escolhida.

Como a maioria das comunidades de fala é composta por centenas, ou até mesmo milhares, de pessoas, é recomendável que “a seleção dos falantes seja aleatória” (SILVA, 2003, p.120). No entanto, há, segundo Silva (2003), dois importantes métodos aleatórios: o simples e o estratificado. No método aleatório simples, todos os indivíduos daquela comunidade terão igual probabilidade de serem escolhidos/sorteados⁵² para o estudo a ser realizado. Contudo, neste método é mais suscetível a realização de uma amostra enviesada, ou seja, a exclusão de informantes a depender do tipo de banco de informantes escolhido, a lista telefônica, por exemplo, deixaria excluído o participante da comunidade de fala que não tivesse telefone fixo a ponto de seu nome constar lá (Cf. SILVA, 2003) ou ficaria de fora do sorteio os informantes não conhecidos pelo pesquisador como pertencentes àquela comunidade de fala.

O método aleatório estratificado consiste em dividir a comunidade de fala em células⁵³ “compostas cada uma de indivíduos com as mesmas características sociais, procedendo-se posteriormente, para preencher cada casa, a uma seleção aleatória” (SILVA, 2003, p.121). Para a utilização do método aleatório estratificado é necessário que o pesquisador selecione, para a montagem das células, aspectos sociais relevantes para a pesquisa que será realizada, as

⁵² Usamos os termos “escolhidos” e “sorteados” tendo em vista que este método pode ser aplicado com base nestas duas formas. Quando escolhidos, os informantes podem ser selecionados a partir de listas telefônicas, catálogos de endereços, dados de senso etc. Quando sorteados, uma identificação de cada indivíduo é posta em um invólucro e vão sendo sorteados até completar o número de informantes desejados.

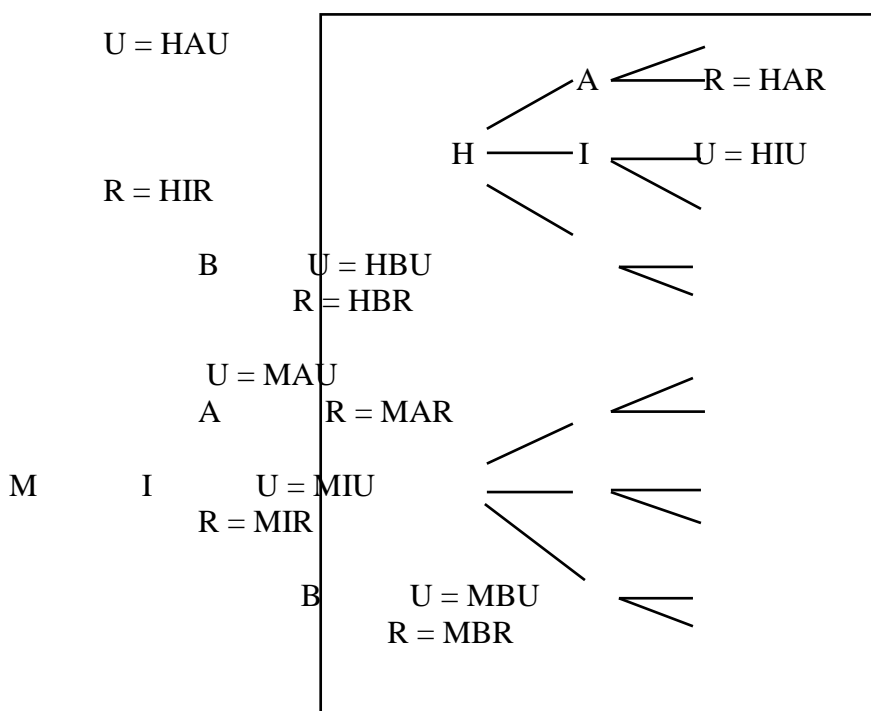
⁵³ Entende-se por células sociais a junção de diferentes fatores de cunho sociais que caracterizam o informante.

estratificações sociais são feitas, costumeiramente, seguindo as seguintes características: gênero, idade (faixa etária) e escolaridade (diferentes níveis de escolarização⁵⁴).

No entanto, há, a depender da pesquisa, outros aspectos sociais que podem ser, e são, utilizados para a seleção dos informantes. Talvez, os mais recorrentes, além dos já citados, sejam a localidade e o tempo em que o informante reside na localidade, o que evidencia, em muitos casos, o pertencimento do indivíduo à comunidade de fala em que o estudo está sendo realizado. Apesar de mencionarmos a localidade como sendo outro fator social a ser levado em conta para a seleção dos informantes, este não é o único, os objetivos de cada estudo, assim como, a comunidade de fala escolhida para tal, podem selecionar variáveis diferentes das citadas aqui que interferem na seleção dos informantes da pesquisa.

Um método interessante de ser utilizado para criação das células sociais é a montagem de um esquema, com letras que identifiquem cada fator social selecionado, pois ao multiplicar o número de fatores, por meio de uma análise combinatória, o pesquisador terá as células e o número delas a serem contempladas no estudo. Silva (2003) traz um exemplo de uma análise combinatória com fatores sociais como sexo (H = homem / M= mulher), 3 classes sociais (A, I, B) e 2 tipos de localidade (U = Urbana / R= Rural), como podemos ver abaixo:

Figura 1: Exemplo de esquema de análise combinatória dos fatores sociais



Fonte: Silva (2003, p.122)

⁵⁴ Há, neste tipo de estratificação social, duas maneiras diferentes de selecionar os informantes: i) nível de escolarização que remete ao ensino fundamental, médio ou superior; ii) tempo de escolarização que remete ao tempo que o informante teve com o processo de escolarização.

A partir da escolha dos aspectos sociais que irão compor as células que servirão para escolha dos informantes e o número destes para cada célula, o pesquisador obterá o número total de informantes de sua pesquisa. Esse é um ponto importante na etapa de seleção dos informantes, pois o pesquisador precisa encontrar um número que seja representativo e não repetitivo, pois um número muito pequeno de informantes para cada célula pode não ser representativo o suficiente, ao mesmo tempo em que um número muito grande de informantes pode ocasionar uma prolixidade nos resultados obtidos. Sendo assim, é necessário que o pesquisador selecione um número que seja ideal. Geralmente, tem-se a quantidade de cinco (5) informantes por célula como sendo ideal, porém este número não é fixo, o banco de dados do projeto Variação Linguística na Região Sul do Brasil – Varsul⁵⁵, por exemplo, tem a distribuição de dois informantes por célula social. A exposição da distribuição dos informantes por células é um procedimento interessante de o pesquisador finalizar esta etapa da pesquisa, pois ela evidencia os fatores sociais selecionados, o número de informantes para cada célula e número total de informantes da pesquisa. Abaixo, expomos a distribuição de informantes do estudo realizado por Valle (2001) que exemplifica bem o que acabamos de descrever.

Quadro 5: Distribuição de informantes do estudo realizado por Valle (2001)

Idade	Escolaridade					
	Primário		Ginasial		Colegial	
	Masc.	Fem.	Masc.	Fem.	Masc.	Fem.
15 a 21 anos	22		2	2	2	2
25 a 50 anos	2	2	2	2	2	2
Mais de 50 anos	2	2	2	2	2	2
	Total de 36 informantes					

Fonte: Valle (2001, p.57)

É importante mencionar que, nesta etapa da pesquisa sociolinguística, a seleção de informantes é feita a partir de dois momentos: o primeiro consiste em selecionar que tipo de informante participará da pesquisa, como mostramos anteriormente, e o segundo consiste em adentrar a comunidade de fala, localizar e solicitar aos informantes, com as características estabelecidas, que eles participem da pesquisa. Para este momento, o papel do pesquisador é

⁵⁵ Este banco de dados pode ser consultado a partir do endereço: <http://www.varsul.org.br/>

fundamental, uma vez que compete a ele não só a aproximação, mas o ato de convencer o indivíduo a participar do estudo como sendo este algo importante para a sociedade e para a produção científica e acadêmica. Afirmamos isso tendo em vista que o pesquisador será, possivelmente, alguém estranho para a comunidade, e, como se trata de uma pesquisa com informantes, é necessário, além da aprovação do Comitê de Ética Institucional, que o informante assine, date e insira seu número do Cadastro de Pessoa Física nos termos solicitados pelo comitê, algo que, para o informante, pode não ser tão simples de fazer a uma pessoa não conhecida em sua comunidade ou círculo de amizades.

Feita a seleção dos informantes, o próximo passo metodológico da pesquisa Sociolinguística é coletar os dados. Discorreremos acerca desta etapa na seção seguinte.

3.2.3 A coleta de dados

Ao coletar os dados para a pesquisa, o pesquisador promove uma consolidação no seu trabalho, ou seja, é a partir desta etapa que não só o estudo realizado por ele terá condições de ser desenvolvido, com as análises a serem desempenhadas, como inúmeros trabalhos futuros que objetivem analisar a comunidade escolhida poderão ser realizados a partir destes dados que foram coletados. Poderíamos afirmar que, neste ponto da pesquisa, o pesquisador sela alguns passos que não poderão mais ser alterados, como a estratificação social dos informantes e/ou a escolha da comunidade de fala, pois finalizou a montagem dos *corpora* de seu estudo, ao tempo em que abre um leque de inúmeras possibilidades de pesquisas futuras, ou seja, para a realização de uma pesquisa sociolinguística, nem sempre é necessário que o pesquisador vá até a uma comunidade de fala e realize a montagem de um *corpus*, como descrevemos até aqui, tendo em vista que, a depender do estudo a ser realizado, um banco de dados linguísticos já existente e que esteja disponível poderá ser usado, desde que seja devidamente autorizado. Além do banco de dados linguísticos do projeto VARSUL, que já citamos anteriormente, há o projeto NURC – *Norma Linguística Urbana Culta*, que disponibiliza um banco de dados linguísticos formado em 1970 com dados de fala de cinco capitais brasileiras, a saber, Porto Alegre, São Paulo, Rio de Janeiro, Salvador e Recife⁵⁶.

Além dos bancos de dados linguísticos já mencionados, Coelho *et al* (2010, p.127) adaptaram e ampliaram um quadro de banco de dados linguísticos citados por Gonçalves

⁵⁶ O projeto NURC do Rio de Janeiro disponibiliza os *corpora* para acesso através do site: <http://www.nurcj.lettras.ufrj.br/>.

(2009). Reapresentamos o referido quadro abaixo, entendendo que o conhecimento da existência destes *corpora* pode ser muito útil para realização de pesquisas futuras.

Quadro 6: Alguns *corpora* disponíveis para estudos na área

Projeto	Procedência	Tamanho da amostra (Número de informantes)	Variáveis sociais e controladas
Programa de Estudos sobre usos da Língua (PEUL)	Cidade do Rio de Janeiro	64	Sexo/gênero, faixa etária e escolaridade
Variação Linguística na Região Sul (VARISUL)	Região sul (12 áreas urbanas)	288 (24 por área)	Sexo/gênero, faixa etária e escolaridade
Variação Linguística na Paraíba (VALPB)	Estado da Paraíba	60	Sexo/gênero, faixa etária e escolaridade
Dialetos Sociais Cearenses	Fortaleza	23	Sexo/gênero, faixa etária, bairro e classe social
Língua Usada em Alagoas	Maceió	32	Sexo/gênero, faixa etária e escolaridade
Discurso e gramática (D&G)	Cidade do Rio de Janeiro, Natal (RN), Juiz de fora (MG) E Rio Grande (RS)	220	Sexo/gênero, faixa etária e escolaridade
Amostra linguística do Interior Paulista (ALIP)	Região Noroeste do estado de SP (7 áreas urbanas)	152	Sexo/gênero, faixa etária, escolaridade e renda familiar
Banco de dados Sociolinguísticos da Fronteira e da Campanha Sul-rio-grandense (BDS – Pampa)	21 municípios da faixa de fronteira (Brasil, Uruguai, Argentina) e campanha gaúcha	600	Sexo/gênero, faixa etária e escolaridade
Banco de dados Sociolinguísticos	Pelotas (RS)	90	Sexo/gênero, faixa etária e classe social

Variáveis por Classe Social (VarX)			
------------------------------------	--	--	--

Fonte: Coelho *et al*(2010, p.127)

É certo que o quadro acima contém importantes bancos de dados linguísticos. No entanto, além destes, gostaríamos de mencionar ainda outros dois *corpora* que, de igual modo, podem ser um importante material de estudos sociolinguísticos⁵⁷. É o caso do PHPB – Projeto *Para História do Português Brasileiro* que traz uma proposta de *corpus* diacrônico, uma vez que é formado por documentos escritos em diferentes séculos, do século XII ao século XX, em alguns estados e regiões do Brasil, podendo ser utilizado, sobretudo, para estudos diacrônicos⁵⁸. Já o Projeto ALiB – *Atlas Linguístico do Brasil*, por sua vez, se propõe a formação de um atlas geral do Brasil que impulse pesquisas em áreas como Sociolinguística e Dialectologia. Em 2014, no III Encontro Internacional de Dialectologia e Sociolinguística, foram publicados os volumes I e II deste importante material que trazem resultados de coletas de dados nacionais de vinte e cinco capitais de estado do Brasil (Cf. CARDOSO, *et al* 2014).

Os bancos de dados linguísticos são importantes opções para o pesquisador sociolinguista. Todavia, é válido ressaltar que nem sempre os já existentes atendem as necessidades do pesquisador e dos objetivos da pesquisa, por fatores como: a comunidade de fala escolhida, a estratificação dos informantes, os contextos de produção da fala/escrita, o período de coleta etc. Quando isso ocorre, é necessário que o pesquisador realize a coleta dos dados e, conseqüentemente, a montagem do *corpus* do estudo. Para tanto, o pesquisador, nesta etapa, terá papel fundamental, pois a maneira como os dados são coletados interfere nos resultados obtidos.

Seja qual for a natureza da situação de comunicação, seja qual for o tópico central da conversa, seja quem for o informante, o pesquisador deverá tentar neutralizar a força exercida pela presença do gravador e por sua própria presença como elemento estranho à comunidade. Tal neutralização pode ser alcançada no momento em que o pesquisador se decide a representar o papel de aprendiz-interessado na comunidade de falantes e em seus problemas e peculiaridades. Seu objetivo central será, portanto, aprender tudo sobre a comunidade e sobre os informantes que a compõem. A palavra ‘língua’ deverá ser evitada a qualquer preço, pois o objetivo é que o informante não preste atenção a sua própria maneira de falar (TARALLO, 2007, p. 21).

⁵⁷ Convém enfatizar que a utilização destes e de qualquer outros *corpora* deve ser precedida pela autorização do(s) organizador(es) do material.

⁵⁸ Os *corpora* de todas as equipes do PHPB estão disponibilizados neste site: <https://sites.google.com/site/corporaphpb/>

Por isso, não é suficiente que o pesquisador se aproxime do informante de qualquer forma, há maneiras para se fazer isto, assuntos que podem, ou não, ser abordados, posturas que merecem ser assumidas e, da mesma forma, outras que devem ser rejeitadas; ou seja, tudo deve ser feito para que, nesta etapa, os melhores resultados possíveis sejam alcançados. Para tanto, Silva (2003) considera que, na coleta de dados, entrevistado e entrevistador têm importante papel a ser desempenhado, pois cabe ao entrevistador ações que vão desde a escolha do tempo de duração da coleta, até o *tipo de contato escolhido*⁵⁹ para se chegar ao informante selecionado, ao entrevistado compete cooperar com a pesquisa de tal forma que venha falar “naturalmente”, assim como se permitir “tornar conhecido do pesquisador no início da pesquisa. O melhor seria ter um longo contato com ele, mas, na prática, torna-se difícil ter esse longo contato prévio e igual para todos os falantes da amostra” (SILVA, 2003, p. 126). Sendo assim, a coleta dos dados de uma pesquisa sociolinguística deve contar com a efetiva participação e colaboração do informante, assim como, a postura e as importantes decisões do entrevistador.

É certo que há métodos diferentes para realizar a coleta de dados que apresentamos até aqui, Silva (2003) afirma, em consonância com o que De Paula (2001) discorre, que há, pelo menos, três métodos diferentes de coletar dados⁶⁰: “interações livres, entrevistas ou testes⁶¹” (DE PAULA, 2011, p.34). Alguns estudiosos, no entanto, consideram que a entrevista é o procedimento mais recorrente. Labov (2008 [1972]) chega a afirmar que a gravação de entrevista é a forma de coletar dados mais satisfatória:

O método básico para se obter uma grande quantidade de dados confiáveis da fala de uma pessoa é a entrevista individual gravada. (...) Em seu conjunto, a entrevista é uma fala pública – monitorada e controlada em resposta à presença de um observador externo. Mas mesmo dentro dessa definição, o investigador pode se perguntar se as respostas numa entrevista gravada são ou não um produto especial da interação entre o entrevistador e o entrevistado. Um modo de controlar isso é estudar a pessoa em seu contexto social natural (LABOV, 2008 [1972], p. 63).

⁵⁹ Chamamos de *tipo de contato escolhido*, tendo em vista que pode ser um contato mais formal entre um pesquisador e alguém que coopera para a realização de uma pesquisa ou um contato mais informal que tem um foco maior em falas espontâneas de assuntos do cotidiano.

⁶⁰ Estes métodos para coleta de dados são, precisamente, utilizados para dados de fala. Quando a pesquisa intenta trabalhar com dados de escrita, a coleta se dá, basicamente, por meio de produções textuais, geralmente, narrativas sobre acontecimentos da vida dos membros da comunidade de fala.

⁶¹ No entanto, pesquisas como Alcântara (2015) mostram que a coleta de dados, a depender do estudo a ser realizado e da comunidade escolhida, pode ter outros métodos, uma vez que, para realização deste estudo, a pesquisadora, além de gravar entrevistas, gravou falas dos informantes durante as apresentações dos programas de rádio e estas fogem da classificação feita por estes autores, pois as falas dos radialistas em um programa de rádio não constituem em interações livres, uma vez que não há um interlocutor preciso para interação, não é uma entrevista, nem tão pouco um teste.

A entrevista requer, obviamente, a presença do entrevistador e do gravador e isso pode interferir na fala dos informantes. Essa situação é conhecida como o paradoxo do observador que, do ponto de vista de alguns teóricos sociolinguistas, pode interferir diretamente nos dados coletados, pois o informante pode, por exemplo, não se sentir a vontade para ter sua fala gravada ao ponto de usar a língua de maneira bastante monitorada, distanciando-se, assim, da fala natural e espontânea que o pesquisador sociolinguista busca coletar. É com base nisso que Santos (2009, p.70) afirma que “captar uma fala natural é um dos maiores desafios encontrados por um sociolinguista, uma vez que este faz uso de recursos tecnológicos como o gravador”. Para neutralizar o paradoxo do observador, há algumas recomendações: Labov (2008[1972]) recomenda que, por meio de pausas ou intervalos maiores, sejam rompidos os constrangimentos típicos de uma entrevista, de forma que o informante não se sinta entrevistado naquele momento. Para Tarallo (2007), uma maneira de erradicar o paradoxo do observador é, no decorrer da entrevista, solicitar ao informante narrativas de experiências pessoais, pois

os estudos de narrativas de experiência pessoal têm demonstrado que, ao relatá-las, o informante está tão envolvido emocionalmente com o que relata que presta o mínimo de atenção ao como. E é precisamente esta a situação natural de comunicação almejada pelo pesquisador-sociolinguista (TARALLO, 2007, p. 22).

Para que a coleta de dados por meio de entrevistas seja satisfatória, é necessário que ela seja desenvolvida com base em um assunto, de preferência, algo de interesse do informante, pois esta será uma maneira de garantir uma melhor participação do entrevistado, o que, certamente, pode influenciar nos dados coletados. Coelho *et al* (2010) retomam o posicionamento de Tarallo (2007) afirmando que a narrativa de experiências pessoais é o mais ideal nas entrevistas

Porque ao envolver o falante em tópicos que recriem emoções fortes vividas no passado (por exemplo, fazendo perguntas como ‘você já passou por uma situação em que correu perigo de morte? Como foi?’), o entrevistador faz com que o informante desvie a atenção de sua própria fala, deixando o vernáculo emergir. O falante deixa de prestar atenção no *como* diz para ficar atento a *o que diz*. Outros estímulos desse tipo podem ser: ‘conte um fato (história) que tenha acontecido com você que tenha sido muito engraçado (ou muito triste, ou muito constrangedor)’. O nome que se dá a esse formato específico de interação, cuja finalidade é a composição de um banco de dados para estudos sociolinguísticos, é entrevista sociolinguística (p.116).

É com base em considerações como esta que expomos que, a despeito de tudo o que já tratamos acerca da entrevista para a coleta de dados em uma pesquisa sociolinguística, a saber: i) é o método mais recorrente; ii) é uma das formas mais satisfatórias, pois por meio delas é possível solicitar ao informante as narrativas de experiências pessoais, o método entrevista possui, na pesquisa sociolinguística, características que lhe são peculiares, a ponto de ser necessário uma especificação do tipo “entrevista sociolinguística” conforme fizeram Coelho *et al* (2010), pois não se trata apenas de perguntas e respostas acerca de algo/alguém, mas de uma maneira de fazer surgir uma fala espontânea. Para tanto, as perguntas deste tipo de entrevista devem provocar a realização do fenômeno a ser estudado e, para que isso aconteça, é importante que o entrevistador monte um roteiro de entrevista, assim como fique atento durante a gravação a eventuais necessidades de adaptações que podem ser realizadas a qualquer momento.

Sendo assim, consideramos que duas colocações são importantes de serem feitas: é pertinente que o próprio pesquisador faça a gravação da entrevista, pois este é o mais habilitado a conhecer e provocar a realização, por meio das perguntas, do fenômeno a ser estudado. O roteiro da entrevista deve ser elaborado e conhecido com muita propriedade pelo pesquisador, pois isso pode fazer com que este fique mais à vontade, encontre-se menos preso ao roteiro, a ponto de passar mais segurança e ter mais naturalidade ao lidar com o informante, ao ponto de deixá-lo falar e não estar preso ao número de perguntas respondidas na entrevista.

As colocações expostas evidenciam, mesmo que de forma indireta, o quanto pode ser significativo para determinado estudo montar seu *corpus* de análise, uma vez que o pesquisador terá oportunidades de lidar com o fenômeno a ser analisado logo na gravação da entrevista, direcionando a realização deste pelo seu informante, algo que não é possível que aconteça quando o pesquisador lida com *corpora* já montados.

Ainda no que tange à entrevista sociolinguística e às perguntas que irão direcioná-la, Silva (2003) atenta para um fator importante: as perguntas que permitirão traçar o perfil social do informante, pois, por exemplo, “não é comum a pessoa dizer espontaneamente a sua situação econômico-financeira (correta)” (p.126). Além disso, respostas acerca de perguntas relacionadas a sexo/gênero, nível ou tempo de escolarização etc., trarão apenas informações sociais pontuais e, dificilmente, provocarão falas naturais do informante. Por isso, o método

de entrevista precisa ser acompanhado de um questionário⁶², escrito, que seja capaz de obter estas informações sociais que são pertinentes, sobretudo para possíveis comparações com os dados de fala obtidos.

Este questionário, também chamado de ficha social do informante, é significativo na etapa de coleta de dados, pois a criação e a utilização dele, na pesquisa sociolinguística, possuem alguns objetivos:

o questionário tem, como principal objetivo, traçar o perfil social do informante. Atribuímos outros objetivos a essa ficha social: ser um primeiro contato antes da entrevista propriamente dita é um papel relevante. ‘Quebrará o gelo’ do informante, familiarizando-se com o gravador. Outra vantagem, ainda, nesse contato prévio, nesse conhecimento do informante, é poder melhor conduzir a entrevista propriamente dita, já que se torna possível saber sobre seus interesses para melhor dirigir a conversa (SILVA, 2003, p.127).

Com isso, podemos considerar que, quando o pesquisador decide coletar os dados da pesquisa por meio da entrevista sociolinguística, é necessário que ele selecione algumas perguntas que sejam pertinentes para o questionário social e outras que sejam mais específicas para a entrevista, ou seja, que componham as perguntas do roteiro de entrevista. No entanto, esta seleção deve ser feita com muito cuidado, por parte do pesquisador, para que o questionário pessoal contemple as perguntas de tal forma que, para o roteiro de entrevista, não reste apenas perguntas sinônimas ou perguntas mais invasivas ao informante⁶³.

As colocações feitas até aqui mostram o quanto o método de entrevista sociolinguística para a coleta de dados é coeso e eficiente. No entanto, é possível que haja estudo com objetivos outros que a entrevista não seja uma forma eficiente para alcançar os dados necessários, é o caso, por exemplo, de pesquisas voltadas para a análise da conversação, uso de pronomes de tratamento, turnos de fala (Cf. SILVA, 2003), é por casos como este que o aporte teórico-metodológico sociolinguístico traz o método de interações livres, como citamos anteriormente, para a coleta de dados.

As interações livres é um tipo de coleta muito válido para as pesquisas sociolinguísticas, pois questões como o paradoxo do observador, que elucidamos

⁶² Não estamos com isso dizendo que o questionário social é usado apenas quando a coleta é feita por meio de entrevista, estamos destacando, na verdade, que, na entrevista, o questionário social é imprescindível.

⁶³ Afirmamos isso por considerar que perguntas como: *você é daqui mesmo?*, que corresponde a pergunta: *qual a sua naturalidade?* nos questionários sociais, podem ser bem mais úteis na entrevista do que na ficha escrita, uma vez pode servir para o entrevistador iniciar a conversa com o seu informante. Sendo necessária, com isso, a escolha, por parte do entrevistador, em que momento utilizar, descartando a ideia de repetição nos dois momentos ou deixando perguntas deste tipo apenas para a ficha social do informante, uma vez que, como dissemos, pode ser bastante útil no roteiro de entrevista.

anteriormente, são mais facilmente administradas, uma vez que este método consiste em gravar a fala de dois ou mais interlocutores interagindo simultaneamente, ou seja, o pesquisador seleciona pares de informantes, geralmente, esses pares são formados por pessoas que já se conhecem e/ou mantêm algo em comum, o trabalho do pesquisador consiste, neste caso, em dar um tema ou direcionar uma conversa e deixar que os informantes interajam entre si. Por tudo isso, parece ser mais fácil a gravação de uma fala mais espontânea e real, assim como, gravações mais extensas, se esta for uma prerrogativa do estudo a ser realizado⁶⁴.

Apesar de destacarmos algumas especificidades e vantagens das interações livres, é importante destacar que há alguns pontos que aparentemente não são tão vantajosos assim, pois, por exemplo, para a coleta de interações livres, o pesquisador precisará selecionar praticamente o dobro dos informantes, pois terá que formar pares para interação. Além disso, a formação destes pares deve ser orientada por grau de conhecimento entre eles e/ou, no mínimo, pontos em comum ao ponto de ser possível uma conversa natural, que flua entre os informantes, pois, dificilmente, dois ‘estranhos’ terão, em um primeiro contato, condições de desenvolver uma conversa capaz de ser significativa ao propósito da pesquisa.

Além disso, as interações livres podem também dificultar um passo metodológico seguinte, que discutiremos de forma mais ampla posteriormente, que é a transcrição dos dados. Isso se dá por motivos como nitidez e superposição de falas, uma vez que, numa interação entre duas ou mais pessoas, é normal que suas falas sejam sobrepostas umas às outras. Tal acontecimento não apenas dificultará a transcrição, como levará o pesquisador a desprezar alguns trechos que, de tanto ter falas sobrepostas, ficarão ininteligíveis.

Mesmo com estes pontos de desvantagens acerca da coleta de dados por meio das interações livres, reiteramos que este método pode ser satisfatório, a depender do estudo a ser realizado, e estes pontos podem ser contornados pelo pesquisador com estratégias como a intervenção durante a conversação, com perguntas do tipo (direcionando-se a um dos participantes) *então você acha que...*, pois, certamente, o outro informante irá ouvir e, em momento oportuno, discordar, concordar ou ampliar o que foi dito, porém depois de o informante requisitado pelo pesquisador se pronunciar. O projeto NURC – *Norma Urbana Culta*, que citamos anteriormente, possui, em seus *corpora*, dados coletados por meio de

⁶⁴ Acreditamos que o estudo que objetiva coletar dados por meio de interações livres deva considerar a possibilidade de trabalhar com comunidade de prática ou rede social, pois isso pode facilitar as interações entre os informantes. Afirmamos isso tendo em vista que, como expomos no quadro de Pereira (2017), na rede social e na comunidade de prática, os participantes possuem contatos a ponto de haver interação entre eles.

interações livres e os problemas aqui elencados foram, aparentemente, sanados pelos pesquisadores.

Dos três métodos citados para a coleta de dados, o que é utilizado de maneira mais específica é o teste, pois este é realizado pelo pesquisador para levantamento dos dados desejados. Ao seguir este modelo de coleta, o pesquisador irá disponibilizar para o informante o dado que ele deseja analisar, consultando o nível de aceitabilidade do dado. Sendo assim, o pesquisador, para montar o teste, terá que lidar com dados introspectivos, ou seja, algo que ele considera real e possível na língua, ou com resultados de estudos anteriores que utilizaram, inclusive, outros métodos de coleta.

O método de testes pode ser de produção ou percepção, ou seja, o pesquisador pode testar sentenças que sejam possíveis de serem produzidas ou a percepção do informante acerca de algo, como o significado das sentenças que estão presentes no teste. De acordo com Coelho *et al* (2003), os resultados destes testes podem divergir, eles afirmam isso ao citar que no

estudo realizado por Labov sobre o inglês vernacular (Black English Vernacular – BEV), teste de repetição da estrutura inglesa padrão I asked him IF he did it (‘Perguntei-lhe se ele fez isso’) mostram que o significado da sentença é percebido pelos falantes, mas que ela é produzida automaticamente como I axed him did he do it (COELHO *et al*, 2003, p.120).

Esta afirmação parece evidenciar que, no teste de produção, a variação é mais fácil de ser identificada/realizada do que no teste de percepção. No entanto, há estudos, como o de Tarallo (1985), em que os testes de produção e percepção se confirmam. Tais considerações apontam para a importância deste tipo de coleta de dados, a depender do foco do estudo a ser realizado.

Para facilitar a compreensão, criamos, nos quadros abaixo, um teste de percepção e outro de produção acerca do uso dos RADs⁶⁵. Por meio deles, será possível inferir diferenças entre um teste e outro, assim como, possíveis objetivos do pesquisador ao utilizar métodos como este.

Quadro 7: Modelo de teste de percepção para estudo com RADs⁶⁶

⁶⁵ Destacamos que não utilizaremos, no presente estudo, tal método de coleta, justamente por isso, criamos este teste para expor o método, uma vez que o modelo de coleta desta dissertação será, detalhadamente, exposto no capítulo seguinte.

⁶⁶ É evidente que um teste de percepção possui inúmeras questões, utilizamos aqui apenas uma questão, porque nosso intuito é exemplificar e mostrar que, em um teste de percepção, com questões como esta, inúmeras

Enumere as frases, de 1 a 4, de acordo com o grau de aceitabilidade, onde 1 deve ser utilizado para a frase mais aceitável; 2 para a frase mais ou menos aceitável; 3 para a frase menos aceitável e 4 deve ser utilizado se a sentença não for aceitável de forma alguma.	
() Aqui é muito bom, <i>né?</i>	() Eu sempre lutei, <i>né?</i>
() Aqui é muito bom, <i>hein?</i>	() Eu sempre lutei, <i>hein?</i>
() Aqui é muito bom, tá <i>certo?</i>	() Eu sempre lutei, tá <i>certo?</i>
() Aqui é muito bom, <i>viu?</i>	() Eu sempre lutei, <i>viu?</i>
() Minha família é de lá, <i>né?</i>	() Gosto dos amigos, <i>né?</i>
() Minha família é de lá, <i>hein?</i>	() Gosto dos amigos, <i>hein?</i>
() Minha família é de lá, tá <i>certo?</i>	() Gosto dos amigos, tá <i>certo?</i>
() Minha família é de lá, <i>viu?</i>	() Gosto dos amigos, <i>viu?</i>

Diferente do teste de percepção, o teste de produção pode ser organizado por meio da solicitação de que o informante produza, e não apenas enumere sentenças, como no teste de percepção, sentenças que lhes sejam aceitáveis, como é possível ver no quadro abaixo.

Quadro 8: Modelo de teste de produção para estudos com RADs

A partir das sentenças dadas, produza uma única sentença.⁶⁷
O homem é o único ser racional, <i>né?</i> O homem age como se não raciocinasse, <i>viu?</i>
_____.

Os quadros acima mostram modelos, de maneira bem resumida, de testes escritos a serem realizados para a coleta de dados. No entanto, atualmente, alguns pesquisadores têm aderido à realização de testes por meio de planilhas, rodadas em diferentes programas computacionais, como *Excel*, que permitem que o entrevistador consulte informantes de diferentes e longínquas localidades sem ter que realizar um deslocamento físico para tal.

A planilha eletrônica diminui alguns dos problemas citados nos métodos anteriores, como o paradoxo do observador e a sobreposição de vozes, nos métodos de entrevista e interação livre, respectivamente. Porém, selecionar o informante a partir do perfil social estabelecido para o estudo, conseguir meios de contatos, como *e-mail*, destes informantes são ações bem difíceis de serem realizadas. Além disso, a planilha com o teste pode ser

análises podem ser realizadas, tais como: o reconhecimento de diferentes formas de RADs; a escolha de uma forma em detrimento de outra como mais ou menos aceitável; o uso do RAD precedido por diferentes classes de palavras etc.

⁶⁷ A produção da sentença pode apontar para o pesquisador aspectos como: *na produção da sentença, o RAD foi mantido?; qual RAD foi escolhido em detrimento de outro?; em que sentença lugar da sentença ele foi inserido?* etc.

devidamente preenchida, após as etapas anteriores terem sido realizadas, mas, por ter sido preenchida sem a presença do entrevistador, não é possível garantir que, por exemplo, o informante tenha respondido sozinho, sem interferências de outros que não tenham o mesmo perfil social, ou seja, os dados linguísticos coletados ao serem cruzados com os dados sociais podem não apresentar resultados tão fidedignos como os esperados pelo pesquisador.

Segundo Labov (2008 [1972]), há, ainda, outra maneira de coletar dados de fala: “é observar o uso público da língua na vida diária fora de qualquer situação de entrevista – para ver como as pessoas usam a língua em contexto quando não existe observação explícita” (p. 63). Embora pareça ser um método eficiente, não é tão utilizado quanto à gravação de entrevista, talvez pelo fato de, ao observar e gravar a língua na vida diária do informante, seja necessário outro contato com os mesmos informantes para coletar informações extralinguísticas, demandando, assim, mais tempo e esforço para a realização da pesquisa.

Com os dados, sociais e linguísticos, coletados, os próximos passos de uma pesquisa sociolinguística consistem em dar um tratamento aos dados, que vão desde fazer um registro gráfico deles, por meio da transcrição, até quantificá-los e analisá-los estatisticamente. Esses passos são descritos na seção seguinte.

3.2.4 Transcrição, seleção, codificação e quantificação dos dados coletados

O primeiro passo a ser dado após a coleta de dados é a transcrição. É ela que vai permitir que o pesquisador tenha acesso a um material físico para a realização da pesquisa, pois, embora seja possível estudar a língua falada, é praticamente impossível realizar análises de fala sem que antes uma transcrição para a escrita seja feita. De Paula (2011) descreve que o objetivo da transcrição é “transpor o discurso falado para um registro gráfico de forma mais fiel possível” (p.36). Ao realizar a transcrição, o pesquisador armazena seu material de estudo, agora em formato de escrita, realizando um registro que pode ser utilizado para estudos futuros. Aliás, os grupos e/ou os projetos de pesquisa, como os já mencionados aqui, como, por exemplo, o VARSUL, chegam até este passo metodológico quando se dispõem a montar um *corpus* de análise: transcrever os dados coletados e armazenarem em bancos, eletrônicos ou não, para, *aposteriori*, disponibilizarem para a comunidade acadêmica realizar inúmeros e diferentes estudos. É interessante destacar que, mesmo que até aqui o pesquisador tenha recorrido a princípios metodológicos da pesquisa sociolinguística, os dados coletados e

transcritos podem ser utilizados em diferentes linhas e tipos de pesquisa. Citamos, como exemplo, estudos em Linguística Histórica, Dialectologia etc., que podem utilizar os dados *decorpora* assim montados para análises dentro destas perspectivas.

Dessa forma, o método de transcrição pode variar a depender da pesquisa que esteja sendo desenvolvida e da forma que os organizadores do banco de dados estejam seguindo. Sendo assim, podemos realizar uma transcrição ortográfica ou uma transcrição fonética⁶⁸. A transcrição ortográfica é a mais utilizada em estudos de língua falada. Geralmente, este tipo de transcrição é utilizado quando o pesquisador objetiva analisar aspectos sintáticos, semânticos ou lexicais, ou seja, a maneira como a palavra é pronunciada não é relevante neste tipo de transcrição. Por tudo isso, podemos destacar que é um modelo que “se, de um lado, compromete a fidelidade dos registros, tem, por outro lado, a vantagem de garantir maior legibilidade da transcrição” (PAIVA, 2003, p.137). Sendo assim, parece ser que a escolha do tipo de transcrição é mais motivada por questões relacionadas aos objetivos da pesquisa do que, simplesmente, o tipo da transcrição em si.

Para estudos relacionados à fonética/fonologia é necessário que seja feita uma transcrição fonética, esta, por sua vez, deve ser realizada a partir do alfabeto fonético internacional (IPA). E isso pode demandar um pouco mais de tempo e atenção para a transcrição, o que pode ser considerado uma dificuldade para o pesquisador, mas, ao mesmo tempo, um interessante modelo por contemplar não apenas o que foi dito, mas como foi dito, tendo em vista que a pronúncia é contemplada para a transcrição.

Mesmo com diferentes tipos de transcrição, o *corpus* montado poderá ser utilizado para diferentes pesquisas, pois, caso o banco de dados tenha apenas as gravações e transcrições fonéticas, por exemplo, e seja do interesse de um pesquisador utilizá-lo para uma análise sintática, tanto o áudio, quanto a transcrição fonética podem auxiliá-lo para a transcrição ortográfica. Caso tenha a transcrição ortográfica e o estudo seja fonético/fonológico, as gravações permitem que a transcrição fonética seja feita.

A transcrição dos dados conduz o pesquisador, em alguns casos, ao início da análise propriamente dita. Na verdade, ao transcrever os dados coletados, o pesquisador parece perceber realizações, aspectos e pontos que até mesmo durante as gravações não tenha dado conta, por isso, Paiva (2003, p.135) pontua que “podemos afirmar que a transcrição pressupõe uma pré-análise dos dados, na medida em que nosso posicionamento teórico

⁶⁸ Dentro destes sistemas de transcrição, convém destacarmos que há convenções realizadas para a marcação de, por exemplo, pausas, interferências, superposição de vozes. No entanto, tais símbolos são, geralmente, elucidados no referido *corpus*.

preestabelece, muitas vezes, a própria unidade de análise a ser considerada”. O que inicia neste ponto da pesquisa é a montagem do envelope de variação, ou seja, o pesquisador principia a percepção, assim como a descrição da variável a ser estudada. Segundo Coelho *et al* (2010), o “[e]nvelope de variação é o nome dado, em um estudo sociolinguístico, à descrição detalhada da 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” (p. 130).

O envelope de variação é comumente chamado de seleção de dados, que é o que de fato acontece, uma seleção. Afirmamos ser uma seleção, pois é muito provável que, após a transcrição dos dados coletados, o pesquisador perceba um número de fenômenos, ou até mesmo de particularidades de um único fenômeno, presentes no *corpus* montado e como o fazer científico pressupõe a delimitação de um objeto de análise, nada mais pertinente a ser feito, então, do que uma seleção. Scherre e Naro (2003) resumem o fazer desta etapa ao dizerem que

Após a delimitação do fenômeno a ser analisado, temos de proceder à identificação e ao levantamento dos dados relevantes. Antes de mais nada, aconselha-se que o pesquisador observe atentamente as fontes dos dados, com o objetivo de delimitar os dados relevantes, da forma que melhor lhe convier (p.153).

Assim, a seleção dos dados é uma maneira de afunilar o objeto de análise ao ponto de prepará-lo para um tratamento estatístico. Para tanto, é necessário que os dados selecionados sejam codificados de acordo com o que é aceitável pelos programas computacionais, pois “codificar é transformar em código identificável pelos programas computacionais disponíveis tudo o que queremos que seja quantificado” (SCHERRE E NARO, 2003, p.155). Percebemos, com isso, que a codificação dos dados consiste em organizar os dados transcritos, preparando-os para o tratamento estatístico que sempre é realizado em estudos sociolinguísticos, por isso é feita posteriormente à transcrição.

Uma ferramenta muito utilizada pelos sociolinguistas para quantificação é o pacote *Varbrul* (*Variable Rules*), que consiste em “um conjunto de programas computacionais de análise multivariada, especificamente estruturado para acomodar dados de variação sociolinguística” (GUY e ZILLES, 2007, p. 105). Com base nisso, a codificação é etapa fundamental, visto que auxilia na pesquisa, na medida em que gera os códigos que serão utilizados, pelo programa, para tratar estatisticamente os dados coletados que, em seguida, serão analisados/interpretados.

Para a codificação, no entanto, é necessário que durante a realização do envelope de variação, o pesquisador tenha selecionado as variáveis dependentes e as variáveis independentes, tanto sociais, quanto linguísticas. Feita a seleção, neste nível metodológico, cabe, apenas, atribuir um código a cada variável, porém, não é um procedimento tão simples assim, tendo em vista que: i) não é possível repetir os símbolos, ou seja, cada variável recebe um símbolo, que pode ser uma letra, maiúscula ou minúscula, números ou caracteres especiais, e isto deve ser caracterizador dela, a ponto de não ser utilizado em outra variável, sob pena de dificultar a leitura estatística apresentada ao fim da rodada computacional; ii) a variável dependente, por exemplo, por conter até mais do que três variantes, variável enéaria e, portanto, irá requerer três ou mais símbolos, quando somado aos símbolos das variáveis independentes gera um código muito extenso, facilitando, assim, a duplicidade ou ausência de algum símbolo, o que provoca erro ao rodar o programa.

Apesar das dificuldades elencadas, há ações que podem facilitar o trabalho de codificação, como, por exemplo, escolher um símbolo que possua alguma relação com a variável, assim como, inserir a sentença ao lado do código, pois, certamente, voltar à codificação, caso o programa apresente erros, por meio de *checkout*, é mais simples do que retomar ao próprio *corpus* para achar o erro. O próprio *checkout* é algo que facilita o trabalho do pesquisador, pois, por meio dele, o programa seleciona erros que devem ser corrigidos para que as rodadas computacionais sejam realizadas.

Seguidos estes passos, a pesquisa ganha agora um novo impulso, é o momento de analisar os dados coletados, transcritos, codificados e com os pesos relativos e as porcentagens de variação estabelecidas pelo programa computacional. É necessário, nesta etapa, que o sociolinguista apenas mostre a ‘quantificação’ das realizações de cada variante, dando unicamente tratamento estatístico, mas analise e interprete o que os números estão mostrando. Para isso, “é fundamental identificar conjuntos de circunstâncias linguísticas e sociais (restrições) que tendem a favorecer ou desfavorecer o uso de uma ou outra variante, assim se correlacionando o uso da variável dependente sob análise” (SCHERRE e NARO, 2003, p. 148). Dito de outra forma, é nesta etapa da pesquisa que os conceitos teóricos devem ser levados em conta para compreender como a língua é usada no grupo de pessoas que está sendo estudado.

Embora pareça que as etapas, e o que fazer em cada uma delas, estejam bem estabelecidas, é possível que, para atender aos objetivos da pesquisa, sejam necessários outros procedimentos metodológicos. Segundo Santos (2009), nem tudo está estabelecido, não há

uma receita pronta e é justamente por isso que “cada pesquisa exigirá os seus procedimentos metodológicos. (...) E é na realização desses procedimentos que o pesquisador perceberá se eles são ou não adequados para a pesquisa” (SANTOS, 2009, p. 71).

Por tudo isso, o capítulo seguinte traz a metodologia seguida para a realização do estudo apresentado nesta dissertação. Nele iremos detalhar outros aspectos não contemplados aqui, da mesma forma que irá exemplificar cada etapa metodológica, o que facilita a compreensão do trabalho realizado nesta pesquisa, assim como, o aporte metodológico da pesquisa sociolinguística apresentado neste capítulo.

4 METODOLOGIA

No presente capítulo, apresentamos a metodologia que seguimos para realização deste estudo que, como citado anteriormente, objetiva identificar as variantes existentes na variável RADs – Requisitos de Apoio Discursivo, uma das nove categorias dos MCs – Marcadores Conversacionais –, na fala de pessoas do Sertão do Pajeú. Embora estejamos baseados nos pressupostos teóricos da Sociolinguística Variacionista e, como já apresentamos no capítulo anterior, este aporte teórico possui um aparato metodológico que lhe é peculiar, destacamos que, para a realização de cada estudo, é necessário que o pesquisador faça adequações que lhe sejam necessárias a depender dos objetivos de seu trabalho. Citamos aqui, por exemplo, o envelope de variação, pois, para cada pesquisa, é imperativo que sejam realizadas especificações próprias a ponto de se tornar compreensível como o pesquisador chegou a cada concepção de variável e variantes apresentadas no estudo. E é fato que, até chegar a esta etapa, a pesquisa seguiu e realizou passos que, sem eles, não seria possível atingir a este ponto.

Sendo assim, apresentamos a metodologia de nosso estudo, expondo cada etapa realizada, fazendo com que, não apenas o método seguido seja exposto, mas o próprio estudo, pois, juntamente com os resultados alcançados com a pesquisa, que serão expostos *aposteriori*, a metodologia principia a apresentação do estudo, elucidando como ele foi realizado. As seções seguintes trazem a descrição da escolha da comunidade de fala, a seleção dos informantes, a coleta de dados, a transcrição, a codificação e a seleção dos dados a serem analisados.

4.1 A ESCOLHA DA COMUNIDADE DE FALA

Escolher uma comunidade de fala para fornecer os dados linguísticos que serão analisados em um estudo sociolinguístico é, como afirmamos no capítulo anterior, fazer um recorte importantíssimo para a pesquisa, pois a significância do estudo pode estar no fato de o fenômeno ainda não ter sido analisado, ou analisado de forma inicial, e a comunidade de fala deve realizar, na fala ou na escrita, tal fenômeno a ponto de fornecer condições (realizações) de o pesquisador analisar o uso real da língua e, mais especificamente, da realização de tal fenômeno, da mesma forma que, ao escolher uma comunidade de fala para analisar um fenômeno que já tenha sido amplamente estudado em outras comunidades, na comunidade de

fala agora escolhida, os resultados podem corroborar estudos anteriores, assim como trazer resultados bem destoantes dos que até então foram apontados por outros estudos, ou seja, a escolha da comunidade de fala, o primeiro passo metodológico a ser dado, é primordial não apenas para os próximos passos, mas, sobretudo, para os resultados alcançados, uma vez que estes serão sempre colocados como os dados presentes na comunidade de fala escolhida.

Antes de apresentarmos a comunidade de fala e o porquê desta escolha, é necessário que venhamos endossar que o presente estudo, tendo em vista seus objetivos, escolheu uma comunidade de fala e não uma comunidade de prática ou rede social para coletar os dados a serem analisados. A razão desta escolha é bem evidente: só a comunidade de fala nos daria condições de levantar formas que estejam atuando como RADs na fala das pessoas da localidade. A comunidade de prática ou a rede social, pelo fato de as pessoas terem, além de uma relação de interação, laços de proximidade, os RADs que porventura venham a utilizar podem ser específicos do grupo, de maneira mais restrita, e não da região – Sertão do Pajeú – que é o que nos propomos analisar no presente trabalho⁶⁹.

Com base nestas assertivas, destacamos que a escolha da comunidade de fala está diretamente ligada aos objetivos do estudo. Como no nosso caso aqui, a pretensão é observarmos e analisarmos as diferentes formas que estão operando como RADs, nada mais significativo do que escolher uma comunidade que até então não foi estudada no que diz respeito ao uso destes elementos linguísticos, a saber, o Sertão do Pajeú⁷⁰.

No entanto, julgamos pertinente mencionarmos que vários estudos linguísticos foram ou estão sendo realizados nesta comunidade, citamos aqui, por exemplo, a pesquisa realizada por Sedrins, Pereira e Siqueira (2015) em que os autores analisaram a *Variação na realização do artigo definido diante de antropônimos em dados de fala e escrita do Sertão de Pernambuco*, assim como o estudo que analisa *A variação na concordância nominal de número na língua falada no Sertão pernambucano*, realizada por Sedrins, Siqueira e Araújo (2015). Contudo, tais estudos, embora tenham sido realizados com dados de fala e escrita do sertão pernambucano, não estão voltados para o uso dos marcadores, como nos propomos analisar aqui. Irineu e Costa (2015) realizaram um estudo com os marcadores utilizados em Pernambuco, e não apenas no Sertão do Pajeú, todavia, o recorte feito no trabalho direcionou

⁶⁹ Contudo, não desprezamos a ideia de analisar o uso dos RADs em comunidades de prática ou rede social, comparando, inclusive, com os resultados obtidos nesta pesquisa em estudos futuros.

⁷⁰ Afirmamos que esta comunidade de fala ainda não foi estudada no que diz respeito ao uso dos Marcadores Conversacionais, mais especificamente dos RADs, tendo em vista que o estudo realizado por Alcântara (2015), embora tenha sido executado em cinco diferentes cidades do Sertão do Pajeú, a pesquisadora coletou dados de fala em uma comunidade de prática, ou seja, os radialistas do Sertão do Pajeú, o que não constitui como sendo a comunidade de fala do Sertão do Pajeú.

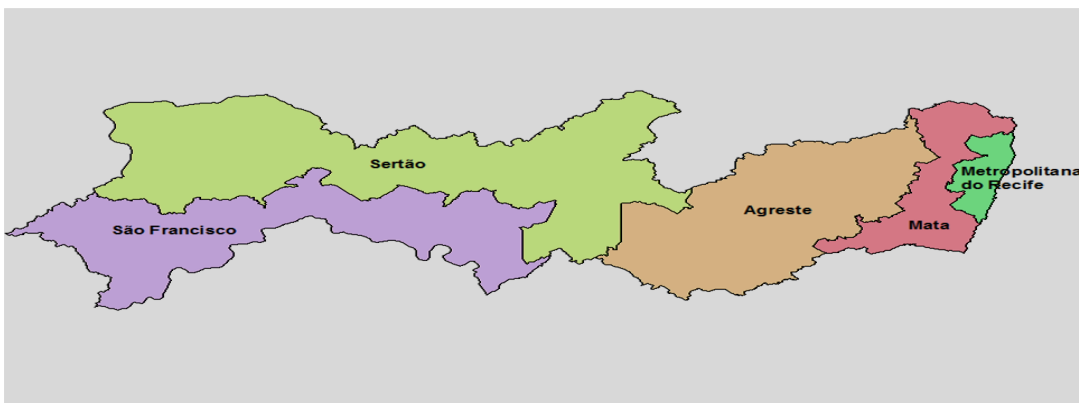
a pesquisa para dados de escrita nos comentários da seção "esportes" coletados no *blog* do Diário de Pernambuco, o que nos leva entender que os marcadores realizados são de uma comunidade específica, de pessoas que, além de terem acesso à *internet*, possuem interesse em esportes e ainda se predispõem a tecer comentários acerca do assunto e isso em muito se distancia da comunidade de fala da presente pesquisa.

Deste modo, a comunidade que escolhemos para realização desta pesquisa é o Sertão do Pajeú. O porquê desta escolha é exatamente o fato de nenhum estudo voltado para o uso dos marcadores e/ou dos RADs ter sido realizado nela, evidenciando, assim, a importância deste estudo. Na seção seguinte, apresentamos a comunidade de fala selecionada, apontando desde a localização geográfica até aspectos de cunho econômico e social.

4.1.1 O sertão do Pajeú

O estado de Pernambuco é formado por cinco mesorregiões, a saber: Metropolitana do Recife, Zona da mata de Pernambuco, Agreste pernambucano, Sertão Pernambucano e São Francisco Pernambucano, conforme pode ser visto no mapa abaixo.

Figura 2: Mapa das mesorregiões do estado de Pernambuco



Fonte: <http://www.baixarmapas.com.br/mapa/estado/pernambuco/>

Essas mesorregiões são subdivididas em regiões menores de acordo com as características geográficas, tais como: clima, relevo, localização. Assim sendo, as cinco mesorregiões de Pernambuco são divididas em 12 microrregiões, conforme podemos ver no mapa abaixo.

Figura 3: Mapa das 12 microrregiões do estado de Pernambuco



Fonte: <http://www.investinpernambuco.pe.gov.br/web/invest/regioes-de-desenvolvimento>

O Sertão do Pajeú é uma das doze regiões do estado de Pernambuco e é esta a região selecionada aqui como uma comunidade de fala para o nosso estudo. Esta região é formada por 17 cidades: Afogados da Ingazeira, Brejinho, Calumbi, Carnaíba, Flores, Ingazeira, Igaracy, Itapetim, Quixaba, Santa Cruz da Baixa Verde, Santa Terezinha, São José do Egito, Serra Talhada, Solidão, Tabira, Triunfo e Tuparetama. Estas cidades possuem, justamente pela sua divisão geográfica, climas semelhantes, assim como características de relevo.

Figura 4: Mapa das cidades que integram a região do Sertão do Pajeú



Fonte: Versycle *et al* (2015, p.18)

Em estudo acerca do clima, economia e desenvolvimento desta região, Versycle *et al* aponta que

a microrregião apresenta clima quente com períodos longos de estiagem e as chuvas irregulares variando entre os meses de janeiro e maio, mas, mesmo assim, a economia local se destaca na agropecuária, sendo a atividade mais desenvolvida no sertão brasileiro (2015, p. 16).

Se os municípios que integram a região do Sertão do Pajeú, conforme destacamos, possuem semelhanças entre si, no que diz respeito às questões geográficas, diferem muito no que diz respeito ao desenvolvimento. Há municípios desta região, como Calumbi, segundo estimativa do IBGE (2017), com uma população de 5.736⁷¹, ou seja, municípios bem pequenos, que não dispõem de uma infraestrutura básica, como faculdades, centros médicos etc. No entanto, há municípios, como Serra Talhada, que se destacam não apenas pelo contingente populacional, estimativa de 85.568 mil habitantes, mas pelo desenvolvimento na área de saúde, educação e economia.

Sendo assim, a comunidade de fala escolhida para o presente estudo é bem mista e ampla, com 17 municípios ao todo, necessitando, por isso, de uma seleção mais específica para a realização do estudo, uma vez que coletar dados em todas as cidades desta região seriaprovável, além de um imenso trabalho, obtermos uma repetição de dados de uma maneira muito acentuada, tendo em vista que parece que os informantes dividem algumas normas da língua entre si ao utilizá-la em alguns espaços em comum⁷².

Portanto, selecionamos, dentre os 17 municípios do Sertão do Pajeú, três dos mais importantes municípios da microrregião. A escolha destes municípios não foi aleatória, na verdade, consideramos estas cidades, a saber: Serra Talhada, Triunfo e São José do Egito, como representativas desta comunidade, tendo em vista seu desenvolvimento econômico, expansão geográfica, crescimento populacional, entre outros. Os municípios citados são apresentados nas seções seguintes.

4.1.1.1 Serra Talhada

Conhecida como uma das mais importantes cidades da região do Sertão do Pajeú, Serra Talhada, distante aproximadamente 415 km da capital do estado, Recife, está entre as

⁷¹ Dado consultado em <https://cidades.ibge.gov.br/brasil/pe/calumbi/panorama>.

⁷² Os membros desta comunidade de fala utilizam a língua em espaços em comum a partir das necessidades que surgem exatamente pela questão do desenvolvimento, como pontuamos anteriormente, do município. Como há cidades como Calumbi, Tabira, Solidão entre outros que não possuem uma estrutura suficiente para seus moradores, as pessoas destas cidades acabam tendo que se descolar, até mesmo diariamente, para trabalhar, estudar, realizar procedimentos médicos, bancários, entre outros. Sendo assim, há um trânsito significativo das pessoas de um município para o outro.

maiores cidades da microrregião, com uma população de aproximadamente 85 mil habitantes, conforme mencionamos anteriormente por meio dos dados do IBGE. Nos últimos anos, com o advento de universidades, centros médicos e ampliação do comércio, a cidade tem experimentado um crescimento significativo. Conta, atualmente, com inúmeras instituições de Ensino Superior, presenciais e na modalidade a distância com pólos de atendimento ao discente e tem se destacado quando o assunto é educação, uma vez que dados do último censo do IBGE mostram que o município teve, em 2010, uma taxa de 96.7 % de escolarização de 6 a 14 anos de idade. Em 2015, o IDEB do município ficou com média 4.1 para os alunos dos anos finais do Ensino Fundamental.

Por sediar inúmeras Instituições de Ensino Superior, em diferentes modalidades, pública, estadual e federal, privadas e educação a distância, Serra Talhada hospeda um grande número de estudantes das cidades circunvizinhas, é o caso de, por exemplo, Santa Cruz da Baixa Verde, Triunfo, Calumbi, Flores, Afogados da Ingazeira, Tabira, entre outras.

No que tange à educação e pesquisa, Serra Talhada tem sido destaque, após a implantação da Universidade Federal Rural de Pernambuco – Unidade Acadêmica de Serra Talhada, a pesquisa, o ensino e a extensão tiveram um importante impulso nessa área. As pesquisas na área de língua/linguagem têm sido realizadas frequentemente, a exemplo, do próprio estudo que apresentamos nesta dissertação, pois este surgiu a partir de pesquisas iniciais realizadas em 2015 nesta instituição de ensino.

Contudo, este município não se destaca apenas nas questões relacionadas à educação, pessoas de diferentes municípios do Sertão do Pajeú transitam por Serra Talhada em busca de atendimento médico, tendo em vista que a cidade é considerada um polo médico no sertão do estado. Esse destaque vem crescendo desde a implantação do curso de Medicina pela Universidade do Estado de Pernambuco – UPE.

Outro ponto que faz com que Serra Talhada tenha um afluente de pessoas sertanejas transitando em seu território é a cultura local. Conhecida como a terra de Lampião, a cultura serratalhadense é voltada para as questões do cangaço e de seu precursor, Virgulino Ferreira – Lampião. A cidade explora bastante as questões relacionadas ao cangaço, sobretudo, por meio de exposições e atividades culturais realizadas no Museu do Cangaço, que recebe, constantemente, visitantes de inúmeros lugares que procuram o espaço para conhecer mais da história do município e do cangaço como um todo, uma vez que estas se entrelaçam.

Figura 5: Museu do Cangaço em Serra Talhada



Fonte: Arquivo pessoal da autora

Anualmente, a secretaria de cultura do município, juntamente com a direção do Museu do cangaço, realiza um espetáculo ao ar livre acerca da morte de Lampião. Intitulado *O massacre do Angico*, o espetáculo recebe pessoas de inúmeras regiões do estado, mais especificamente, dos sertanejos circunvizinhos.

Figura 6: Cenário ao ar livre para apresentação do espetáculo



Fonte: Arquivo pessoal da autora

Figura 7: Apresentação do espetáculo O massacre do Angico



Fonte: <http://g1.globo.com/pe/caruaru-regiao/noticia/o-massacre-de-angico-sobre-vida-de-lampiao-e-encenado-em-serra-talhada.html>

Partindo destas considerações, entendemos que a cidade de Serra Talhada é um importante município da região do Sertão do Pajeú e é capaz de ser representativa da comunidade de fala escolhida para ser analisada neste estudo linguístico, tendo em vista que é uma cidade que recebe visitantes de outras cidades da região e seus habitantes possuem um repertório linguístico capaz de estabelecer uma comunicação e compartilhar marcas, como o uso dos marcadores conversacionais, entre si. Outra importante cidade e que foi selecionada aqui como sendo capaz de representar a região do Sertão do Pajeú é Triunfo, os dados deste município e o porquê de sua escolha são apresentados na próxima seção.

4.1.1.2 Triunfo

O município de Triunfo fica a aproximadamente 400km da capital Recife, com uma população pouco maior que 15 mil habitantes, 15.221, mais precisamente, segundo dados do IBGE em 2017. Mas não é pelo número de habitantes ou extensão territorial que Triunfo tem se destacado, pois está entre os municípios menores da região, mas sim pelo turismo local, motivos estes que atraem visitantes de outras cidades da região para a localidade.

Conhecida como o "oásis do sertão", Triunfo “possui o ponto mais alto do estado com mil duzentos e sessenta metros, tem um microclima muito especial procurado pelos turistas pelas suas condições climáticas diferenciadas no contexto semiárido” (VERSYCLE *et al*, 2015, p.19). Sendo assim, Triunfo, por ser uma cidade favorecida por questões

climáticas, como por exemplo, a temperatura, pois por ser uma cidade serrana conta com baixas temperaturas praticamente durante todo o ano, conta com um transitar significativo de pessoas que não vão à cidade por questões relacionadas à educação e à saúde, como destacamos ser o caso de Serra Talhada, mas para o turismo e diversão.

Figura 8: Visão parcial da cidade de Triunfo



Fonte: <https://www.guiadoturismobrasil.com/cidade/PE/653/triunfo>

Figura 9: Pico do Papagaio - ponto mais alto do estado de Pernambuco localizado na cidade de Triunfo



Fonte: <http://blogdescalada.com/quais-sao-os-pontos-culminantes-de-cada-estado-brasileiro/>

Se o clima atrai visitantes à cidade, alguns aspectos culturais são responsáveis pelo entretenimento destes, dentre eles, destacamos a apresentação dos *caretas*⁷³ em ocasiões de fluxo alto de turistas, como no período carnavalesco, o Festival de inverno que acontece anualmente na cidade e o festival de cinema realizado e apresentado em um dos prédios que é considerado patrimônio histórico do município, o teatro Guarani.

Figura 10: Apresentação dos caretas em frente ao teatro Guarani



Fonte: <https://cultura.pe.gov.br>.

A imagem acima nos mostra algo que também é muito peculiar da cidade: a arquitetura urbana. A cidade conta com prédios históricos que, além de atividades culturais, são realizadas atividades de cunho religioso, como é o caso dos conventos históricos espalhados pela cidade.

A cidade conta com instituições de ensino fundamental e médio, mas o Ensino Superior ainda não é ofertado, gerando com isso, um deslocamento diário de seus habitantes para cidades próximas, como Serra Talhada, por exemplo, distante cerca de 35 km. Sendo assim, embora próximas uma da outra, o fluxo de pessoas do sertão pernambucano nestas cidades, além de intenso, é motivado por questões distintas, aumentando, assim, o transitar das pessoas do Sertão do Pajeú nestas cidades, assim como, a possibilidade de existirem variações linguísticas entre as cidades desta região.

⁷³Caretas é o nome dado aos personagens que são montados, sobretudo em períodos carnavalescos e ou períodos de alta temporada de turistas na cidade. A montagem dos personagens, chamados de caretas, é feita a partir de chocalhas, placas com sátiras, um chicote que eles utilizam para estalar e chamar a atenção das pessoas, assim como das máscaras com semblante triste que não permitem que o fantasiado seja identificado pelas pessoas que assistem sua apresentação.

As considerações acerca da cidade de Triunfo mostram o quanto este município pode ser representativo da região do Sertão do Pajeú, pois, embora tenha uma população pequena e um espaço geográfico não muito extenso, cerca de 181,4 Km², segundo dados do IBGE (2017), é um município que oferece atrações turísticas e culturais para as pessoas da região, se tornando para muitos uma opção de passeio e lazer.

Por tais questões, a cidade de Triunfo integra o grupo de cidades escolhidas para serem representativas da comunidade de fala aqui analisada, o Sertão do Pajeú. Outra cidade bastante importante desta região é a cidade de São José do Egito, o que motiva a escolha desta é exposto na seção seguinte.

4.1.1.3 São José do Egito

O município de São José do Egito é conhecido em todo o estado de Pernambuco, e até mesmo fora dele, como sendo o berço imortal da poesia. Das três cidades escolhidas, São José do Egito é uma cidade mediana no que diz respeito à extensão e a localidade, é o terceiro município, segundo o IBGE, mais populoso da região sertaneja⁷⁴.

Na extensão territorial, é um município pouco maior que Triunfo e um pouco menor que Serra Talhada, São José do Egito possui uma área territorial de 794,143 Km², com um número de 33.704 pessoas de acordo com o censo de 2017 do IBGE⁷⁵. Esse número de habitantes não conta, obviamente, com o número de pessoas que transitam na cidade, pois São José do Egito fica localizada entre alguns municípios do Sertão do Pajeú, a saber: Tuparetama, Itapetim, Brejinho. Santa Terezinha, Ingazeira e Tabira. Estes municípios são menores, a ponto de serviços bancários e comércio em geral, por exemplo, só serem oferecidos a população na cidade de São José do Egito, ocasionando, com isso, um fluxo significativo de pessoas circulando pela cidade.

Apesar de receber um número considerável de pessoas, os egipsienses, esse é o gentílico dos habitantes desta localidade, também precisam ir até outros municípios do Sertão do Pajeú em busca de estudos, pois, embora conte com um ensino de qualidade, a taxa do IDEB do município é cerca de 4.3 para os anos finais do ensino fundamental, conforme IBGE (2017), maior, inclusive, do que a taxa do IDEB de Serra Talhada, a cidade não possui um ensino superior que atenda às demandas da população, uma vez que oferece apenas alguns

⁷⁴ Segundo o IBGE, o município mais populoso é Serra Talhada, seguido de Afogados da Ingazeira e, em terceiro lugar, São José do Egito.

⁷⁵ Tais dados foram obtidos em: <https://cidades.ibge.gov.br/brasil/pe/sao-jose-do-egito/panorama>

poucos cursos a distância. Sendo assim, as pessoas deslocam-se para as cidades de afogados da Ingazeira ou Serra Talhada para cursarem uma graduação.

Outra razão que leva a movimentar a cidade, com um número significativo de pessoas, é a cultura local. Esta é notória desde que o indivíduo adentra a cidade. Como podemos ver na imagem do portal de entrada de São José do Egito.

Figura 11: Portal de entrada da cidade de São José do Egito



Fonte: <https://www.ferias.tur.br/cidade/5443/sao-jose-do-egito-pe.html>

A viola, os músicos e a saudação de acolhida mostram que a cultura do lugar está relacionada à música e à poesia. Anualmente, a cidade realiza o *Festival de Cantorias*, uma festa cultural, que conta com apresentação de violeiros, repentistas e poetas da cidade e de municípios vizinhos em apresentações que são ovacionadas pelo público.

A arte de cantar repentes, tocar violas e fazer poesias é algo inerente às pessoas da localidade, de crianças a idosos. Todos gastam boa parte do tempo ocioso da vida interiorana fazendo ou apreciando estas artes que identificam a cidade de São José do Egito. A viola e o cantador se tornaram tão famosos na localidade, que uma escultura foi posta em um ponto estratégico, mostrando aos visitantes o quanto esta arte é propagada no município.

Figura 12: Escultura posta no meio da cidade que enaltece a viola e o cantador egipsiense



Fonte: <https://www.ferias.tur.br/cidade/5443/sao-jose-do-egito-pe.html>

Por tudo isso, notamos que a cidade de São José do Egito recebe diariamente um vasto público de sertanejos, assim como, seus habitantes também circulam em outros municípios do sertão pernambucano em busca de estudos e para propagar a arte produzida na cidade, fazendo com que o egipsiense possua uma comunicação frequente com pessoas de diferentes municípios, motivo que nos leva a reconhecer esta cidade como representativa da comunidade de fala do Sertão do Pajeú.

Os dados expostos acerca da região do Sertão do Pajeú, bem como, das cidades escolhidas para representar esta comunidade, mostram algumas das principais características da comunidade de fala a ser analisada na presente pesquisa. Na seção seguinte, apresentamos a seleção dos informantes feita nesta comunidade de fala.

4.2 A SELEÇÃO DOS INFORMANTES

Ao realizar um estudo sociolinguístico, grande parte dos passos metodológicos que o pesquisador deve seguir são realizados por meio de seleção, como é o caso da seleção da comunidade de fala, da seleção dos informantes e da seleção dos dados. Na seção anterior, descrevemos como a seleção da comunidade de fala foi realizada e, tendo selecionado-a, é necessário selecionar os informantes que irão participar da pesquisa.

A escolha dos informantes deve ter critérios, pré-estabelecidos pelo pesquisador, que permitam selecionar indivíduos que representem a comunidade de fala (Cf. SILVA, 2003). Os

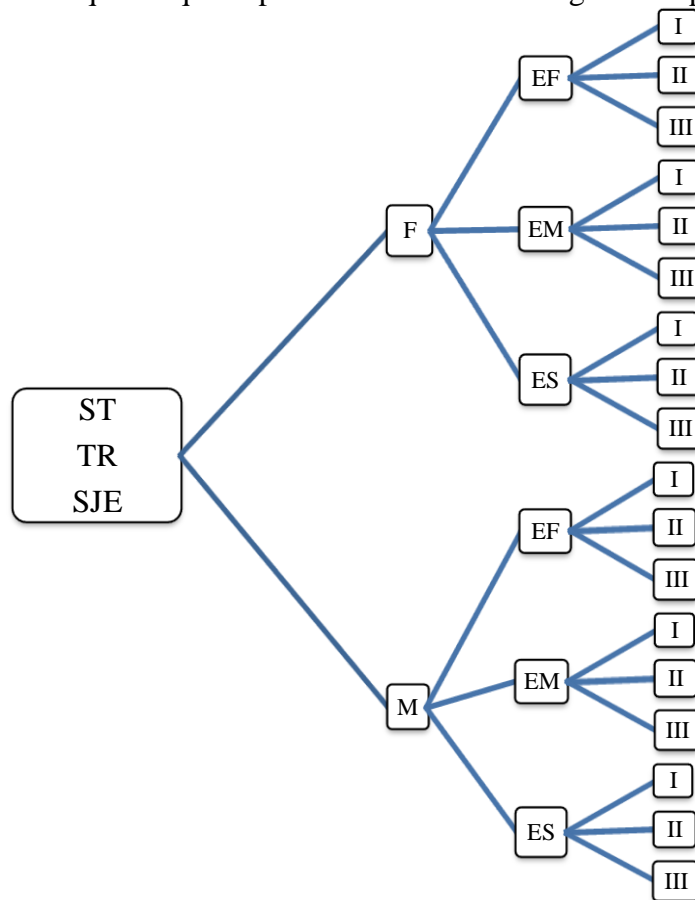
critérios mencionados vão desde a naturalidade até a faixa-etária do entrevistado, pois em estudos que buscam analisar a fala de pessoas de determinada comunidade de fala, por exemplo, se o informante é natural de outra localidade e/ou é recém chegado ao local, pode fazer uso de elementos linguísticos que não sejam recorrentes na comunidade de fala em análise, mas da de origem do informante. Podemos exemplificar isso com o exemplo do estudo que realizamos nesta dissertação, pois caso o informante seja natural de outra cidade/estado e/ou tenha chegado recentemente ao município em que a pesquisa está sendo realizada, é provável que o uso dos RADs que porventura possa fazer, pode ser mais característico de sua cidade de origem do que da cidade em que reside atualmente⁷⁶.

No entanto, antes de expormos os critérios que utilizamos para a seleção dos informantes, convém destacarmos o método de seleção que adotamos neste estudo. Como vimos no capítulo anterior, há, na metodologia sociolinguística, dois métodos para selecionar os informantes: a seleção aleatória simples e a seleção aleatória estratificada. Nesta pesquisa, descartamos a possibilidade de realizarmos a seleção aleatória simples pelo fato de os informantes serem selecionados por meio de escolha ou sorteio, e tal método poderia não atender aos critérios de seleção que estabelecemos para alcançar os objetivos do estudo.

Assim, o método que seguimos é a seleção aleatória estratificada, em que a estratificação foi realizada de acordo com os critérios pré-estabelecidos para o estudo. Estes critérios, no entanto, estão diretamente ligados às variáveis extralinguísticas que serão levadas em consideração para análise que posteriormente será feita nos dados coletados. As variáveis extralinguísticas que selecionamos são: gênero, faixa-etária, escolaridade e localidade. Para expor, com mais clareza, tais variáveis, montamos um esquema, baseado no que foi proposto por Silva (2003), e apresentamos na figura abaixo.

⁷⁶ É pertinente destacarmos que tais afirmações são feitas levando em consideração que o informante não esteja na comunidade há muito tempo, pois há casos em que, embora seja natural de outra localidade, o informante já reside no município há muito anos.

Figura 13: Esquema que expõe as variáveis extralinguísticas para seleção dos informantes



O esquema mostra as casas, também chamadas de células, que irão compor a nossa mostra de informantes. Estas casas, no entanto, são formadas a partir das variáveis extralinguísticas selecionadas para a presente pesquisa. Conforme podemos ver na imagem, a seleção se deu por meio da localidade. Na figura, exemplificamos isso por meio da sigla ST, para o município de Serra Talhada, TR para o município de Triunfo e SJE para a cidade de São José do Egito. Os informantes de Serra Talhada, por sua vez, dividem-se em dois gêneros que rotulamos F para pessoas do gênero feminino e M para pessoas do gênero masculino, cada grupo, masculino e feminino, divide-se em três grupos: EF que corresponde às pessoas que possuem apenas Ensino Fundamental, EM para o grupo de pessoas que tem Ensino Médio e ES para pessoas que têm Ensino Superior. A partir da variável escolaridade, os informantes são selecionados, ainda, com base em outro critério: a faixa-etária. No esquema, representamos a faixa-etária pelos números romanos, em que I corresponde ao grupo de informantes com uma faixa-etária de 12 a 18 anos, a faixa-etária II conta com informantes

com idade entre 19 e 40 anos e, por fim, o III grupo que é formado por pessoas com idade acima de 40 anos⁷⁷.

Como proposto por Silva (2003), a montagem de um esquema como este mostra não apenas os critérios de seleção dos informantes, que também correspondem às variáveis extralinguísticas, mas nos auxiliam na montagem das células sociais, pois nos permite ver o tamanho da amostra e a quantidade de informantes selecionados para a pesquisa. A partir do esquema acima, criamos uma amostra dos informantes, expondo, por meio das células sociais, o número de informantes de cada localidade, assim como, de cada célula social, conforme podemos ver no quadro de informantes abaixo.

Quadro 9: Distribuição da amostra de informantes da pesquisa

Localidade:		Escolaridade					
		Ensino Fundamental		Ensino Médio		Ensino Superior	
Serra Talhada São José do Egito Triunfo							
Idade ↓	Gênero→	F	M	F	M	F	M
		Faixa-etária I	1	1	1	1	1
Faixa-etária II	1	1	1	1	1	1	1
Faixa-etária III	1	1	1	1	1	1	1
Total de informantes por células de faixa-etária / gênero / escolaridade		3	3	3	3	3	3
Informantes na localidade = 18 informantes por município							
TOTAL DE INFORMANTES NA PESQUISA: 54 Informantes							

Conforme exposto, após a seleção da comunidade de fala, os informantes foram selecionados a partir de quatro características, localidade⁷⁸, gênero⁷⁹, escolaridade⁸⁰ e faixa-

⁷⁷ O porquê de escolher cada variável desta será exposto na seção que abordaremos acerca do envelope de variação, pois nesta discutiremos acerca das hipóteses do presente trabalho.

⁷⁸ Participaram deste grupo pessoas nascidas e/ou residentes no município a mais de 10 anos.

⁷⁹ Nomeamos este grupo de gênero e não de sexo, partindo da concepção de gênero enquanto característica sociológica do indivíduo e não apenas por sua identificação biológica, que seria o caso de sexo. No entanto, nenhum dos informantes participantes desta pesquisa afirma, no questionário social, que sua identificação biológica difere de sua identificação sociológica.

⁸⁰ Este fator social foi dividido em três grupos, pois tínhamos o intuito de contemplar os diferentes níveis de escolaridade, a saber: Ensino Fundamental, Ensino Médio e Ensino Superior.

etária⁸¹, permitindo assim, chegarmos ao número total de informantes da pesquisa, conforme a distribuição em cada célula é apresentada no quadro acima. Realizada esta etapa, é necessário que façamos a seleção/identificação dos informantes que tenham estas características sociais agora na comunidade de fala solicitando-os a participarem da pesquisa. A seção seguinte traz os passos seguidos para a coleta de dados com os informantes da comunidade de fala escolhidos para este estudo de acordo com o que apresentamos até aqui.

4.3 A COLETA DE DADOS

Realizar um estudo de campo na área da Linguística, como é o nosso caso aqui, é um trabalho, como já defendido por muitos autores, desafiador, sobretudo, no que tange à coleta e análise dos dados, pois, diferente do trabalho de um biólogo, por exemplo, que coleta e analisa materiais externos e visíveis, o trabalho de um linguista, por sua vez, é realizado com um material que está interno e não é palpável ao pesquisador. Discorrendo acerca disso, Castilho (2016) afirma que os estudiosos que se propõem à realização de tal ação, ou seja, estudar a língua: “operam com um objeto guardado em sua mente e na mente dos indivíduos de sua comunidade, lidando com uma propriedade interna a ele, não evidente no mundo real” (p.41). Sendo assim, a coleta de dados é o momento em que o pesquisador procura uma maneira de encontrar e registrar este objeto que está guardado na mente de seus informantes.

Além disso, consideramos que a coleta de dados de um estudo sociolinguístico não é uma etapa simples, talvez seja o momento, um dos poucos durante a pesquisa linguística inclusive, que o pesquisador sai de seu âmbito de estudo e adentra a comunidade de fala em busca de informantes que, como vimos, anteriormente, foram pré-selecionados a partir de suas características sociais. No entanto, até chegar ao momento de coletar/capturar o dado linguístico há etapas a serem obedecidas sob pena de comprometer todos os resultados de um estudo caso haja um descumprimento destas.

Uma das etapas a serem seguidas e que entendemos ser um importante passo, pois por meio dela todo o processo necessário para a coleta de dados de um estudo de campo com material humano é estabelecido, é a submissão e aprovação da realização do estudo por um

⁸¹ As três faixas-etárias (de 12 a 18 anos, de 19 a 40 anos e acima de 40 anos) foram escolhidas seguindo a ideia de que estas podem representar grupos distintos de pessoas, como: jovens; adultos que estão em busca de uma estabilização acadêmica e profissional; adultos que estão em fase de total estabilização acadêmica e profissional, respectivamente. Muito embora não tenhamos como afirmar que isso se dá com todos os informantes, mas acreditamos que é mais provável que alguém, por exemplo, com mais de 40 anos já tenha norteado sua vida acadêmica e seus objetivos profissionais, de maneira mais enfática, do que um adulto com idade um pouco mais que 19 anos.

Comitê de Ética em Pesquisa, pois todo estudo que, para sua realização, necessite de envolvimento de indivíduos, no que tange a um processo de angariação de material humano carece que o Comitê de Ética Institucional, este por sua vez ligado a um Comitê de Ética Nacional, aprove o modo como estes serão coletados e usados pelo pesquisador

A aprovação pelo Comitê de Ética é a uma ação importante, pois para alcançar esta aprovação, o pesquisador precisa, além de toda a documentação pessoal impressa e o registro na Plataforma Brasil com os documentos solicitados, apresentar o projeto de pesquisa de maneira detalhada, informando: quantos seres humanos irão participar do estudo; em que período e localidade ele será realizado; o que será usado do participante pelo pesquisador; como o sigilo do informante será mantido na pesquisa; em que lugar, físico e/ou eletrônico, e por quanto tempo, estes materiais serão armazenados, quais os termos que o voluntário da pesquisa deve assinar e possuir após a aceitação de participar do estudo etc.

Após a apresentação do projeto de pesquisa e de todos estes materiais citados, o Comitê analisa e julga como aprovado ou não a pesquisa que foi proposta. Caso seja aprovada, o comitê emite um termo consubstanciado e disponibiliza para o pesquisador, por meio do seu acesso à Plataforma Brasil, um número de identificação do estudo, o CAE. Por meio deste número de identificação, qualquer pessoa pode ter acesso à autorização do Comitê de Ética Institucional, comprovando que a coleta de dados e em que condições ela está apta a ser realizada.

A presente pesquisa submeteu a proposta metodológica e científica para a realização deste estudo ao Comitê de Ética em Pesquisa – CEP, da Universidade Federal de Pernambuco, comitê este que é vinculado a Comissão Nacional de Ética em Pesquisa – CONEP e teve parecer aprovado em 8 de junho de 2017 com a emissão do termo consubstanciado e o CAE de número 67485517.6.0000.5208. Após a aprovação pelo Comitê de Ética em Pesquisa, a coleta de dados foi realizada seguindo o que foi proposto no projeto detalhado apresentado ao referido comitê e descrito aqui, uma vez que o que foi apresentado ao CEP está em consonância com a coleta de dados de um estudo sociolinguístico, como o que realizamos nesta dissertação.

A submissão do projeto para a realização deste estudo ao CEP foi necessária uma vez que optamos por montar o *corpus* a ser analisado na pesquisa⁸². Esta decisão foi motivada por

⁸² Quando o pesquisador se propõe a utilizar um *corpus* que foi montado a partir de uma pesquisa aprovada pelo CEP, ele deve solicitar acesso e uso do material ao mesmo comitê em que aprovou a realização da coleta de dados. Os procedimentos diferem um pouco, mas são necessários, uma vez que o CEP havia dado parecer consubstanciado apenas para o projeto anterior e não para a nova pesquisa que surge.

inúmeros fatores, dentre eles, destacamos o fato de alguns bancos de dados linguísticos montados para realização de estudos nesta região⁸³, Sertão do Pajeú, não contemplavam questões que entendemos ser primordiais para a realização de RADs na fala, é o caso das perguntas feitas durante a entrevista, pois a depender do tipo de pergunta, o uso dos RADs pode não ser tão favorecido, uma vez que, estes elementos estão voltados para interação falante-ouvinte.

Sendo assim, optamos por submeter à pesquisa ao CEP e realizar a coleta e montagem do *corpus* a ser analisado. Para tanto, realizamos a coleta por meio de entrevistas gravadas com os informantes selecionados para a pesquisa. Ao selecionarmos o método de entrevista, desprezamos os métodos de interação e testes, embora sejam métodos eficientes, parece-nos que só por meio da entrevista teríamos condições de fazer o levantamento das formas que estejam atuando como RADs na comunidade de fala, pois, nas interações, os RADs utilizados poderiam ser característicos de uma rede social que os informantes pertencem e, nos testes, teríamos que sugerir as formas que porventura poderiam estar atuando como RADs e isso nos distanciaria bastante do objetivo principal deste estudo.

No entanto, antes de realizarmos a entrevista, seguindo a metodologia sociolinguística, solicitamos ao informante o preenchimento do questionário social⁸⁴. Montamos este questionário com perguntas que poderiam nos ser significativas nas análises a serem realizadas posteriormente e que não necessitariam estar no roteiro de entrevista a ser gravada, pois boa parte das respostas poderia ser dada com expressões como *sim* ou *não*, apresentando, portanto, um perfil mais tendencioso para um questionário escrito do que um oral, que é o caso das perguntas que utilizamos para a obtenção dos dados de fala.

Além do questionário escrito, antes da gravação para a coleta de dados, seguindo agora determinações do CEP, todos os informantes acompanharam a leitura e assinaram o Termo de Consentimento Livre e Esclarecido, doravante TCLE. Este termo possui um modelo estrutural dado pelo CEP da instituição em que o estudo está vinculado e deve ser, após a leitura do informante ou do pesquisador, assinado, datado e ter dados como o de Cadastro de Pessoas Físicas - CPF inseridos no fim do documento, assim como, a assinatura de duas testemunhas acerca da gravação e participação do informante no estudo. Para os informantes com menos de 18 anos de idade, além do TCLE assinado pelos responsáveis, foi disponibilizado, para aprovação, assinatura e inserção do número de CPF do menor, o Termo de Assentimento Livre e Esclarecido, doravante TALE.

⁸³ Alguns destes estudos foram mencionados na seção 3.1 deste capítulo.

⁸⁴ O questionário social que montamos para a coleta de dados encontra-se nos apêndices desta dissertação.

Reconhecemos a importância do CEP solicitar estes termos, assim como a assinatura deles antes da gravação da entrevista, afinal, para um estudo sociolinguístico o material a ser utilizado é humano, a fala. No entanto, para realização de tal procedimento acabamos por transgredir recomendações do aporte teórico-metodológico da Sociolinguística para a coleta de dados, pois conforme Labov (1972), o pesquisador sequer deve se apresentar como fazendo parte de uma universidade, pois isso pode gerar uma monitoração do informante em sua fala. Além disso, Silva (2003, p.118) afirma que não é interessante que o pesquisador dê explicações acerca do estudo, “como, por exemplo, sou linguista e estudo a queda do –r no infinitivo” e tudo isso são ações solicitadas pelo CEP a serem realizadas antes da coleta de dados, pois, no TCLE e TALE, previamente analisados e aceitos pelo comitê, estas informações devem constar no documento e precisam ser enfatizadas pelo pesquisador.

Com o intuito de amenizar os efeitos das informações dadas por meio dos termos, procedemos da seguinte forma: os entrevistados realizaram a leitura e assinatura dos termos, após isso, solicitamos o preenchimento do questionário social. Durante este preenchimento, algumas conversas informais foram realizadas, sem gravação, algo como: “se tiver alguma dúvida, pode falar” e “esta questão aqui quer saber coisas do tipo...”, decorremos assim tentando tirar a atenção do informante do fato de a fala dele está sendo observada, mesmo ele estando ciente disto.

Após estas etapas mencionadas até aqui, o gravador foi ligado e a gravação foi realizada, todos os informantes forneceram entrevistas com um tempo entre 13 e 15 minutos, e, em muitos casos, algumas perguntas complementares foram realizadas⁸⁵, tendo em vista a necessidade de manter o assunto abordado uma vez que a fala acerca dele mostrava-se como sendo espontânea, pois o informante expressava sua opinião e interagiu com o entrevistador. É o caso, por exemplo, das perguntas acerca do país, muitos informantes apresentavam seu ponto de vista, assim como, solicitavam uma aceitação dele por parte do pesquisador e esse é um momento propício, de acordo com Macedo e Silva (1996) para a utilização dos RADs.

A coleta de dados se deu no período de 3 de julho a 31 de agosto de 2017. No entanto, pelo fato de a comunidade de fala ser cidades que distam entre si, a coleta em cada cidade aconteceu em períodos diferentes, a saber: em Serra Talhada, os dados foram coletados no período de 3 de julho a 24 de julho; em São José do Egito, a coleta aconteceu no período de 25 de julho a 11 de agosto; e em Triunfo, os dados foram coletados entre os dias 14 e 31 de agosto.

⁸⁵ As perguntas que compõem o roteiro de entrevista encontram-se nos apêndices desta dissertação.

Após as gravações, os dados coletados foram transcritos, selecionados e codificados para o tratamento estatístico a ser dado na análise dos RADs. Estes procedimentos metodológicos são abordados na seção seguinte.

4.4 A TRANSCRIÇÃO, SELEÇÃO, CODIFICAÇÃO E QUANTIFICAÇÃO DOS DADOS

Com os dados de fala coletados, é necessário que o pesquisador transforme a produção falada em produção escrita, pois só por meio dela é que os dados podem ser selecionados, quantificados e analisados. É a etapa chamada de transcrição dos dados.

No presente estudo, realizamos a transcrição ortográfica dos dados coletados, uma vez que, para o que objetivamos nesta pesquisa, este tipo de transcrição é suficientemente capaz de identificar o dado a ser analisado, não necessitando, por isso, da transcrição fonética. No entanto, é necessário destacarmos que realizamos a transcrição ortográfica dos dados coletados seguindo os parâmetros de transcrição dado por este modelo, ou seja, de grafar a sentença tal e qual ela foi produzida, citamos como exemplo os casos como o (18) (19) e (20).

(18) Mas os outro só de ano e ano, *sabe?* (Inf002 – SJE)

(19) Tem bons médico, *né?* (Inf029 – TR)

(20) Os jovens poderia ser mais animados, *viu?* (Inf039 – ST)

Os exemplos apresentados mostram que há alguns desvios da norma padrão no que diz respeito à concordância, no entanto, realizamos a transcrição sem fazer alterações e/ou correções de acordo com a Gramática Tradicional da língua, pois, além de o modelo de transcrição ortográfica primar por escrever a sentença exatamente como foi produzida, se retirássemos ou alterássemos o que estivesse fora do padrão gramatical, o próprio fenômeno em análise, ou seja, os MCs, seriam excluídos da transcrição, pois o uso destes não é prescrito e nem tão pouco normatizado pela Gramática Tradicional da Língua Portuguesa.

Com a transcrição de todos dados realizada, selecionamos os dados a serem utilizados no estudo. Apesar de termos notamos, durante a escuta e a transcrição do material coletado, o uso de MCs de diferentes classes⁸⁶, selecionamos apenas os RADs para serem quantificados e

⁸⁶ É possível ter uma noção do que afirmamos aqui no exemplos citados de (1) a (9) expostos na introdução desta dissertação, pois ao citarmos as divisões feitas por Macedo e Silva (1996) dos MCs, utilizamos sentenças retiradas do próprio *corpus* para exemplificarmos cada categoria.

analisados neste trabalho. A razão desta escolha está fundamentada no seguinte aspecto: descrever e analisar o uso das nove categorias propostas por Macedo e Silva (1996) resultaria em um trabalho que, de tão extenso, comprometeria a análise do fenômeno em tela, uma vez que seria um número de formas e grupos muito grande para serem estudados, especulamos que talvez seja esta a mesma razão pela qual Valle (2001) selecionou apenas três formas dos RADs para a sua análise, a saber, as formas *sabe?*, *não tem?* e *Entende?*

Para a seleção dos dados, no entanto, um aspecto foi importante e por isso devemos mencioná-lo ao descrevermos a metodologia deste trabalho, trata-se do fato de selecionarmos as formas que estejam atuando como RADs na comunidade de fala em análise. Isso porque, diferente de outros estudos que se propõem a analisar aspectos de variação e mudança linguística com formas, de prestígio ou estigmatizadas, que coexistem no sistema linguístico de determinada comunidade de fala, o presente estudo busca, antes de realizar tal análise, identificar as formas que funcionam como RADs no Sertão do Pajeú. Sendo assim, na seleção dos dados a serem analisados foi necessário fazermos uma pré-análise dos dados coletados, no intuito de identificar, nas sentenças, as formas que podem ser classificadas como RADs evidenciando, assim, o quanto esta etapa é significativa em um estudo sociolinguístico como este.

Portanto, durante a seleção dos dados, identificamos as formas que atuam como RADs na fala dos sertanejos, algo que, de maneira geral, é bastante simples, pois os RADs possuem características que, de tão peculiares, facilitam sua identificação, dentre elas destacamos o fato de se localizarem sempre no fim da sentença, tendo em vista que possuem um contorno interrogativo, conforme aponta Macedo e Silva (1996), Marcuschi (1989), Valle (2001) e Freitag (2008).

Contudo, o critério de entonação interrogativa deve ser seguido com cautela, uma vez que os RADs, segundo os autores já citados, advêm de algumas classes de palavras da língua portuguesa, como verbos, por exemplo, e, em alguns casos, a palavra pode ser usada não como RAD, mas como um verbo mesmo, como mostram as sentenças (21) e (22) citadas abaixo.

(21) Inglês eu quero aprender mesmo, *sabe?* (Inf010 – SJE)

(22) Eita, o nome daquele ator rapaz que morreu no rio, *sabe?* (Inf048 – ST)

Ao observarmos casos como estes, percebemos que em (21) a forma *sabe?* parece ser um RAD e não um verbo, uma vez que se inseríssemos um sujeito e um complemento na sentença em que ele é realizado, algo que é defendido pela Gramática Tradicional da Língua Portuguesa, ou seja, a ordem Sujeito-Verbo-Complemento, a construção ficaria algo como: “*Inglês eu quero aprender mesmo, você sabe disso?*” e não teria tanto significado para o interlocutor, estando, talvez, mais voltada para uma pergunta retórica, ou seja, que não requer do informante uma resposta acerca do que foi perguntado, ao ponto de apenas a forma *sabe?*, sem sujeito e complemento, como foi realizada pelo informante da pesquisa, ser suficiente para realizar a pergunta retórica, algo tão característico dos RADs. Ponderamos isso partindo do pressuposto que a sentença “*Inglês eu quero aprender mesmo*”, produzida pelo informante, é uma afirmação até então desconhecida pelo entrevistador, que é a pessoa a quem a pergunta está sendo direcionada, pois as pessoas envolvidas nesta conversa não possuem, conforme o entrevistado afirmou no questionário social, nenhum grau de conhecimento pessoal⁸⁷.

Algo diferente acontece com a ocorrência exposta em (22), pois se inseríssemos os elementos que não foram realizados, ou seja, o que estão ocultos na sentença, a saber, sujeito e complemento, teríamos uma sentença como “*Eita, o nome daquele ator rapaz que morreu no rio, você sabe qual é?*”, a ideia de entonação interrogativa parece ser mais evidente, ou seja, não parece ser uma pergunta retórica, consideramos isso porque estamos partindo do pressuposto que a primeira sentença já traz um caráter mais interrogativo que é expresso pela oração que, sendo iniciada pelo marcador conversacional *eita*, já expressa uma dúvida a ser respondida, ou seja, trata-se de uma pergunta a ser respondida, mostrando, assim, que aparenta ser uma dúvida expressa na oração e que é possível de ser respondida pelo interlocutor, ainda que os participantes da conversa, entrevistador e entrevistado, não possuam nenhum grau de conhecimento pessoal.

Assim, reiteramos a ideia de que, embora a palavra apareça após uma sentença dada e tenha um caráter interrogativo, como exemplificado em (21) e (22), nem sempre podemos considerar como um RAD, como exemplificamos em (21), pois pode haver casos, como mostrado em (22) que a forma parece mais atuar como um verbo e não como um RAD.

⁸⁷ Chamamos de conhecimento pessoal uma relação de proximidade entre as pessoas participantes da conversa. Este conhecimento pode ser de diferentes formas, pode ser resultado de uma amizade, de conversas esporádicas ou até mesmo de, algum momento anterior, estas pessoas terem tido um contato que possa resultar em um conhecimento partilhado entre eles. Algo que, como destacamos, não acontece com o informante e o entrevistador.

Os exemplos expostos acima mostram como os dados coletados e transcritos foram selecionados para o estudo, pois selecionamos apenas as formas que entendemos ser RADS. Além disso, realizações como estas nos permitiram chegar à escolha de uma das variáveis linguísticas da presente pesquisa, a entonação interrogativa, pois ao selecionarmos os dados a serem analisados, notamos peculiaridades em algumas construções, como as exemplificadas em (23) e (24).

(23) E⁸⁸: Você tem ídolos? Quem são seus ídolos?

I⁸⁹: Meus ídolos são os profissionais da música, *né?* (Inf004 – SJE)

(24) E: O que você mais gosta de fazer?

I: Eu adoro viajar, *viu?* (Inf037 – ST)

As realizações expostas em (23) e (24), mostram que as sentenças produzidas pelos informantes são respostas dadas a uma pergunta feita pelo entrevistador. Estas respostas são construções afirmativas/declarativas, em face da pergunta feita acerca de algo que o entrevistador desconhece, porém o RAD aparece no fim da sentença trazendo, sob ele, uma marca interrogativa. Ou seja, os informantes respondem a pergunta, declarando quem são seus ídolos e o que gosta de fazer, como mostramos respectivamente em (23) e (24), mas utiliza-se do RAD para pedir a anuência do entrevistador. Em casos como estes, bastante recorrentes no *corpus* em análise, inclusive, defendemos que há uma entonação interrogativa na sentença, porém ela se dá apenas por meio do, ou no, RAD e não em toda a sentença em que ele está inserido, pois, sem comprometer o sentido ou a gramaticalidade da sentença, as perguntas feitas pelo entrevistador em (23) e (24), poderiam ser respondidas por meio de orações como (25) e (26)⁹⁰:

(25) I: Você tem ídolos? Quem são seus ídolos?

E: Meus ídolos são os profissionais da música.

⁸⁸ E = Entrevistador.

⁸⁹ I = Informante.

⁹⁰ Os códigos inseridos para identificação dos informantes foram retirados das sentenças (25) e (26) porque estas foram alteradas no que diz respeito a forma como elas foram realmente produzidas, como estão citadas em (23) e (24). Essa alteração foi realizada, porque objetivamos mostrar que a ausência do RAD, nestes casos, retiraria toda a marca interrogativa da sentença. Diante destas alterações, as letras inseridas antes das sentenças correspondem a I = Inqueridor e E = entrevistado.

(26) I: O que você mais gosta de fazer?

E: Eu adoro viajar.

Além disso, realizações como estas, apontam para uma consideração importante que entendemos ser pertinente e, por isso, discorreremos aqui. Trata-se da semelhança dos RADs com as construções sintáticas da Língua Inglesa chamadas, em alguns estudos, de *question tags*.

Conhecidas como perguntas retóricas e/ou sentenças que buscam iniciar/manter a conversação entre os interlocutores, as *question tags* são, segundo Murphy (2004, p.104), classificadas como “Mini-questions that we often put on the end of a sentence in spoken English. In question tags, we use an auxiliary verb. Normally we use a negative question tag after a positive sentence. And a positive question tag after a negative sentence⁹¹”. Com base nesta concepção, ponderamos que os RADs, em situações como as exemplificadas em (23) e (24), possuem algumas semelhanças com as *question tags*, pois são comuns na língua falada e aparecem após os enunciados afirmativos ou negativos, fazendo uma pergunta, quase sempre retórica.

No entanto, é válido mencionar que, no caso das *question tags*, elas são produzidas com base no oposto do que foi realizado no enunciado, ou seja, se a sentença é afirmativa, a *question tag* será negativa, se a sentença é negativa, a *question tag* é afirmativa. Algo que parece não ser comum aos RADs, pois formas que expressam um sentido negativo, como o *num é?*, são realizados após sentenças negativas, como podemos ver em (27), distanciando, assim, um pouco das semelhanças entre os RADs da Língua Portuguesa e as *question tags* da Língua Inglesa.

(27) Não, nós não fomos criados por minha mãe, *num é?* (Inf005 – SJE).

Ainda assim, acreditamos que estudos posteriores devem buscar investigar essas semelhanças e diferenças entre os RADs e as *question tags*, ainda que, para tanto, dados de fala, em língua portuguesa e língua inglesa, devam ser coletados e analisados de maneira comparativa, só assim, serão evidenciadas, ou não, as possíveis relações entre estes fenômenos linguísticos.

⁹¹ Pequenas perguntas que nós frequentemente inserimos no fim da sentença do inglês falado. Em *questions tags*, nós usamos um verbo auxiliar. Normalmente nós usamos uma *question tag* negativa após uma sentença afirmativa. E uma *question tag* positiva após uma sentença negativa. (Tradução nossa).

A despeito dos exemplos apresentados em (23) e (24), identificamos, ainda durante a seleção dos dados, realizações que nos conduziram na escolha da variável linguística citada acima, ou seja, a entonação interrogativa, pois nos dados coletados, encontramos ocorrências como (28) e (29) que em muito nos pareceram distintas das ocorrências mencionadas em (23) e (24). Vejamos.

(28) E: Certo, muito bem eee...

I: Tá perto de terminar, *hein?* (Inf049 – ST)

E: Tá sim.

(29) E: Muita coisa pode melhorar, sempre. Mas e aqui? o que pode mudar?

I: Na nossa cidade, *né?* (Inf027 – TR)

E: Sim, aqui na sua cidade mesmo.

Ao compararmos estas ocorrências, com realizações como (23) e (24), notamos que parece ser que as sentenças produzidas pelos informantes em (28) e (29) possuem uma marca interrogativa já no enunciado que antecede o RAD, fazendo com que a presença do RAD reforce a entonação interrogativa que a sentença como um todo já possui. Consideramos isso tendo como base pontos como: i) o entrevistador acaba por responder a pergunta feita pelo informante, pois tem condições de assim proceder⁹²; ii) a não realização do RAD deixaria a sentença, dentro deste contexto de produção, ainda interrogativa, algo que não parece não acontecer nas sentenças expostas em (23) e (24). Para tentar mostrar isso, mantivemos as sentenças produzidas, retirando apenas os RADs produzidos nela pelo informante, conforme apresentamos em (30) e (31).

(30) I: Certo, muito bem e...

E: Tá perto de terminar?

I: Tá sim.

(31) I: Muita coisa pode melhorar, sempre. Mas e aqui? o que pode mudar?

E: Na nossa cidade?

⁹² Afirmamos isso porque em (23), por exemplo, o entrevistador não tinha condições de afirmar qualquer coisa, pois a pergunta realizada recai apenas no RAD que parece pedir ao entrevistador apenas uma aquiescência acerca do que estava sendo dito e não uma resposta, propriamente dita.

I: Sim, aqui na sua cidade mesmo.

Tais exemplos, nos fazem notar que a entonação interrogativa é mantida a ponto de a apresentação de uma resposta, por parte do inquiridor, não deixar a sentença sem sentido e/ou agramatical, compreendendo, assim, a presença de uma pergunta por parte da sentença realizada pelo informante.

Por tudo isso, ao escolhermos os dados para a análise, temos selecionados, além das variáveis sociais já escolhidas para a seleção dos informantes⁹³, uma variável linguística a ser considerada e analisada no presente estudo: a entonação interrogativa, para esta variável independente, conforme defendemos até aqui, por meio das ocorrências encontradas no *corpus* em análise, temos duas possibilidades distintas para as realizações dos RADs, a saber: entonação interrogativa em toda a sentença ou entonação interrogativa apenas no RAD.

Com base no que apresentamos até aqui, as variáveis sociais do estudo foram apresentadas por ocasião da montagem das células sociais, a saber: localidade, gênero, faixa-etária e escolaridade, e a seleção dos dados nos permitiram escolhermos uma variável linguística a ser utilizada, a entonação interrogativa. As demais variáveis linguísticas foram escolhidas após a seleção dos dados, para que a codificação pudesse ser realizada, contudo, elas não foram escolhidas com base em ocorrências no *corpus* montado, mas são resultados de importantes considerações feitas acerca dos RADs em estudos anteriores.

No texto de Marcuschi (1989), já citado aqui como um material pioneiro acerca dos MCs, o autor divide estes itens em quatro grupos, a saber: os MCs simples, os MCs compostos, os MCsprosódicos e os MCs oracionais. Nestas divisões, um aspecto importante deve ser destacado, no que tange aos RADs, trata-se da massa fônica diretamente ligada à extensão do RADs. Para o autor, os MCs simples possuem pouca massa fônica, pois possuem apenas um lexema e são formados por palavras originárias de distantes classes, como interjeição, advérbios, verbos, adjetivos, conjunções e pronomes. Já os MCs compostos possuem duas ou mais formas, possuindo, portanto, mais massa fônica.

Sendo assim, ao intentarmos descrever as formas que atuam como RADs na fala dos sertanejos, selecionamos como uma variável linguística independente, a massa fônica, com o intuito de analisar se estas formas tendem a ser simples ou compostas. Além destas duas variáveis linguísticas independentes, escolhemos duas outras variáveis, a saber: a classe de palavra originária do RAD e a classe de palavra que antecede o RAD.

⁹³ Contudo, as hipóteses relacionadas à escolha de cada variável social serão expostas na última seção deste capítulo.

A escolha destas variáveis se deu a partir de algumas considerações feitas por Macedo e Silva (1996) e retomadas por Valle (2001) e Freitag (2008), dentre elas, destacamos a afirmação de Freitag que os RADs são, para a autora, originários de verbos ou adjetivos (2008, p.2). Sendo assim, é pertinente analisar qual a classe de palavras em que os RADs utilizados no Sertão do Pajeú são originários.

Contudo, ao realizarmos as transcrições dos dados coletados, identificamos a ocorrência da forma *hein?*, uma das formas até então não analisadas em estudos com outras comunidades de fala. Esta forma teve ocorrências significativas no estudo realizado com a comunidade de prática de radialistas do Sertão do Pajeú por Alcântara (2015), a saber, um total de 6,88%, calculado a partir do número total de ocorrências dos RADs, contudo, a classe de palavra originária desta forma não foi estudada. Sendo assim, procuramos encontrar neste RAD alguma característica que nos permitisse classificar sua origem⁹⁴, uma vez que descartamos a ideia de ser originária de um adjetivo ou verbo, pois sua forma não nos permite assim considerar.

Para tanto, realizamos uma pré-análise das ocorrências da forma *hein?* no *corpus* analisado e notamos que esta forma teve ocorrências em sentenças interrogativas, semelhantes às exemplificadas em (28) e (29), em que a entonação interrogativa não se dá apenas no RAD, mas em toda a sentença. Como nos exemplos citadas abaixo.

(32) E você já viu algum bom, *hein?*(Inf050 – ST)

(33) Tu acha, *hein?*(Inf047 – ST)

(34) Ele morde, *hein?*(Inf049 – ST)

Sendo assim, parece que esta forma é utilizada para expressar dúvidas ou até mesmo uma incerteza, algo que, segundo Castilho (2006, p.542), é uma dentre as quatorze espécies de advérbio da língua portuguesa. Além disso, Castilho afirma que, do ponto de vista sintático, “os advérbios são palavras relacionadas ao verbo, ao adjetivo ou a outro advérbio... e podem igualmente aplicar-se aos substantivos...” (CASTILHO, 2016, p.543) e estas foram, segundo os resultados apresentados por Alcântara (2015), as classes de palavras que mais antecederam as ocorrências da forma *hein?* na comunidade de prática analisada⁹⁵.

⁹⁴ Procuramos por características explícitas na forma, tendo em vista que desconhecemos estudos que descrevam e analisem a forma *hein?* enquanto RAD.

⁹⁵ Alcântara (2015, p.56) afirma que a forma *hein?* foi antecederida por substantivos em 33,33% das ocorrências, 22,22% por adjetivos, 24,45% por verbo e 20% por advérbios.

Partindo destas considerações apresentadas, consideramos pertinente selecionar a classe de advérbios como sendo a classe de palavra que originou a forma *hein?* realizada por informantes da comunidade de fala analisada. Desta forma, para a variável linguística independente nomeada de classe de palavra originária do RAD, separamos três possibilidades: i) advérbios, ii) verbos e iii) adjetivos.

No que tange à variável linguística denominada como classe de palavra que antecede o RAD, baseamo-nos nas concepções de Macedo e Silva (1996) que consistem em considerar os RADs como ocupando a posição final da sentença. Portanto, por ocupar esta posição, é válido observar quais as classes de palavras que podem favorecer o uso dos RADs, uma vez que estes se encontram no fim do enunciado.

Apresentamos estas considerações acerca do porquê em escolher cada umas das variáveis linguísticas independentes citadas aqui, porque, após a seleção dos dados coletados para a pesquisa, é necessário que o pesquisador codifique os dados para inseri-los no programa computacional a ser utilizado para o tratamento estatístico a ser dado na pesquisa. Desta forma, já elucidamos aqui quais são as variáveis independentes, sociais e linguísticas, e o porquê de estas terem sido selecionadas para o estudo, apesar de deixarmos a apresentação das hipóteses levantadas a partir de cada uma delas para serem apresentadas na última seção deste capítulo.

Tendo escolhido as variáveis independentes, necessárias para um estudo sociolinguístico e, conseqüentemente, para as rodadas computacionais do programa estatístico. É necessário identificar a variável dependente da pesquisa, esta identificação nos parece um tanto quanto peculiar quando o assunto é o estudo dos MCs, mais especificamente, dos RADs.

A peculiaridade se deve ao fato de que é considerado como sendo variável dependente para a Sociolinguística Variacionista. O conceito, conforme apresentamos no segundo capítulo desta dissertação refere-se ao fenômeno estudado, o fenômeno que se encontra em variação por meio da ‘disputa’ realizada pelas variantes, podemos exemplificar o que afirmamos com o exemplo de um suposto estudo⁹⁶ com a variação no sistema pronominal do Português Brasileiro a partir das formas *tu* e *você*. Neste caso o fenômeno em análise seria o pronome de segunda pessoa do singular, a variável dependente seria a presença/ausência do pronome *você*, que é o pronome considerado como sendo uma variante do pronome de

⁹⁶Chamamos a atenção para a expressão suposto estudo, pois o criamos apenas para exemplificarmos o que vem a ser os termos de variável dependente e variantes. Não correspondendo, assim, a um estudo real a ponto de referenciado aqui. A suposição deste estudo é apresentado neste o parágrafo seguinte.

segunda pessoa do singular, sendo assim, as variantes seriam as formas *tu* e *você* que podem estar concorrendo no sistema linguístico da comunidade de fala.

Porém ao atentarmos para as possibilidades de variáveis dependentes de alguns estudos sociolinguísticos, notamos que elas trazem, na maioria dos casos, uma possibilidade de variação que foge da prescrição gramatical e uma forma que está de acordo com a prescrição gramatical⁹⁷, elucidando, assim, a variação que pode estar ocorrendo e talvez apresentando indícios de mudança linguística, em que a forma padrão de prestígio esteja caindo em desuso e a forma estigmatizada, sobretudo pela gramática, esteja sendo em constante uso na comunidade de fala. Ou ainda, as variáveis dependentes são selecionadas a partir de duas ou mais formas prescritas pela Gramática Tradicional que apresentam variação no uso, ou seja, formas que possuem duas possibilidades de realizações a partir dos preceitos gramaticais. No entanto, estabelecer uma variável dependente de um elemento linguístico que não é abordado pela Gramática Tradicional da língua, como é o caso dos RADs, é, antes de tudo, analisar uma possível variação de um elemento que, do ponto de vista gramatical, passa despercebido por ela.

Sendo assim, nos casos dos MCs, e de maneira mais específica no caso dos RADs, temos, de forma bem evidente, o fenômeno linguístico em análise, os RADs e as variantes deste fenômeno, ou seja, as diferentes formas que atuam como RADs, mas as possibilidades da variável dependente não são tão claras assim, uma vez que a variável RADs já é estigmatizada pela Gramática Tradicional, impedindo, assim, que tenhamos uma opção padrão e outra estigmatizada ao abordarmos o uso destes.

Diante desta problemática e da necessidade de selecionarmos uma variável dependente para a variável linguística em análise, retomamos a ideia central do trabalho – identificar se há formas de RADs na fala dos sertanejos pernambucanos que diferem das formas já identificadas como RADs da Língua Portuguesa em outros estudos já realizados – e consideramos como sendo a variável dependente da presente pesquisa a presença/ausência de RAD já identificado na Língua Portuguesa. Embora reconheçamos que as formas já reconhecidas, ou não, sejam estigmatizadas pela Gramática Tradicional, as selecionamos como variável dependente tendo em vista que o próprio fenômeno em análise, embora real na

⁹⁷ As autoras Scherre e Naro (2003) apresentam um levantamento de alguns estudos para mostrar quando as variáveis dependentes são binárias, tais estudos podem também exemplificar o que estamos afirmando aqui acerca da variável dependente, os estudos são Scherre e Naro (1997) que analisa a alternância de concordância de número a partir do plural explícito *vs* plural zero; a presença *vs* ausência da semivogal nos ditongos (PAIVA, 1996); alternância das formas pronominais *seu* e *dele* (OLIVEIRA E SILVA, 1996a.), entre outros.

língua, também seja exprobrado pelo manual de Gramática da língua em que se materializa.

Concluída esta etapa, as variáveis independente e dependente selecionadas, os dados coletados precisam ser codificados para serem rodados no programa computacional que oferece tratamento estatístico aos dados linguísticos. Procedemos à codificação seguindo as regras do próprio programa, ou seja, sem repetição de códigos, distinção para letras, consideradas pelo programa como código, maiúsculas e minúsculas, assim como, a inserção de números como códigos serem usados para lidos pelo programa.

A escolha dos códigos é arbitrária, ou seja, não é motivada por nada relacionado à variável. No entanto, ao codificar uma sentença, o pesquisador inicia a análise do dado coletado, pois precisa considerar, como no caso do fenômeno aqui analisado, se a forma realmente é um RAD e, sendo, analisar qual seria a classe de palavra originária do RAD, assim como julgar qual a classe da palavra que antecede o RAD.

Sendo assim, realizamos a codificação atribuindo códigos a cada variável selecionada, conforme já discorremos aqui acerca da seleção de cada uma. A codificação de cada variável pode ser observada no quadro abaixo.

Quadro 10: Codificação dos dados coletados

VARIÁVEL DEPENDENTE	
A (= presença de RAD já descrito na língua portuguesa) B (= ausência de RAD já descrita na língua portuguesa)	
VARIÁVEIS INDEPENDENTES EXTRALINGUÍSTICAS	
Gênero 4 (=Feminino) 5 (=Masculino)	Faixa etária 1 (= faixa etária 1) 2 (= faixa etária 2) 3 (= faixa etária 3)
Localidade 6 (= Serra Talhada) 7 (= São José do Egito) 8 (= Triunfo)	Escolaridade 0 (= Ensino Fundamental) 9 (= Ensino Médio) F (= Ensino Superior)
VARIÁVEIS INDEPENDENTES LINGUÍSTICAS	

Classe de palavra originária do RAD O (= advérbios) P (= verbos) Q (= adjetivos)	Massa fônica R (= forma composta) S (= forma simples)
Classe de palavra que antecede o RAD s (= substantivo) a (= adjetivo) p (= pronome) v (= verbo) o (= advérbio)	Entonação interrogativa M (=Entonação interrogativa só no marcador) N (= Entonação interrogativa em toda a sentença)

Após a codificação, os dados foram inseridos no programa computacional. No nosso caso aqui, utilizamos o programa do pacote VARBRUL, mais precisamente, a versão *GoldVarbX*, versão 3.0, tendo em vista que este programa traz, como produto final das rodadas realizadas a partir dos códigos inseridos, resultados numéricos e pesos relativos de cada fator, informando se estes são significativos do ponto de vista estatístico.

Ainda com relação à codificação, o programa que a recebe tem como opção receber os dados codificados ladeados das sentenças em que o fenômeno foi realizado, essa é uma estratégia que pode auxiliar ao pesquisador, pois, caso os códigos inseridos sejam apontados pelo programa como tendo algum erro, como, por exemplo, duplicidade ou ausência de um código, a sentença estando expressa facilita a identificação e reparação do erro, uma vez que o pesquisador não precisa voltar ao *corpus* completo para fazer a identificação do que o programa acusou.

Para a rodagem dos dados, seguimos esta estratégia que abordamos acima. Para tanto, os códigos no programa precisam ser colocados após um parêntese(, tendo em vista que o programa fará a leitura de tudo o que estiver após a abertura do parêntese como pertencente ao código e as sentenças, que produziram cada código, devem ser expostas após um espaço em branco que seja capaz de separar o código da sentença. Assim, a codificação e as sentenças em que os RADs foram realizados no presente estudo foram inseridos no programa seguindo o modelo apresentado abaixo:

(B4270PSvM	Eu só ia mais pra bagunçar e arengar, <i>visse?</i>
(A427/PSsM	Muita superação, <i>sabe?</i>
(B4169OSvN	Como gostar, <i>hein?</i>

Todas as sentenças que tiveram ocorrências de RADs foram codificadas, a partir dos códigos estabelecidos, conforme expomos no quadro 10 e exemplificamos acima. Os resultados das rodadas computacionais serão apresentados no quarto capítulo desta dissertação, pois estes nos permitirão analisar os dados coletados.

Contudo, para realizarmos a análises, temos que partir de hipóteses para as variáveis selecionadas, estas serão apresentadas na seção seguinte deste capítulo.

4.5 AS VARIÁVEIS E AS HIPÓTESES LEVANTADAS

Acreditamos que a variável a que nos propomos analisar nesta pesquisa já tenha sido devidamente exposta nas seções e capítulos anteriores. Ainda assim, retomamos nesta seção a variável linguística, assim como, as variáveis independentes e dependente do presente estudo com o intuito de apresentarmos as hipóteses levantadas para cada uma destas.

A variável linguística, ou seja, o fenômeno que buscamos analisar, é a categoria de RADs, esta é uma das nove categorias de MCs e está voltada para a interação entre falante e ouvinte. Acreditamos que já está evidente que, segundo os resultados a que Valle (2001) e Freitag (2008) chegaram em suas pesquisas sobre os RADs, apresentados no primeiro capítulo desta dissertação, esta categoria pode ser considerada uma variável linguística e os itens que pertencem a ela são também considerados, variantes desta variável. Sendo assim, para esta variável, Macedo e Silva (1996) elencam seis variantes: *né?, tá?, sabe?, entendeu?, viu?* e *não é mesmo?*. Estas seis formas estão presentes no português, porém temos como hipótese que podem existir outras formas que atuam como RADs na fala de pessoas do Sertão do Pajeú – Pernambuco que estejam sendo utilizadas em co-ocorrências com estas formas já existentes. Com base nisso, selecionamos como Variável dependente a presença *vs* ausência de RAD já catalogado na Língua Portuguesa.

O *corpus* desta pesquisa é formado por cinquenta e quatro informantes do Sertão do Pajeú estratificados por localidade (Serra Talhada, São José do Egito e Triunfo), gênero, (masculino e feminino), escolaridade (Ensino Fundamental, médio e superior) e faixa-etária (I – de 12 a 18 anos; II – de 19 a 40 anos e III – Acima de 40 anos), sendo, portanto, as variáveis extralinguísticas desta pesquisa.

Levando em consideração estas variáveis, temos como hipótese para a variável localidade que, embora seja cidades próximas e pertencentes a uma mesma comunidade de

fala, como defendemos neste capítulo, podem existir formas distintas operando como RAD em cada uma destas cidades que integram a comunidade de fala aqui analisada.

No que diz respeito à variável gênero, nossa hipótese é que as formas de RADs realizadas no sertão pernambucano são mais usadas por mulheres, visto que algumas pesquisas já mencionadas neste trabalho apontam para o fato de as mulheres usarem mais os MCs como um todo, não sendo diferente, portanto, no uso dos RADs, em nossa concepção.

Com relação à faixa-etária, temos como hipótese que as pessoas mais jovens utilizam mais os RADs. Além disso, hipotetizamos também que, entre os jovens, o número de formas de RADs pode ser maior, pois, por frequentarem diferentes esferas da sociedade, escola, trabalho, praças, clubes, academias, shows etc., utilizam de maneira muito significativa a língua para interagirem com seus pares e com amizades novas, algo tão comum na vida de um jovem⁹⁸.

Para a variável escolaridade, embora não tenhamos nenhum dado de estudos anteriores que nos motive a apresentar alguma hipótese acerca desta variável social, achamos pertinente analisar esta variável hipotetizando que quando maior o nível de escolaridade, menos o uso dos RADs, pois, como afirmamos anteriormente, o uso dos RADs é estigmatizado e em sendo assim pode ser que pessoas com um nível de instrução maior evite usar estes elementos, uma vez que eles não são aceitos pela Gramática Normativa da Língua Portuguesa.

Até aqui, destacamos a variável estudada e as variáveis externas à língua que podem influenciar o uso das variantes e as hipóteses que temos para cada um destes fatores. Cabe agora destacarmos como as variáveis internas à língua, ou variáveis linguísticas, com também são chamadas, podem influenciar o uso de uma forma em detrimento de outra.

Como objetivamos analisar as formas que operam como RADs no Sertão do Pajeú, selecionamos como variável linguística a classe de palavra que parece originar o RAD, pois hipotetizamos que estas formas podem ser originárias de outras classes de palavras, além das já descritas por autoras como Macedo e Silva (1996), a saber, os verbos e os adjetivos, pois, da mesma maneira que formas novas surgem no grupos dos RADs, é possível que estas surjam de outras classes de palavras da Língua Portuguesa.

Ainda em relação às classes de palavras, selecionamos como variável interna à classe de palavra que antecede o RAD, tendo em vista que eles são realizados no fim da sentença,

⁹⁸ Não estamos, com isso, afirmando que pessoas de mais idade não frequentem tais espaços e/ou não façam novas amizades, estamos apenas considerando que isso parece ser mais assíduo na vida do jovem, principalmente quando comparado com uma pessoa adulta de uma cidade interiorana, como é o caso dos informantes desta pesquisa.

temos como hipótese que são antecidos por palavras que pertencem à classe como substantivo, adjetivo, pronome, verbo ou advérbio. Uma vez que as demais classes, como as interjeições, artigos, preposições, numeral e conjunções não parecem ocupar a posição final de um enunciado. Sendo assim, temos como hipótese que estas classes de palavras mencionadas podem favorecer o uso das diferentes formas de RADs, daí procurarmos observar quais destas classes podem favorecer mais as ocorrências de RADs na fala dos sertanejos.

No que tange à massa fônica, selecionada como variável interna, temos como hipótese que as formas não descritas podem ser mais recorrentes enquanto formas compostas, pois as formas já descritas podem ser formas simples uma vez que podem estar a mais tempo no sistema linguístico e em diferentes comunidades de fala, tornando o uso destas formas mais recorrente e, conseqüentemente, sendo formas que, de tão usadas, já passaram por um processo de economia reduzindo a massa fônica da variante.

Selecionamos também, como variável linguística independente, a entonação interrogativa, com base nas ocorrências exemplificadas em (23), (24), (28) e (29). Para esta variável temos como hipótese que os RADs, embora possuam uma entonação interrogativa, como apontam os estudos já mencionados, eles são frequentemente utilizados em sentenças que não expressam uma pergunta, mas uma afirmação/declaração, tendo, assim, a entonação interrogativa apenas no RAD e não na sentença em que é utilizado.

Baseados em todas estas considerações, apresentamos, no capítulo seguinte, as análises feitas a partir dos dados coletados, pois estas análises nos permitirão corroborar, ou não, as hipóteses levantadas.

5 ANÁLISE DOS DADOS

No presente capítulo, apresentamos os dados coletados e as análises que realizamos destes dados, procurando observar quais das hipóteses levantadas foram ou não confirmadas. Para tanto, dividimos o capítulo em seções para apresentação das análises realizadas. Iniciaremos a análise por meio de uma comparação das variantes encontradas no *corpus* e das variantes identificadas nos estudos que utilizamos como referências neste trabalho, pois temos como objetivo preceder as análises elucidando o porquê de considerarmos determinada forma como sendo peculiar ou não da comunidade de fala em estudo.

Na segunda seção, os números de realizações serão expostos e analisados, ou seja, cada variante encontrada será apresentada, por meio de números, como foi realizada pelas pessoas participantes da pesquisa. Para tanto, levaremos em consideração as variáveis independentes linguísticas e sociais no que diz respeito ao uso destas formas.

Na terceira seção deste capítulo, apresentaremos os números obtidos a partir das rodadas realizadas no programa computacional *GoldVarbX*, com intuito de discorreremos as análises a partir dos pesos relativos e porcentagens calculados pelo programa a partir da variável dependente e das variáveis independentes.

Na última seção, articularemos acerca das hipóteses que foram ou não confirmadas após as análises realizadas. Após estas seções, os resultados da pesquisa serão expostos por meio das considerações finais.

5.1 AS VARIANTES ENCONTRADAS NO SERTÃO DO PAJEÚ

Nesta seção serão apresentadas as variantes dos RADs que encontramos na fala das pessoas do Sertão do Pajeú. Não trataremos das realizações por meio de números de ocorrências, pois nosso objetivo nesta primeira seção, não é elucidar as formas que são mais e as que são menos utilizadas, mas como e por que consideramos cada forma como sendo uma variante, da variável linguística em estudo, específica ou não da comunidade de fala em análise.

Convém destacar, portanto, que ao apresentarmos as formas que estão sendo usadas como RADs nesta região, partiremos de três pontos importantes: i) se estas formas possuem ocorrências de uso em outra comunidade de fala; ii) se há formas que não tiveram ocorrências

nesta comunidade, mas que já foram apontadas em outros estudos como sendo RADs; e iii) se há formas, nesta comunidade, que não tiveram realizações em outras comunidades de fala.

Para facilitar essa apresentação e compreensão, é necessário que relembremos as variantes que os estudos realizados com RADs, mencionados anteriormente nesta pesquisa, catalogaram nas comunidades de fala/prática analisadas. Para isso, montamos o quadro abaixo para exposição destas variantes.

Quadro 11: Variantes de RADs já estudadas por diferentes pesquisadores.

	Marcuschi (1989) ⁹⁹	Macedo e Silva (1996)	Valle (2001)	Freitag (2008)	Alcântara (2015)
Né	X	X	X	X	X
Tá	X	X			X
Sabe	X	X	X		X
Entendeu	X	X	X		X
Viu	X	X			X
Não é mesmo	X	X			
Não foi	X				
Pronto				X	
Repare				X	
Não tem			X		
Tá bom					X
Hein					X
Tá certo					X
Ok					X

Algumas considerações importantes devem ser apreendidas a partir do que está exposto no quadro acima. Em primeiro lugar, destacamos o uso e descrição deles nos estudos mencionados no quadro acima: no caso, por exemplo, da variante *né?*, esse RAD teve ocorrências em todos os estudos mencionados. Já as formas *tá?* *sabe?* *entendeu?* e *viu?* tiveram ocorrências em três ou mais, dos cinco estudos citados, evidenciando, assim, um uso considerável destas formas, enquanto RADs, em diferentes comunidades de fala. Essa observação nos faz considerar, ou rotular, as variantes encontradas no Sertão do Pajeú como sendo comuns na Língua Portuguesa e/ou específicas da comunidade de fala em análise.

⁹⁹ É válido destacar que as formas que mencionamos como estudadas por Marcuschi (1989), consideradas aqui RADs, são apontadas, no seu estudo, como *finalizadores*, uma vez que o autor separa em sua pesquisa as formas que atuam como finalizadores ou iniciadores. No entanto, estamos os considerando aqui como RADs, tendo em vista que são as mesmas formas apontadas por Macedo e Silva (1996), quando as autoras dividem os grupos de iniciadores e finalizadores em nove subgrupos como já mencionamos nesta dissertação.

Outro ponto importante a ser destacado é que as variantes de RADs citadas por Freitag (2008) são formas encontradas pela pesquisadora no estudo comparativo que fez com os dados alcançados por Valle (2001). Sendo assim, não podemos afirmar que são apenas estes RADs realizados em Itabaiana – Sergipe, mas são as formas *pronto* e *repare* que apareceram no *corpus* analisado em Itabaiana e não apareceram no *corpus* analisado em Florianópolis – Santa Catarina.

Pontuamos que, embora tenhamos citado o estudo feito por Alcântara (2015), não iremos selecioná-lo aqui como capaz de apontar formas já estudadas dos RADs, tendo em vista que o estudo foi realizado em uma comunidade de prática do Sertão do Pajeú, os radialistas. Sendo assim, as formas apresentadas no seu estudo podem ser variantes dos RADs na comunidade de fala também e não apenas na comunidade de prática estudada. Assim, as formas encontradas no estudo de Alcântara (2015) serão comparadas com as formas realizadas no *corpus* aqui em análise, para que, assim, tenhamos condições de perceber quais as variantes de RADs podem ser da comunidade de fala Sertão do Pajeú e quais são da comunidade de prática formada por radialistas do Sertão do Pajeú.

Entendemos que estas colocações são pertinentes de serem feitas antes de apresentarmos as variantes encontradas na fala dos sertanejos pernambucanos, pois o que consideramos aqui como variantes já descritas na Língua Portuguesa ou como variantes que parecem ser específicas da comunidade de fala em análise é baseado nos estudos já realizados com os RADs.

Após a coleta, transcrição e seleção dos dados, identificamos as variantes de RADs em ocorrências¹⁰⁰ como as que expomos no quadro abaixo:

Quadro 12: Variantes e suas ocorrências no *corpus* em análise

VARIANTES	OCORRÊNCIAS
<i>Né?</i>	Na verdade nunca fui lá visitar eles, <i>né?</i> (Inf024 – TR) A família minha é boa, estável, <i>né?</i> (Inf004 – SJE)
<i>Sabe?</i>	Só a violência mesmo, <i>sabe?</i> (Inf040 – ST) Eu acho que deve existir mais respeito, <i>sabe?</i> (Inf002 – SJE)
<i>Viu?</i>	Eu vou começar pelos bregueiros, <i>viu?</i> (Inf025 – TR) Aqui às vezes é ruim, <i>viu?</i> (Inf039 – ST)
<i>Tá ligado?</i>	Já vi a pessoa querendo ser mais do que a outra, <i>tá ligado?</i> (Inf001 – SJE) Eu num tenho amigos, <i>tá ligado?</i> (Inf039 – ST)

¹⁰⁰ Grafamos o RAD precedido por vírgula e sucedido pelo sinal de interrogação tendo em vista que reconhecemos que eles são realizados após uma pausa na fala tendo uma entonação interrogativa, quer seja apenas no RAD, quer seja na sentença como um todo.

Entendeu?	Ele é humilde, <i>entendeu?</i> (Inf038 – ST) A gente se dá bem, <i>entendeu?</i> (Inf017 – SJE)
Num sabe?	O problema é que ele vive por conta própria, <i>num sabe?</i> (Inf025 – TR) Eu reconheço que sou ruim, <i>num sabe?</i> (Inf057 – ST)
Num é?	Faz é criticar, <i>num é?</i> (Inf043 – ST) Nós fomos criados por minha mãe, <i>num é?</i> (Inf005-SJE)
Hein?	Como gostar, <i>hein?</i> (Inf047 – ST) Tá perto de terminar, <i>hein?</i> (Inf049 – ST)
Tá?	Ai, num quero falar disso não, <i>tá?</i> (Inf046 – ST) Alguns homens também, <i>tá?</i> (Inf018 – SJE)
Visse?	Não conto com ninguém de minha família, <i>visse?</i> (Inf015 – SJE) Já faz muito tempo, <i>visse?</i> (Inf008 – SJE)
Certo?	Deixando as famílias aflita, <i>certo?</i> (Inf027 – TR) Miguel Falabela, <i>certo?</i> Sônia Braga, <i>certo?</i> (Inf027 – TR)

O quadro mostra que identificamos na comunidade de fala 11 variantes para a variável RADs. Aos considerarmos estas formas variantes de uma mesma variável, estamos afirmando que, segundo os pressupostos da Sociolinguística Variacionista, estas formas podem ser utilizadas em qualquer uma das sentenças expostas no quadro e, ainda assim, teriam o mesmo valor de verdade, ou seja, serviriam para solicitar a aquiescência do interlocutor.

Para evidenciarmos isso, selecionamos, aleatoriamente, três das sentenças produzidas por três diferentes informantes, de três diferentes cidades e inserimos no final das sentenças as formas que, no *corpus*, identificamos como variantes da variável RAD. Conforme está apresentado abaixo.

Quadro 13: Ponderação acerca das formas como sendo variantes de uma mesma variável

NA VERDADE NUNCA FUI LÁ VISITAR ELES, NÉ? (inf024 – TR)	
Sabe?	Na verdade nunca fui lá visitar eles, <i>sabe?</i>
Viu?	Na verdade nunca fui lá visitar eles, <i>viu?</i>
Tá ligado?	Na verdade nunca fui lá visitar eles, <i>tá ligado?</i>
Entendeu?	Na verdade nunca fui lá visitar eles, <i>entendeu?</i>
Num sabe?	Na verdade nunca fui lá visitar eles, <i>num sabe?</i>
Num é?	Na verdade nunca fui lá visitar eles, <i>num é?</i>
Hein?	Na verdade nunca fui lá visitar eles, <i>hein?</i>
Tá?	Na verdade nunca fui lá visitar eles, <i>tá?</i>
Visse?	Na verdade nunca fui lá visitar eles, <i>visse?</i>
Certo?	Na verdade nunca fui lá visitar eles, <i>certo?</i>
ELE É HUMILDE, ENTENDEU? (inf038 – ST)	
Né?	Ele é humilde, <i>né?</i>
Sabe?	Ele é humilde, <i>sabe?</i>
Viu?	Ele é humilde, <i>viu?</i>

<i>Tá ligado?</i>	Ele é humilde, <i>tá ligado?</i>
<i>Num sabe?</i>	Ele é humilde, <i>num sabe?</i>
<i>Num é?</i>	Ele é humilde, <i>num é?</i>
<i>Hein?</i>	Ele é humilde, <i>hein?</i>
<i>Tá?</i>	Ele é humilde, <i>tá?</i>
<i>Visse?</i>	Ele é humilde, <i>visse?</i>
<i>Certo?</i>	Ele é humilde, <i>certo?</i>
	ALGUNS HOMENS TAMBÉM, TÁ? (inf018 – SJE)
<i>Né?</i>	Alguns homens também, <i>tá?</i>
<i>Sabe?</i>	Alguns homens também, <i>sabe?</i>
<i>Viu?</i>	Alguns homens também, <i>viu?</i>
<i>Tá ligado?</i>	Alguns homens também, <i>tá ligado?</i>
<i>Entendeu?</i>	Alguns homens também, <i>entendeu?</i>
<i>Num sabe?</i>	Alguns homens também, <i>num sabe?</i>
<i>Num é?</i>	Alguns homens também, <i>num é?</i>
<i>Hein?</i>	Alguns homens também, <i>hein?</i>
<i>Visse?</i>	Alguns homens também, <i>visse?</i>
<i>Certo?</i>	Alguns homens também, <i>certo?</i>

Podemos compreender que, por meio do teste mostrado acima¹⁰¹, as formas parecem ser variantes de uma mesma variável linguística, uma vez que a troca de uma forma pela outra não compromete nem o significado da sentença nem o caráter de solicitação da anuência do interlocutor ao que está sendo dito pelo informante. É partindo desta conclusão que, a partir de então, nos juntamos ao posicionamento de Valle (2001) e Freitag (2008) e consideramos que há formas que atuam como variantes dos RADs, sendo este uma variável linguística.

Assim, partindo desta concepção, procedemos as análises deste capítulo, pois esse entendimento nos é necessário, tendo em vista que todo o tratamento dado aos RADs a partir de agora, partem deste princípio, ou seja, de que eles possuem variantes linguísticas.

Dentre as formas identificadas como RADs, reconhecemos algumas que já foram apontadas como presentes em outras comunidades de fala, como é o caso do *né?*, da mesma sorte que há variantes ainda não mencionadas em outros estudos, mas que tiveram ocorrências aqui, como é o caso de *hein*¹⁰²?. Além disso, notamos a ausência, nesta comunidade de fala, de formas que foram recorrentes em outras comunidades, mas que, nesta em análise, não foi

¹⁰¹ Destacamos, contudo, que não é um teste para a coleta de dados como mencionamos anteriormente, mas trata-se de uma espécie de experimento que realizamos com as sentenças produzidas pelos informantes, com o intuito de perceber se trocássemos a forma de RAD, se a sentença ficaria sem sentido ou agramatical.

¹⁰² Afirmamos que a forma *hein?* não é compreendida aqui como uma variante já identificada tendo em vista que esta forma apareceu na fala de uma comunidade de prática desta mesma comunidade de fala, ou seja, na pesquisa realizada por Alcântara (2015) com radialistas do Sertão do Pajeú.

realizada, como é o caso de *não é mesmo?*, apontado, como sendo um RAD, nos estudos realizados por Marcuschi (1989) e Macedo e Silva (1996).

Para facilitar esse entendimento, ou seja, a identificação das formas que já foram ou não mencionadas como RADs, reorganizamos, em um único quadro, os quadros 11 e 12 expostos acima, retirando apenas o estudo realizado por Alcântara (2015), pois para este é pertinente uma análise contrastiva separada, tendo em vista que a autora realizou o estudo em uma comunidade de prática pertencente a comunidade de fala aqui estudada. Analisemos a reorganização dos quadros 11 e 12¹⁰³, apresentada no quadro 14 abaixo.

Quadro 14: Os RADs encontrados na comunidade de fala em análise

	Marcuschi (1989)	Macedo e Silva (1996)	Valle (2001)	Freitag (2008)	Estudo em tela
Né	X	X	X	X	X
Tá	X	X			X
Sabe	X	X	X		X
Entendeu	X	X	X		X
Viu	X	X			X
Não é mesmo	X	X			
Não foi	X				
Pronto				X	
Repare				X	
Não tem			X		
Hein					X
Tá ligado					X
Num sabe					X
Num é					X
Visse					X
Certo					X

As informações apresentadas no quadro 14 nos fazem chegar aos seguintes pontos: há variantes no Sertão do Pajeú que também são utilizadas e reconhecidas como RADs em outras comunidades de fala, é o caso das variantes *né?*, *tá?*, *sabe?*, *entendeu?* e *viu?*, e isso pode apontar para algo como, após a ampliação dos estudos com RADs em Língua Portuguesa, a existência de formas que perduram no sistema linguístico brasileiro, tendo em vista que ao comparamos, por exemplo, a pesquisa pioneira de Marcuschi (1989) com o presente trabalho, temos um intervalo de 28 anos, e neste significativo intervalo as formas *né?* *tá?*, *sabe?*,

¹⁰³ Retiramos do quadro 12 os exemplos de realizações, tendo em vista que o nosso foco aqui é apresentarmos uma comparação das variantes em cada *corpus* analisado.

entendeu? e *viu?* permanecem no sistema linguístico do falante, como no caso das pessoas do Sertão do Pajeú.

Em contrapartida, a forma *não é mesmo?*, catalogada e estudada por Marcuschi (1989) e Macedo e Silva (1996), não apareceu no *corpus*, pelo menos nas comunidades de fala estudadas, a partir do século XX, pelas pesquisadoras Valle (2001) e Freitag (2008), pois, embora não tenha sido o foco destas pesquisadoras fazer um levantamento detalhado das variantes de RADs nas falas estudadas, as variantes encontradas na comunidade de fala foram citadas nos estudos, como é o caso do *né?*, por exemplo, que nem Valle (2001) nem Freitag (2008) se propuseram a analisar, no entanto, mencionaram que o uso desta é muito frequente nos grupos estudados. Entendemos que, se a variante *não é mesmo?* tivesse ocorrências nos estudos realizados por estas pesquisadoras, ainda que não tenha sido seu foco de análise, o registro e a menção deste uso seria citado na pesquisa, algo que não se deu, fazendo-nos compreender, de maneira especulativa, que nos estudos com RADs no século XXI, a variante *não é mesmo?* tem mostrado uma possibilidade de desuso.

Algo semelhante acontece com a forma *não foi?*, catalogada e estudada por Marcuschi (1989). De todos os estudos realizados com RADs citados aqui, este foi o único em que esta forma foi apresentada como variante dos RADs em Língua Portuguesa, pois, como podemos ver através dos quadros 11, 12 e 14, os estudos não trazem mais ocorrências desta variante para a variável RAD¹⁰⁴.

No caso da forma *não foi?*, algo nos chama a atenção: trata-se da variante *num é?*, presente no *corpus* em análise e não antes mencionada como RAD em estudos realizados com este fenômeno, pois parece ser que a variante *num é?* pode ter passado a ser realizada a partir do desuso da forma *não foi?*, visto que temos aí uma variação da formano *num é?* > *nume* a mudança de tempo do verbo ‘ser’, enquanto marcador, ocorrida entre *é* e *foi*. Porém, é necessário um estudo diacrônico para evidenciar isso. De outro modo, o que fazemos aqui consiste, apenas, em olhar para o dado e entender essa possibilidade, uma vez que ao notar as realizações, percebemos formas que talvez tenham caído em desuso e outras que entraram na fala das pessoas. No entanto, nosso foco aqui não é um estudo diacrônico, mas um estudo sincrônico das formas que estão presentes na comunidade de fala em análise.

Em contrapartida a isto, ou seja, diferente das formas que parecem se manter recorrentes na língua e das que aparentam ter caído em desuso, há variantes que não foram mencionadas em outras pesquisas, mas que aparecem na comunidade de fala aqui analisadas,

¹⁰⁴ Enfatizamos a nossa afirmação de maneira especulativa, porque só um estudo diacrônico e com mais comunidades de fala poderia confirmar tal especulação.

é o caso das variantes *hein?*, *tá ligado?*, *Num sabe?*, *num é?*, *visse*¹⁰⁵? e *certo?*. O que está posto no quadro 14 nos permite compreender que estas formas integram¹⁰⁶ o grupo dos RADs na fala dos sertanejos, da mesma forma que a variante *não tem?* Integram o grupo de RADs em Florianópolis – Santa Catarina (Cf. VALLE, 2001) e as variantes *pronto* e *repare*¹⁰⁷ (Cf. FREITAG, 2008) que são realizadas pelas pessoas de Itabaiana– Sergipe.

Acerca destas variantes, é necessário destacarmos algumas considerações, pois a forma *visse?*, por exemplo, é considerada um marcador muito frequente na fala das pessoas de Recife, capital do estado em que esta comunidade de fala está inserida, a saber, Pernambuco, conforme afirmam Irineu e Costa (2015). No entanto, na comunidade de fala em análise, o Sertão do Pajeú, esta variante teve ocorrências de uso. Tal dado nos faz considerar a possibilidade de *visse?* ser uma variante presente, além de na capital, em algumas cidades do estado de Pernambuco, não sendo, portanto, um RAD específico da cidade do Recife¹⁰⁸.

Além disso, atentamos para a variante *Tá ligado?* que, embora não tenha ocorrências registradas em estudos anteriores com RADs, não parece ser uma forma estranha para os falantes da Língua Portuguesa, sobretudo, em algumas comunidades de prática formadas por jovens. Trazemos estas considerações motivadas por duas razões: i) desconhecemos estudos linguísticos que investiguem o uso da expressão *tá ligado?*; ii) enquanto falantes da língua portuguesa, consideramos que esta é uma forma bem conhecida, talvez por pertencer a determinadas comunidades de prática, pois a construção não nos parece estranha ao ponto de não a reconhecermos como pertencente a Língua Portuguesa.

No que tange à variante *hein?*, como citado por Alcântara (2015), é uma forma que, embora não seja costumeiramente identificada como RAD, é recorrente na fala dos sertanejos. No presente estudo, as ocorrências de *hein?* evidenciam o quanto esta forma parece ser peculiar no Sertão do Pajeú, pois conta com ocorrências tanto na comunidade de prática

¹⁰⁵ Cabe destacar aqui que consideramos a forma *visse?* como uma variante de RAD ainda não estudada por levarmos em consideração que o estudo de Irineu e Costa (2015) em que aponta a forma *visse?* como pertencente ao grupo dos marcadores utilizados em Pernambuco. Essa forma teve sua análise realizada em produções escritas e por meio de uma perspectiva funcionalista, não abordando, portanto, a forma como uma variante linguística. Os próprios autores finalizam o texto sugerindo análises futuras em que o *visse?* seja analisado como variante e receba um tratamento teórico-metodológico da Sociolinguística Variacionista, que é o que nos propomos fazer aqui.

¹⁰⁶ E para considerarmos estas formas como integrantes dos grupos dos RADs, realizamos a análise que mencionamos quando os dados da pesquisa foram selecionados, conforme apresentamos no capítulo anterior.

¹⁰⁷ Grafamos estas formas sem interrogação seguindo a maneira como a autora menciona em seu estudo, pois embora afirme que estão atuando como RADs na fala dos sertanejos, solicitando a atenção do interlocutor, a autora não marca com entonação interrogativa as formas citadas.

¹⁰⁸ É válido destacar que a realização de *visse?* se mostrou tão significativa para o presente estudo, que dedicamos uma seção deste capítulo para analisá-la. Sendo assim, enfatizamos que retomaremos e aprofundaremos a discussão acerca desta variante posteriormente nesta dissertação.

analisada por Alcântara (2015), como na comunidade de fala aqui estudada, ambas do Sertão do Pajeú.

No que diz respeito às variantes *num sabe?* e *num é?*, são formas que nos parece ter uma relação com *sabe?*, e com *né?*, respectivamente, já analisadas enquanto RADs em outras comunidades de fala. Afirmamos isso levando em consideração que *sabe?* já era considerado um RAD da língua portuguesa e o *num é?* pode ter sido a forma intermediária da variante *né?*: *não é mesmo?* → *num é?* → *né?*. Sendo esta forma intermediária, como sugerimos aqui, e que só pode ser corroborada por meio de estudos diacrônicos específicos com estas formas, o *num é?* esteve em competição, em determinado momento, com a forma *não é mesmo?* que, como já mencionamos, parece não aparecer em estudos posteriores aos de Macedo e Silva (1996). Sendo assim, o *num é?* parece ter mantido as ocorrências nesta comunidade de fala, o que não se deu em outras já mencionadas, a ponto de considerarmos como uma forma nova e específica da comunidade aqui estudada. Se esta forma é uma redução de *não é mesmo?* ou uma ampliação da variante *né?* só uma análise diacrônica poderia, por meios de dados de fala, confirmar ou refutar tal hipótese. Lembramos, assim, que esta hipótese vai de encontro o que citamos anteriormente, mesmo que por proposição também, com relação às variantes *num é?* e *num foi?*.

O número de suposições em torno das formas *não é mesmo?*, *num foi?*, *num é?* e *num sabe?* e *né?* sugere, seguindo a metáfora de Macedo e Silva (1996) acerca da floresta virgem, um passo à frente desta mata desconhecida. No entanto, reconhecemos que este passo só pode ser dado se o pesquisador estiver equipado com as ferramentas necessárias para um estudo diacrônico da língua, algo que não é nosso foco nesta dissertação.

A apresentação das variantes encontradas no *corpus* aqui analisado por meio de uma comparação, como fizemos por meio dos quadros 11, 12 e 14, nos foi necessária para explicitarmos o porquê de considerarmos determinadas variantes como sendo já identificadas como RADs, assim como, considerarmos outras formas como pertencentes à comunidade de fala em análise, ao ponto de considerarmos que estas formas podem ser peculiares no Sertão do Pajeú. Além disso, para endossar esta concepção é necessário que comparemos também as variantes encontradas neste *corpus* montado com dados de fala da comunidade de fala Sertão do Pajeú e as encontradas no *corpus* montado com dados de fala de uma comunidade de prática do Sertão do Pajeú (Cf. ALCÂNTARA, 2015). Para tanto, montamos o quadro abaixo que traz as formas deste e do estudo anterior para que possamos proceder com as análises comparativas.

Quadro 15: Variantes encontradas no *corpus* analisado por Alcântara (2015) e no presente estudo

	Alcântara (2015)	Estudo em tela
Né	X	X
Tá	X	X
Sabe	X	X
Entendeu	X	X
Viu	X	X
Tá bom	X	
Hein	X	X
Tá certo	X	
ok	X	
Tá ligado		X
Num sabe		X
Num é		X
Visse		X
Certo		X

As realizações de *né?*, *tá?*, *sabe?*, *entendeu?* e *viu?* nestas comunidades de prática e de fala do Sertão do Pajeú, assim como em algumas das outras comunidades de fala estudadas por outros pesquisadores, como já citamos anteriormente, apontam para uma possibilidade de permanência de algumas formas dos RADs, ou seja, formas que possuem ocorrências em diferentes lugares em que seus falantes utilizam a Língua Portuguesa.

No entanto, as variantes apresentadas no quadro acima nos fazem atentar para o seguinte dado: há formas que atuam como RADs no Sertão do Pajeú diferentes das formas que são consideradas RADs em outras regiões¹⁰⁹. É o caso, por exemplo, da variante *hein?*, que não foi realizada em outras comunidades estudadas, mas teve ocorrências tanto na fala dos radialistas, quanto na fala das pessoas selecionadas aleatoriamente na região, o que parece ser um indício de que esta forma seja recorrente na região do Sertão do Pajeú; há variantes, como *tá bom?*, *Tá certo?* e *Ok?* que, embora tenham tido realizações no estudo desenvolvido por Alcântara (2015), foram usadas apenas no contexto considerado mais formal, ou seja, no programa de rádio, e não foram realizadas nas entrevistas, tratadas como um contexto mais informal no referido estudo, o que pode justificar a não-ocorrência destas

¹⁰⁹No entanto, temos destacado que nesse estudo, a pesquisadora analisou dados de fala de uma comunidade de prática, a saber, os radialistas do Sertão do Pajeú e não pessoas do Sertão do Pajeú estratificadas aleatoriamente a partir de variáveis sociais pré-estabelecidas, como fazemos aqui.

formas no presente trabalho, pois neste não coletamos dados em dois contextos de formalidades distintas, mas em uma única situação, a entrevista informal. Há formas que tiveram ocorrências apenas quando os dados de fala da comunidade foram analisados a partir de uma coleta do vernáculo, como é o caso das variantes *tá ligado?*, *num sabe?*, *num é?*, *visse?* e *certo?*.

Todos estes dados evidenciam o quanto as pesquisas com este fenômeno linguístico são significativas para os estudos linguísticos da Língua Portuguesa, assim como, o quanto ainda há a ser investigado, pois os RADs e suas variantes parecem mudar de uma região para outra, de uma comunidade de fala para outra, de uma comunidade de prática para outra, de um contexto de formalidade para outro e assim por diante.

Por tudo isso, podemos resumir as considerações aqui tecidas a partir das comparações realizadas como: as onze formas que tiveram realizações na fala dos sertanejos pernambucanos podem ser consideradas variantes linguísticas que, em alguns casos, possuem as mesmas formas que estudos anteriores já apontaram como integrantes do grupo dos RADs, assim como, em outros casos são formas que parecem ser típicas da comunidade em análise, quer seja uma comunidade de prática, é o caso do *tá bom?*, *tá certo?* e *ok?*, quer seja uma comunidade de fala, é o caso das variantes *tá ligado?*, *num sabe?*, *num é?*, *visse?* e *certo?*. Além disso, as análises comparativas oferecem alguns *insights* para estudos posteriores, sobretudo, na perspectiva diacrônica de análise da língua.

A despeito dos números percentuais de uso de cada variante, que iremos expor na seção seguinte, analisar estas formas sob esta concepção binária, ou seja, formas típicas da comunidade de fala e formas típicas da Língua Portuguesa como um todo, nos faz atentar para algo interessante: as formas já descritas¹¹⁰ e as formas não descritas em estudos anteriores em outras comunidades de fala¹¹¹ parecem ter alguma relação entre elas e não apenas entre as variantes *num é?*, *né? sabe?* e *num sabe?* que já mencionamos anteriormente. Montamos o quadro abaixo com o intuito de mostrar o quanto parece existir uma relação entre estas formas dos RADs na fala dos sertanejos.

¹¹⁰ Rotulamos assim com o intuito de identificarmos as variantes encontradas nesta comunidade de fala e não produzidas em outras comunidades de fala já estudadas.

¹¹¹ Nomeamos assim com o mesmo intuito que o fizemos com as formas chamadas de nova, ou seja, para identificar as que já foram estudadas como RADs em outras comunidades.

Quadro 16: Possíveis relações entre as variantes dos RADs

Formas já catalogadas	Formas não catalogadas
<i>né?num é?</i>	
<i>viu?visse?</i>	
<i>Sabe?num sabe?</i>	
<i>tá?tá ligado?</i>	
Tácerto¹¹²?certo?	

Apesar de nos parecer muito clara certa relação entre as formas já descritas e as formas não catalogadas ainda, só é possível afirmarmos se elas co-ocorreram no sistema linguístico e se uma advém da outra após um estudo diacrônico, como citamos anteriormente.

Munidos destas considerações, apresentaremos, na seção seguinte, as variantes encontradas no *corpus* deste estudo, assim como as análises que podemos depreender delas, no que tange aos números percentuais.

5.2 OS NÚMEROS DE OCORRÊNCIAS DAS VARIANTES¹¹³

Não apenas os RADs, mas os MCs como um todo, são elementos da língua que chamam a atenção dos linguistas pelo alto índice de ocorrências na fala¹¹⁴. É fato que esse

¹¹² No caso desta variante, consideramos, neste quadro específico, como uma forma mais antiga, mesmo não apresentando realizações em estudos pioneiros acerca dos RADs, uma vez que foi realizada enquanto RAD na comunidade de prática analisada por Alcântara (2015). Nessa linha de comparação, optamos por considerar a variante *tá certo?* como mais antiga que a variante *certo?* que só teve realizações no *corpus* aqui analisado.

¹¹³ A criação e o desenvolvimento desta seção foram necessários tendo em vista o que expomos anteriormente acerca da variável dependente no que tange ao estudo aqui realizado com os RADs, pois o programa computacional utilizado, *GoldVarbX*, para realizar as rodadas em que os percentuais e pesos relativos são gerados, necessita que as variáveis dependentes sejam estabelecidas e codificadas e, como colocamos no capítulo anterior, as variáveis dependentes são sempre resultados de duas ou mais possibilidades de uso da variável estudada, sendo que, nestas variáveis, há, geralmente, a presença/ausência de formas padrão e estigmatizadas da língua, permitindo, assim, estabelecer as variáveis dependentes. No caso da variável linguística em tela, os RADs, todas as formas de uso são estigmatizadas, pois a Gramática Tradicional não os reconhece (ser ou não reconhecido pela GT é um ponto para estigmatização?) como elementos da língua. Sendo assim, para estabelecermos a variável dependente recorreremos a nossa hipótese principal de que pode haver formas específicas atuando como RAD na fala dos sertanejos. Sendo assim, a variável dependente selecionada para este estudo com variável linguística de um fenômeno totalmente estigmatizado foi presença/ausência de formas já estabelecidas, por estudos linguísticos, como sendo RADs. Partindo desta concepção, os resultados a serem apresentados pelas rodadas do *Varbrul*, expostos na seção seguinte, serão sempre a partir desta variável binária. Portanto, antes de analisarmos estes resultados, decidimos por analisar, em termos percentuais, o uso de cada variante encontrada no *corpus*, justificando, assim, a criação e o desenvolvimento desta seção neste capítulo.

número se mostra ainda maior quando todas as formas de MCs pertencentes aos nove subgrupos que Macedo e Silva (1996) estabeleceram, são contabilizadas em um mesmo *corpus*, apesar de que proceder a uma análise, em um único estudo, de todas as formas de MCs seja um trabalho que, de tão extenso, apresenta-se como sendo praticamente impossível de ser realizado, caso não sejam feitos recortes que possibilitem as análises destes itens.

No caso dos RADs, o uso deles é sempre muito recorrente em situações de interações como a que propomos durante a coleta de dados para o presente estudo. Após gravação, transcrição e seleção dos dados coletados, contabilizamos um número total de 673 ocorrências de RADs na comunidade de fala em análise. Esse número nos parece significativo por ser tratar do número de realizações apenas dos RADs e não de todos os MCs realizados na fala dos sertanejos.

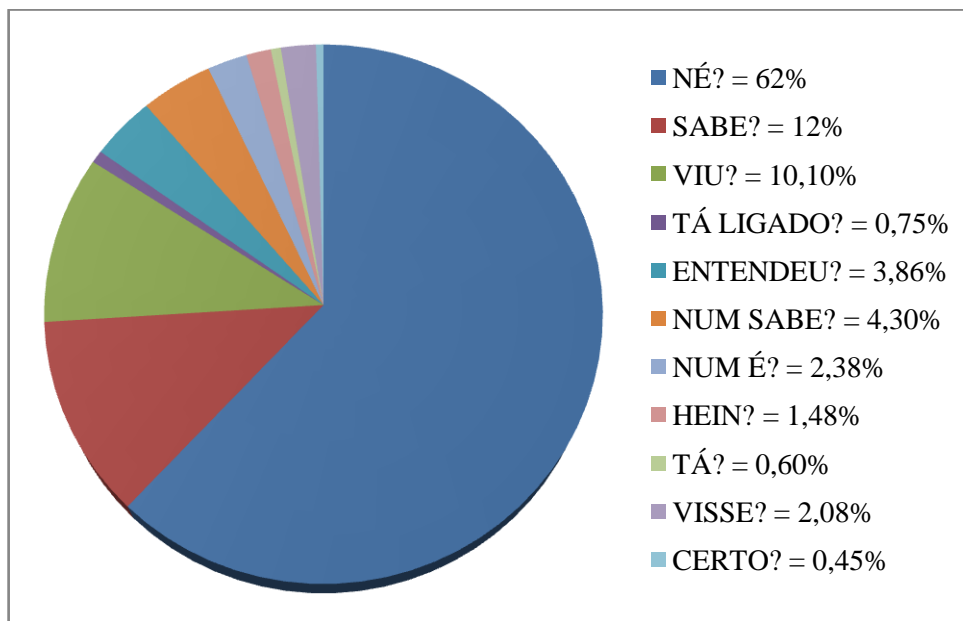
No entanto, a maneira como estes RADs se apresentam é bem peculiar, pois utilizar um recurso, por meio de diferentes formas, torna-se bem mais discreto que utilizar este recurso por meio de uma única forma. É caso, por exemplo, de um determinado nome em um texto, repeti-lo inúmeras vezes é um tanto quanto cansativo para o interlocutor, contudo, se este nome é substituído, pelo produtor do texto, por sinônimos ou pronomes como um todo, a sua presença torna-se, embora real, menos marcada. É o que parece acontecer com os RADs e suas variantes, pois destas 673 realizações, temos onze formas, no caso da comunidade de fala analisada, que se alternam no uso desta variável linguística.

O número de realizações de cada variante foi bastante distinto, conforme podemos ver em termos percentuais no gráfico abaixo¹¹⁵.

¹¹⁴ Não estamos com isso afirmando que os MCs, e os RADs, são utilizados apenas na fala, pois é possível que, a depender do gênero da produção escrita, estes itens multifuncionais da língua portuguesa sejam realizados, pois podem atender à necessidade do produtor em algum fim específico.

¹¹⁵ Em números reais, as variantes tiveram um número de ocorrências conforme expomos aqui: 417 →*né?*, 81 →*sabe?*, 68 →*viu?*, 5 →*tá ligado?*, 26 →*entendeu?*, 29 →*num sabe?*, 16 →*num é?*, 10 →*hein?*, 4 →*tá?*, 14 →*visse?* e 3 →*certo?*.

Gráfico 2: O percentual de uso das formas de RADs encontradas no *corpus*



Ao analisarmos os dados apresentados no gráfico 2, notamos que a variante *né?*, neste estudo, apresentou o maior número de ocorrências e esta tem sido uma constatação feita em todos os estudos realizados com RADs, o *né?* parece ser uma forma canônica de RAD na Língua Portuguesa, pois tem sido comum e muito frequente em diferentes estudos e em diferentes comunidades de fala.

As variantes *sabe?* e *viu?* também se mostraram com um número significativo de realizações no *corpus*, pois, depois do *né?*, estas foram as variantes mais recorrentes do RADs neste estudo. É interessante destacar que estas são formas que são citadas como variantes dos RADs desde os estudos iniciais com os MCs e em sendo estas, *né?* *sabe?* e *viu?*, formas que claramente mostram sua forma verbal, podemos compreender como pertinente a afirmação feita por Macedo e Silva (1996) acerca da origem verbal dos RADs.

Mas é válido atentar para o fato de que, além destas formas que consideramos aqui bem recorrentes, *né?*, *sabe?* e *viu?*, as demais formas, consideradas como peculiares da comunidade de fala ou não, tiveram um percentual de uso bem pequeno, nenhuma teve ocorrências acima de 5% das realizações. Porém, se atentarmos para um dado interessante, podemos notar algo que pode ser significativo: o percentual de uso das três variantes já citadas corresponde a um total de 84,1%, restando, assim, um número correspondente a 15,9% nos 100% de uso das formas, um número maior, inclusive, que as formas *sabe?* e *viu?* que já destacamos. Sendo assim, o baixo número de realizações destas variantes parece não

ser motivado por questões que porventura estejam no entorno da não realização destas formas enquanto RADs, mas pela disponibilidade de inúmeras variantes no sistema linguístico do falante.

Consideramos isso tendo como ponto de partida o seguinte dado: o *né?* é a forma mais recorrente, 62% das ocorrências, seguido do *sabe?* e do *viu?*, com 12% e 10,1%, respectivamente. Se estivéssemos lidando com um fenômeno que possui, por exemplo, duas, três ou quatro variantes, como é frequente em alguns fenômenos da Língua Portuguesa, teríamos para a quarta variante uma média de 15,9% de uso. No entanto, estamos lidando com um fenômeno linguístico que parece possuir, nesta comunidade de fala, onze formas se alternando no sistema linguístico do informante e isso pode justificar o baixo índice de realizações das variantes encontradas na comunidade de fala.

A despeito disso, no entanto, atentamos para as variantes *entendeu?* e *tá?*, com um índice de ocorrência de 3,86% e 0,60%, respectivamente. Apesar de estas variantes estarem com este pequeno índice de recorrência, elas são formas que atuam como RADs em diferentes comunidades de fala, de acordo com os estudos de Marcuschi (1989) e Macedo e Silva (1996).

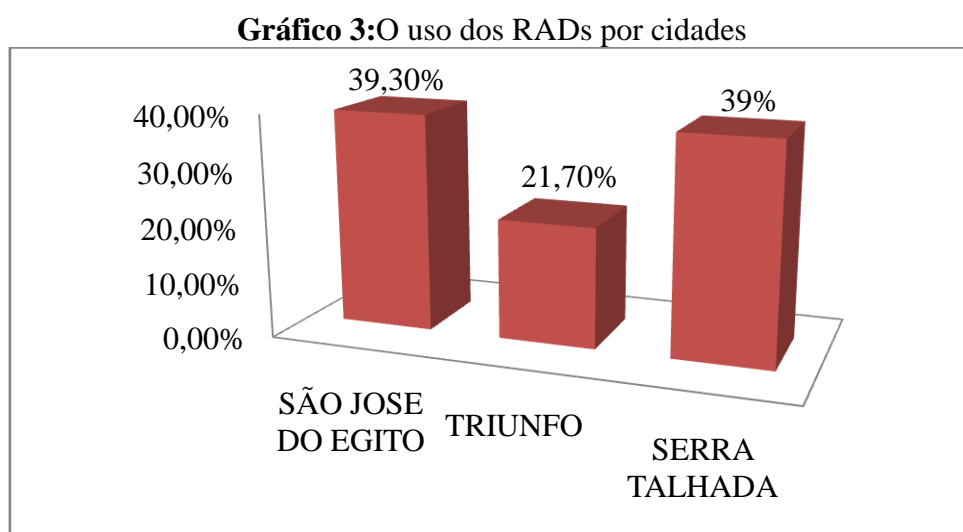
Os números de realizações das variantes consideradas aqui como sendo específicas da comunidade em análise são bem pequenos. No entanto, atentamos para o fato de que, mesmo tendo um índice pequeno, são formas que estão, de alguma forma, presentes na língua e tem seu uso condicionado por fatores linguísticos e sociais. Estamos nos referindo a um número ínfimo, contudo, só as formas *tá ligado?* e *certo?* tiveram um percentual de uso abaixo de 1%, a saber, 0,75% e 0,45%, respectivamente, pois formas como *num sabe?* com um percentual de uso de 4,30% é considerada a quarta variante mais utilizada na comunidade de fala. O percentual de uso da variante *num é?* também, embora pequeno, de 2,38%, nos chama a atenção, pois, sendo uma forma nova, possui o número de ocorrências maior que a variante *tá?*, que é aqui considerada uma forma antiga de RAD, uma vez que esta teve um percentual de uso de 0,60%, como citamos anteriormente.

Quanto ao número de ocorrências da variante *hein?*, 1,48%, não o consideramos tão pequeno assim, tendo em vista que, como já dito, dos estudos realizados com RADs, esta variante teve realizações, apenas, na comunidade de prática de radialistas do Sertão do Pajeú, ou seja, ter ocorrências de 1,48% de uma variante já nos é suficiente para perceber que o uso dela é legítimo na língua falada nesta comunidade de fala.

Algo semelhante acontece com a variante *visse?*, considerada, como mencionamos na seção anterior, específica do falar recifense. Afirmamos ser semelhante porque o percentual de 2,08% de ocorrências nesta comunidade de fala é suficiente para entendermos que o uso de *visse?* enquanto RAD no sertão do Pajeú é real¹¹⁶.

Até aqui apresentamos alguns números e considerações acerca destas variantes no que diz respeito à comunidade de fala em análise. Contudo, como mencionamos no capítulo anterior, a comunidade de fala aqui escolhida é representada por três diferentes cidades, São José do Egito, Serra Talhada e Triunfo, que distam entre si e possuem características culturais, geográficas e sociais diferentes. Portanto, atentamos para o uso dos RADs em cada cidade que aqui representa a comunidade estudada.

Conforme podemos ver no gráfico abaixo, o número de ocorrências em cada cidade possui diferenças não muito acentuadas, exceto na cidade de Triunfo, que teve quase 20% a menos de ocorrências dos RADs. Notamos que Serra Talhada e São José do Egito tiveram, equilibradamente, um número de realizações de RADs bem próximos.



O número de ocorrências de RADs na cidade de Triunfo foi algo que nos chamou a atenção desde a coleta dos dados. No entanto, só na transcrição e codificação atentamos para algo que parece motivar esse baixo número de realizações de RADs: o uso dos MCs do grupo dos iniciadores.

¹¹⁶ Adiante retomaremos a análise em torno da variante *visse?* e ficará ainda mais notório que esta é uma variante de RAD não específica do Recife, mas também é presente no Sertão do Pajeú.

Retiramos alguns exemplos do *corpus* em análise para tentar mostrar o quanto o uso deste tipo de marcador foi frequente na fala do triunfense¹¹⁷.

- (35) Bom, eu não mudaria de vida por ninguém. (Inf024 – TR)
- (36) Olha, é uma cidade linda. (Inf029 – TR)
- (37) Olha, sempre foram todos bons. (Inf029 – TR)
- (38) Bom, eu acho que se não tivesse corrupção (Inf029 – TR)
- (39) Assim, eu não costumo viajar. (Inf030 – TR)
- (40) Assim, eu num era uma coluna. (Inf031 – TR)
- (41) Tipo, um exemplo. (Inf031 – TR)
- (42) Bem, eu gosto dessa cidade por conta do clima dela. (Inf033 – TR)
- (43) Olha só, em toda cidade existe problema. (Inf033 – TR)
- (44) Olha só, o nome mesmo não lembro. (Inf033 – TR)
- (45) Assim, eu passarei pouco tempo. (Inf033 – TR)
- (46) Bom, pra mim é bom. (Inf035 – TR)
- (47) Então, eu num mudaria não. (Inf035 – TR)
- (48) Assim, só na graduação. (Inf036 – TR)

O uso recorrente dos iniciadores em Triunfo nos chamou atenção pelo seguinte motivo: apenas em uma sentença o RAD foi realizado com a presença de outro marcador, a forma *ahh* pertencente ao grupo dos iniciadores, de acordo com o que podemos ver na sentença abaixo.

- (49) Ahh e eu não posso esquecer meu orientador Ferrari, *né?* (Inf036 – TR)

Contudo, esta sentença produzida por um informante que traz as marcas de um falar não tão específico da cidade de Triunfo, pois ao consultar o questionário social do informante, tendo em vista que buscamos averiguar o porquê de esta ser a única a ter uma diferença, que, nesse caso, parece ser significativa. Notamos que, embora o informante seja de Triunfo, faz pós-graduação na Cidade de João Pessoa – PB e trabalha como professora na cidade de Serra

¹¹⁷ Como não tínhamos como objetivos analisar os iniciadores, não contabilizamos o número de ocorrências. Contudo, a alta frequência nos chamou atenção, sobretudo, quando identificamos um número pequeno de realizações dos RADs.

Talhada e Santa Cruz da Baixa Verde, ou seja, possui contatos linguísticos que podem ser capaz de influenciar a sua maneira de falar.

Afirmamos tudo isso justamente por causa das ocorrências dos iniciadores concomitantemente com os RADs, nas demais cidades, como podemos ver nas sentenças abaixo, algo que não aconteceu em Triunfo.

(50) Vixe, começou, *num é?* (Inf044-ST)

(51) Oxe, eu mesmo deixo, *sabe?* (Inf050 – ST)

(52) Ahh viajar é bom, *né?* (Inf050 – ST)

(53) Ahh, foi na hora da doença, *né?* (Inf004 – SJE)

(54) Ahh, eu acredito que são muitas coisas, *né?* (Inf016-SJE)

(55) Ahh, amigos eu tenho muitos, *viu?* (Inf017 – SJE)

Sendo assim, parece que o número de ocorrências de RADs em Triunfo é menor, se comparado com as demais cidades representativas da comunidade de fala em análise, por causa do aparente grande número de realizações dos iniciadores, uma vez que nesta cidade, as realizações de RADs e iniciadores de forma simultânea parecem não acontecer, pois identificamos apenas uma ocorrência e esta ainda foi produzida por um informante que transita em diferentes comunidades de fala.

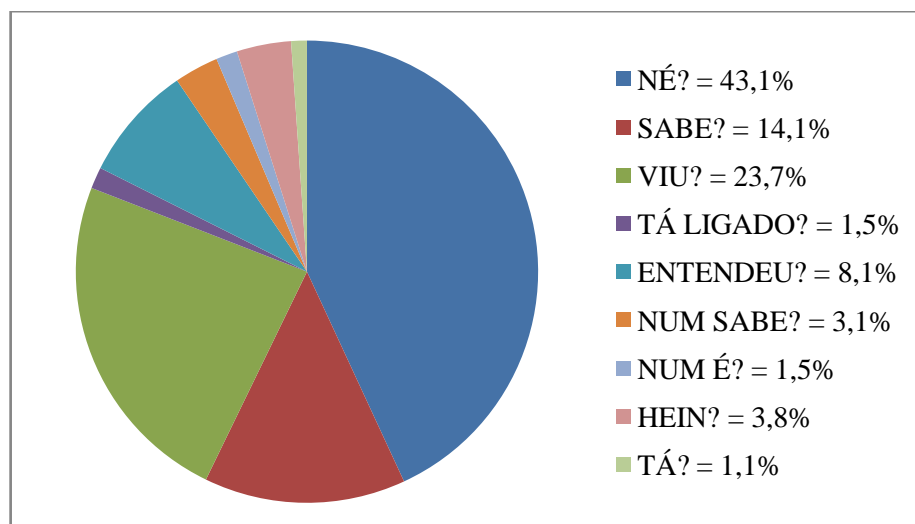
Esta peculiaridade nas realizações dos RADs na cidade de Triunfo nos levou a procurar analisar as variantes em cada cidade, pois, a partir do número de realizações dos RADs, notamos que parece ser que, embora as três cidades sejam de uma mesma comunidade de fala, o Sertão do Pajeú, há particularidades em cada uma delas. Se assim se dá é o que veremos nas seções seguintes.

5.2.1 Serra Talhada

No município de Serra Talhada, o número de realizações total e de variantes foi bem significativo, pois, como apresentamos na seção anterior, a cidade teve um percentual de uso de 39% das ocorrências total de RADs, ou seja, 262 realizações, das onze variantes encontradas na fala dos sertanejos, nove delas têm ocorrências na fala dos serratalhadenses. Isso mostra o quanto esta cidade é representativa da comunidade de fala

analisada. O gráfico abaixo nos auxilia na compreensão de quais são as variantes e o percentual de uso de cada uma delas em Serra Talhada. Analisemos.

Gráfico 4: Os RADs em Serra Talhada



Os números postos no gráfico nos permitem perceber que o número de variantes que consideramos aqui como típicas da comunidade de fala estão em equilíbrio com o número de formas que já atuam como RADs em outras comunidades de fala, 4 e 6, respectivamente. Contudo, em termos percentuais de uso, as variantes antigas, ou seja, já consideradas RADs, têm um número de realizações significativamente maior, pois tiveram um percentual de 90,1% de ocorrências, enquanto que as formas que apareceram nesta comunidade de fala específica contabilizaram um número 9,9% .

Sendo assim, na cidade de Serra Talhada, tem-se um número de variantes significativo, contudo, as variantes que consideramos como sendo formas que atuam enquanto RADs na região do Sertão do Pajeú possuem um pequeno número de realizações. Isso pode ser visto quando analisamos formas como *entendeu?* que, nas realizações em ST, TR e SJE, teve um baixo índice de ocorrência, 3,86% mais precisamente, mas, como em Serra Talhada as formas já estudadas em outras comunidades e que atuam como RADs parecem ser mais frequentes na fala, este número sobe para 8,1% dentro do município, ou seja, esta variante tem um percentual de uso ínfimo na comunidade de fala como um todo, porém em Serra Talhada este uso ainda é acentuado, tendo em vista que das nove, a variante *entendeu?* estaria entre um das quatro variantes mais usadas na fala pelas pessoas de Serra Talhada.

Uma mudança no índice de uso também acontece em Serra Talhada, se compararmos com os números da comunidade de fala como um todo. Trata-se das variantes *sabe?* e *viu?*. Na

seção anterior, elucidamos que estas formas tiveram um alto índice de ocorrência, 12% e 10,10%, respectivamente, em toda a comunidade de fala, porém, em razão do baixo índice de realizações das demais variantes, no município de Serra Talhada, este índice sobe de maneira considerável, sobretudo no que diz respeito à variante *viu?* com um percentual de uso de 23,7%.

Já a variante *sabe?* não aparece como sendo a mais utilizada, após a variante *né?*, que teve um percentual de 43,1%, pois no caso da forma *sabe?* o percentual de uso foi de 14,1% para esta variante, ou seja, com um índice de uso menor do que a forma *viu?*, como apresentamos aqui.

Além destes dados, atentamos para o número de ocorrências das variantes *tá ligado?* e *tá?*, na cidade de Serra Talhada, pois, na comunidade de fala, estas formas tiveram um percentual de uso bem abaixo de 1%, como apresentamos na seção anterior, porém, em Serra Talhada, a forma *tá?* e *tá ligado?* tiveram um percentual de realização de 1,1% e 1,5%, ou seja, as formas que parecem ter um baixo índice de realização no Sertão do Pajeú podem não ter índices semelhantes em algumas localidades que integram esta comunidade de fala.

Se o número de realizações das variantes *tá?* e *tá ligado?* elevaram-se em Serra Talhada, é esperado que determinadas formas tenham um número menor de uso e foi isso que parece ter acontecido com as formas *num sabe?* e *num é?*. Com porcentagem de 3,1% e 1,5%, respectivamente, estas variantes foram utilizadas bem menos em Serra Talhada do que na comunidade de fala em estudo. Esse dado nos aponta para uma consideração importante acerca dos RADs e que motiva a realização deste e de futuros estudos: o uso dos RADs necessita de análises e descrições na Língua Portuguesa, pois há inúmeras particularidades em suas variantes dentro das comunidades de fala em que eles são utilizados.

Um ponto interessante acerca das variantes na comunidade de fala é a ausência de algumas formas. Em Serra Talhada, por exemplo, as variantes *visse?* e *certo?* não tiveram ocorrências enquanto RADs. Inclusive, no estudo realizado por Irineu e Costa (2015), a forma *visse?* aparece como típica do falar pernambucano. No entanto, consideramos que a não-ocorrência desta variante em um dos municípios no *corpus* aqui analisado nos leva a considerar que há formas atuando enquanto RADs em uma comunidade de fala, mas, ainda assim, parece haver variantes que são específicas de determinadas lugares que integram a comunidade de fala, como é o caso de Serra Talhada.

Afirmamos que os RADs possuem particularidades, não apenas pelos dados apresentados acima, mas por oscilações no uso de algumas variantes, como é o caso da

variante *hein?*. Como citado, é uma forma que estudos anteriores não apontaram como pertencente à classe dos RADs, porém, na comunidade de fala em análise, tivemos um percentual de uso de 1,48%, 10 ocorrências, no total. Contudo, ao analisamos as variantes de cada cidade, notamos que esta forma teve ocorrências apenas na cidade de Serra Talhada.

É certo que em números percentuais, dentro da comunidade de fala e no município em que é produzida *hein?*, esta é uma variante com um número de realizações bem pequeno, 1,48% e 3,8%, respectivamente. No entanto, é algo real no uso da língua e, de maneira mais específica, é uma forma que atua como variante no município que integra a comunidade de fala em análise. Tal assertiva nos motivou a analisar, dentro desta seção intitulada Serra Talhada, a variante que parece ser específica desta localidade que integra a comunidade de fala. É o que procederemos a seguir.

5.2.1.1 *Hein?*

Para procedermos às análises em torno da variante *hein?*, realizada em Serra Talhada, lançamos mão de algumas das variáveis independentes¹¹⁸ linguísticas e extralinguísticas selecionadas nesta pesquisa, a saber: gênero, faixa-etária, escolaridade, classe de palavra que antecede e entonação interrogativa¹¹⁹.

No que tange às variáveis extralinguísticas, esta variante apresenta alguns possíveis condicionamentos de uso, pois, por exemplo, para a variável gênero, o número de realizações foi de 100% na fala das pessoas do gênero feminino. Este dado parece corroborar com os resultados, já citados no primeiro capítulo desta dissertação, que apontam que as mulheres usam mais os RADs. É evidente que estamos tratando de 100% de uso de uma única variante, no entanto, o fato de não ser realizada por homens nesta localidade, pode ser um indicativo que as mulheres usam, não apenas com mais frequência os RADs, mas usam um número de variantes maior, pois no caso desta, a variante *hein?* apenas mulheres a utilizaram. Para ratificar esta consideração, analisamos o *corpus* e não identificamos nenhuma outra variante

¹¹⁸ A variável dependente não foi levada em consideração, pois a forma *hein?*, como já discutido, é um RAD ainda não identificado como tal na língua portuguesa em outros estudos com MCs.

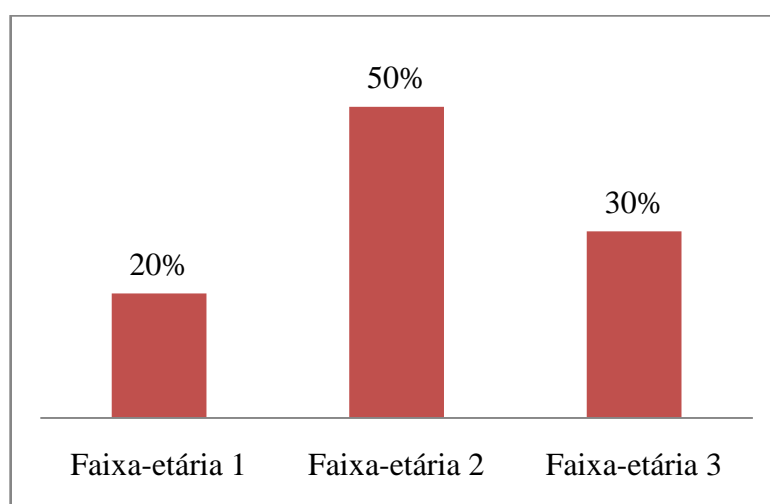
¹¹⁹ Ao selecionamos apenas estas variáveis independentes, desprezamos as demais variáveis que integram este estudo tendo em vista que: i) a variante *hein?*, em relação à massa fônica, é uma forma simples; ii) esta variante é originária da classe de palavras advérbios, consideramos assim conforme apresentamos no capítulo de metodologia; e iii) a localidade em que foi realizada foi apenas em Serra Talhada. Se considerássemos estas variáveis teríamos resultados redundantes.

que apenas foi utilizada pelo gênero masculino, o que endossa ainda mais a ideia de que o gênero feminino parece usar um número maior de RADs e de suas variantes.

A variável escolaridade também se mostrou significativa para esta variante, pois nenhum informante com ensino superior utilizou a forma *hein?*, o que poderíamos considerar que esta variante apresenta-se como muito estigmatizada na comunidade em análise, tendo em vista que pessoas com ensino superior não a utilizaram durante as entrevistas. Contudo, se assim fosse, o maior índice de realização estaria entre pessoas do ensino fundamental e não foi o que os dados de fala apresentaram, pois as ocorrências de *hein?* estratificadas a partir da escolaridade tiveram um percentual de 40% para o ensino fundamental e 60% para o ensino médio, elucidando uma variação muito pequena em torno destes fatores. Sendo assim, ainda mantemos a concepção de que a variante *hein?* seja estigmatizada ao ponto de pessoas com maior grau de formação não a utilizem, contudo este estigma não se estende para as pessoas de níveis médio e fundamental, tendo em vista que, para estes, o uso mostrou ser bem equiparado.

Para a variável faixa-etária acreditamos que uso, assim como para a variável escolaridade, foi bem equiparado, sobretudo para as faixas-etárias 1 e 3, que compreende a pessoas com 12 a 18 anos e acima de 40 anos, respectivamente. Vejamos o gráfico abaixo.

Gráfico 5: Percentual de uso da variante *hein?* estratificado por faixa-etária



As realizações de 20% e 30% para as faixas-etárias 1 e 3, intermediadas por 50% de uso na faixa-etária 2 nos levam a considerar que a variante *hein?* parece ser condicionada pela faixa-etária, pois as pessoas adultas, com idade entre 19 e 40 anos, faixa-etária 2, usam, de maneira significativa, esta forma. As realizações feitas pelas demais faixas podem estar

indicando o período de transição, ou seja, os jovens que a utilizam fazem isso tendo em vista a aproximação da inserção deles na faixa-etária 2 e os mais velhos, acima de 40 anos, a utilizam por perpetuar, de alguma forma, com o uso de uma variante que, no *corpus* analisado, parece ser mais recorrente na fala de pessoas em fase adulta, a faixa-etária 2. Sendo assim, o significativo percentual de 50 % de uso da variante por pessoas com idade entre 19 e 40 anos apontam para uma possibilidade desta forma ser condicionada por esta variável e, assim sendo, evidencia o que consideramos como hipótese deste estudo, ou seja, as variantes dos RADs podem ter seu uso condicionado por variáveis linguísticas e sociais, como são os casos das variáveis gênero, escolaridade e faixa-etária.

No que diz respeito às variáveis independentes linguísticas selecionadas nesta pesquisa, conforme já mencionamos no capítulo 3 desta dissertação, por ocasião da identificação da classe originária desta variante, reconhecemos que o *hein?* parece ser originário da classe dos advérbios, possui uma forma simples, ou seja, não é um RAD composto por duas ou mais formas e em todas as sentenças, 100% das ocorrências, a entonação interrogativa estava presente em toda a sentença e não apenas no RAD, segundo o critério de análise que selecionamos para o presente estudo.

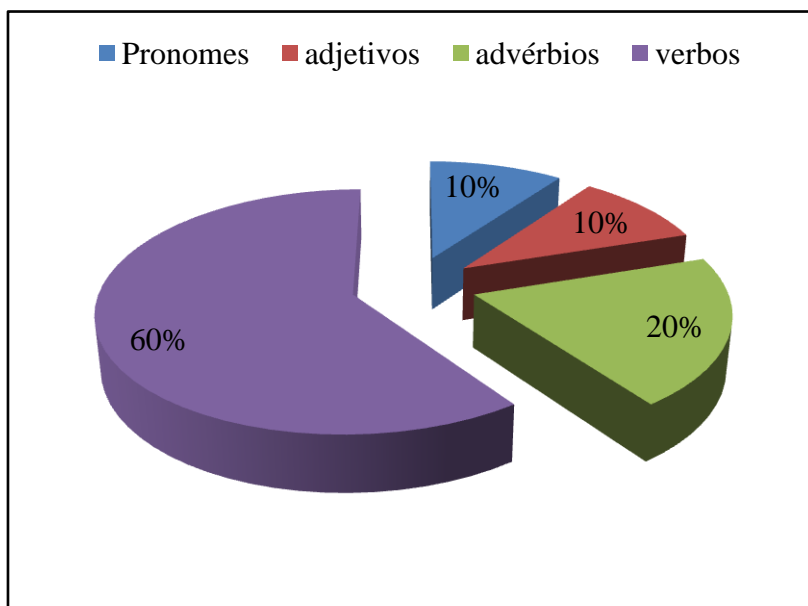
A despeito destas variáveis, um dado é pertinente de ser destacado no que tange à variante *hein?*, trata-se da classe de palavra que antecede o RAD, pois os substantivos foram a única classe de palavras que não antecederam nenhuma das ocorrências dos RADs. Atribuímos esse dado ao fato de esta variante ter sido realizada em sentenças interrogativas, tendo sido, inclusive, usada em estruturas pequenas compostas de verbos e pronomes interrogativos apenas, conforme podemos observar na sentença abaixo.

(56) Será que é, *hein?* (Inf053 – ST)

(57) Como gostar, *hein?* (Inf047 – ST)

Porém, as demais classes de palavras que consideramos ser possíveis de antecederem os RADs, a saber, pronomes, verbos, advérbios e adjetivos, conforme podemos ver no gráfico abaixo, as sentenças foram, a maior parte, finalizadas, antes dos RADs, por verbos.

Gráfico 6: Percentuais de uso estratificados pela classe de palavras que antecede a variante *hein?*



Este dado pode ser atribuído ao que destacamos acima quando mencionamos acerca das sentenças em que o *hein?* apareceu, ou seja, estruturas pequenas compostas, em sua maioria, por verbos e pronomes interrogativos, o que parece justificar, também, o uso do RAD após o pronome em apenas um dos casos das realizações, 10% do total de uso. Nesta sentença, conforme apresentada abaixo, o uso do pronome se dá antes do RAD pelo fato de a informante usar o pronome interrogativo no final da sentença e não no seu início (como poderia acontecer com "Sem parente, o que vou fazer, hein?"), de acordo com o que destacamos em (58).

(58) Sem parente vou fazer o que, *hein?* (Inf053 – ST).

(59) O que vou fazer sem parente, *hein?*¹²⁰

Com base nessas considerações, notamos que a variante *hein?*, realizada nesta comunidade de fala apenas no município de Serra Talhada, parece ter seu uso condicionado por variáveis sociais, como o gênero, escolaridade e faixa-etária, bem como mostrou ser uma variante de RAD usada em sentença interrogativas precedidas, em sua maioria, por verbos. Tais constatações, levam-nos a ponderar a possibilidade de cada variante dos RAD possuir condicionamentos distintos para a suas ocorrências. No entanto, esta possibilidade só poderá

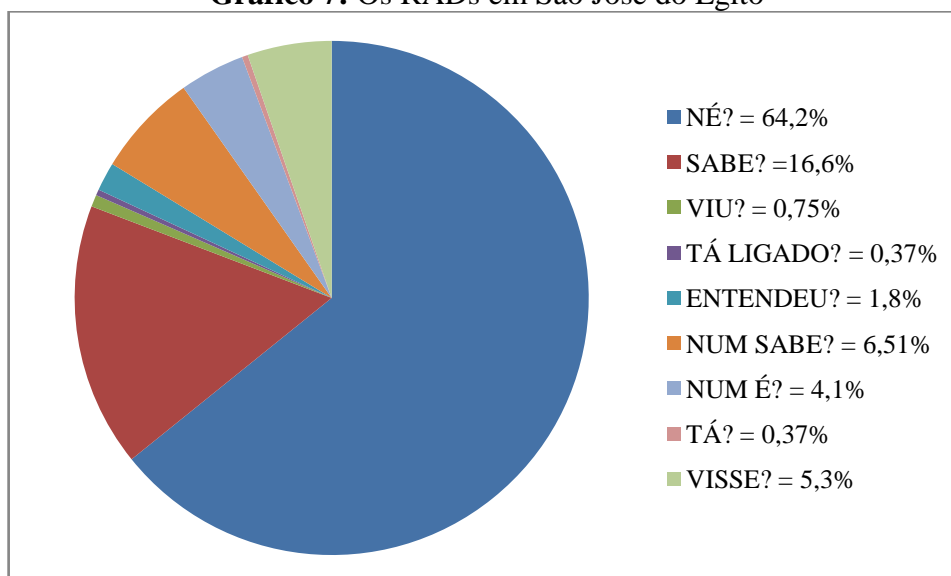
¹²⁰ Esta sentença foi elaborada com base (58) para exemplificar a ordem inversa que o informante usou ao produzir a sentença.

ser acatada ou desprezada após as análises seguintes serem realizadas. Na seção seguinte, apresentamos as variantes e análises realizadas a partir dos dados coletados em São José do Egito.

5.2.2 São José do Egito

Os dados coletados em São José do Egito, em termos de números, em muito se assemelham aos de Serra Talhada. Com um percentual de 39,3%, o equivalente a 265 realizações de RADs na fala do egipsienses. O número de variantes também é semelhante ao de Serra Talhada, pois no município foram contabilizados exatamente nove variantes distintas de RAD, atuando na fala das pessoas desta localidade, conforme está explicitado no gráfico.

Gráfico 7: Os RADs em São José do Egito



As variantes que consideramos aqui como típicas da comunidade de fala em análise, a saber, *tá ligado?*, *num sabe?*, *num é?* e *visse?*, tiveram, em São José do Egito, um índice de ocorrência pequeno, 16,28% no total, ou seja, estas formas juntas somam um percentual de 16,28% divididas em diferentes percentuais de uso, conforme mostra o gráfico¹²¹, para quatro variantes. Esse números foram bem semelhantes aos que foram levantados em Serra Talhada. No entanto, há particularidades nas variantes em São José do Egito que consideramos relevantes de destacar.

¹²¹ 0,37% de ocorrências de *tá ligado?*, 6,51% de *num sabe?*, 4,1% de *num é?* e 5,3% de *visse?*, totalizando 16,28%.

A forma *né?*, em São José do Egito, teve um percentual de uso superior ao percentual de uso desta variante dentro da comunidade de fala como um todo, apontando, assim, para um alto índice de realizações do *né?* neste município, 64,2%, mais precisamente. São José do Egito, das três cidades selecionadas, foi a que os informantes utilizaram mais a variante *né?*. Algo que também aconteceu com a variante *sabe?*, pois, em São José do Egito, teve, de maneira proporcional, um índice de ocorrências maior do que a da comunidade de fala Sertão do Pajeú, com um percentual de 16,6%.

Esses números de ocorrências mais altos para estas variantes sinalizam a possibilidade de algumas formas terem uma possível queda no uso. Talvez seja o caso de as formas consideradas aqui serem típicas da comunidade de fala, pois, como já expomos, estas têm apresentado um baixo índice de realização. Contudo, não foi bem assim que aconteceu.

Ao observarmos no gráfico, as variantes *entendeu?*, *tá?* e *viu?*, formas já consideradas RADs por estudos anteriores, tiveram um índice de ocorrências muito pequeno, é caso da variante *tá?*, com 0,37%, e de *entendeu?*, com 1,8%, pois, das cidades em que foram realizadas, São José do Egito foi onde teve o menor número de realizações.

Sendo assim, parece que o número de uso maior nas variantes *sabe?* e *né?* provocaram uma queda nas realizações das variantes já analisadas enquanto RADs, conforme exemplificamos a partir das variantes *tá?* e *entendeu?*. Ainda acerca da forma *tá?*, algo precisa ser destacado, pois como pontuamos na seção anterior, esta variante parecer ter uma relação com *tá ligado?*. Em São José do Egito, o percentual de uso das duas variantes foi exatamente igual, 0,37%, ou seja, os dados coletados mostram que, em São José do Egito, estas formas parecem estar no mesmo nível de competição, pois o número de ocorrências está bem equiparado. Algo que notamos que se dá em toda a comunidade de fala, ao compararmos as duas variantes, *tá?* e *tá ligado?*, pois, na comunidade de fala Sertão do Pajeú, *tá?* e *tá ligado?* tiveram um índice de 0,75% e 0,60%, respectivamente. Em Serra Talhada, estas formas tiveram uma porcentagem de 1,5% e 1,1% e, em São José do Egito, como destacamos, 0,37%, nas duas variantes. Em Triunfo, não houve ocorrência de nenhuma destas formas. Estes números de ocorrências bem próximos, assim como a não realização em Triunfo, mostram, de alguma forma, o nível de competição destas formas, o que impulsiona, segundo os pressupostos teóricos da Sociolinguística Variacionista, a mudança linguística.

Além dos números destas variantes já apresentados, o número de ocorrências de algumas das variantes consideradas típicas da comunidade de fala, em São José do Egito, teve um percentual de uso superior a essas formas. É o caso das formas *num sabe?* e *num é?*,

pois, das três cidades, São José do Egito foi o município em que estas variantes tiveram um maior número de ocorrências, 6,51% e 4,1%, respectivamente¹²². Elucidando, assim, o que apresentamos na seção anterior quando afirmamos que o uso de cada variante dentro da comunidade de fala possui variações significativas, neste caso, no que diz respeito ao percentual de uso de cada forma.

Das variantes encontradas na comunidade de fala, assim como em Serra Talhada, duas destas variantes não tiveram realizações, *certo?* e *hein?*, a variante *hein?*, como destacamos, teve realizações apenas em Serra Talhada, conforme as análises realizadas. Já para a variante *certo?*, de acordo com o que apresentaremos na seção seguinte, só foi realizada em Triunfo, das três cidades representativas da comunidade de fala. Elucidando, assim, a concepção de que as variantes dos RADs possuem variações entre localidade pertencentes à mesma comunidade de fala.

Além das particularidades das variantes apresentadas até aqui e que foram utilizadas pelos informantes de São José do Egito, outro dado nos chamou a atenção. Trata-se da realização da variante *visse?*, pois, além de ter sido mais recorrente do que variantes já realizadas em outras comunidades de fala, como é o caso da variante *entendeu?*, que teve um percentual de uso tão baixo quanto as variantes consideradas aqui como sendo específicas da comunidade de fala em análise, em São José do Egito, ocorre o mesmo que aconteceu com a variante *hein?* em Serra Talhada, ou seja, esta forma foi realizada apenas neste município.

Com um percentual de uso de 5,3%, *visse?* é uma forma recorrente em São José do Egito, o uso desta variante apenas neste município nos motiva a analisá-la de maneira mais detalhada, como fizemos com a variante *hein?*, em Serra Talhada.

Contudo, antes de apresentarmos a seção que destinaremos à análise desta variante, julgamos pertinente uma comparação com a variante *viu?*, pois esta forma, em São José do Egito, teve um número de ocorrências bem pequeno, 0,75%. Esse número se mostra ainda mais ínfimo quando comparamos com o índice de realizações em outras cidades, como é o caso de Serra Talhada e Triunfo, que, de todas as variantes, foi a segunda mais utilizada, com 23,7% e 2,73%, respectivamente. Daí julgamos pertinente compararmos com a variante *visse?*, pois, em São José do Egito, *visse?* se mostrou ser uma forma típica da comunidade e a variante *viu?*, considerada uma forma já estudada dos RADs, com base nos que já pontuamos, teve um baixo índice de ocorrências. Tal constatação pode ser pertinente para os estudos

¹²² Para facilitar a constatação do que estamos afirmando aqui, expomos os percentuais de ocorrências destas formas nas três cidades, respectivamente. Serra Talhada: 3,1% e 1,5%; Triunfo: 2,73% e 0,69%; São José do Egito: 6,51% e 4,1%.

diacrônicos que sugerimos quando afirmamos, na seção anterior, que as formas típicas da comunidade e as formas já estudadas como RADs podem possuir uma relação de co-ocorrência ou de até mesmo uma forma advir da outra, uma vez que nos municípios em que *visse?* não teve ocorrências, a variante *viu?* teve o percentual de uso abaixo apenas da variante *né?*, como é o caso dos dados coletados em São José do Egito.

A despeito destas considerações, optamos por analisar as ocorrências de *visse?* tendo em vista que esta variante mostrou ser peculiar no município de São José do Egito, integrante, aqui, da comunidade de fala em estudo. Tais análises são apresentadas na seção abaixo.

5.2.2.1 *Visse?*

A variante *visse?*, encontrada no *corpus* em análise, foi algo que em muito chamou nossa atenção, pois esta variante já era conhecida enquanto marcador, como destacamos anteriormente, quando discorremos acerca do trabalho de Irineu e Costa (2015). No entanto, ao ser realizada nesta comunidade de fala, a ideia de que é uma variante específica da cidade do Recife começa a ser questionada. Tal questionamento nos fez observar de maneira mais detalhada as ocorrências de *visse?*, sobretudo no que diz respeito aos falantes que a utilizaram.

Com base no que destacamos na seção anterior, *visse?* teve ocorrências apenas no município de São José do Egito e as pessoas que a utilizaram, de acordo com o questionário social preenchido por eles, não costumam ter o hábito de viajar por outras cidades do estado ou ter contato frequente com pessoas de outras localidades¹²³. Tal dado nos pareceu ainda mais interessante, pois estamos diante de ocorrências de uma variante que tem sido apresentada em estudos e popularmente conhecida como típica de determinada cidade, no entanto, tem realizações em outra comunidade de fala em que seus informantes afirmam não ter nenhum contato com aquela localidade. No entanto, este uso é facilmente compreendido através de informações e noções obtidas durante a coleta de dados do estudo, daí termos enfatizado, no capítulo de metodologia desta dissertação, a importância de o pesquisador adentrar à comunidade de fala e realizar a coleta de dados, pois foi por meio destas práticas que tomamos conhecimento que, em São José do Egito, há um fluxo muito significativo de pessoas de cidades vizinhas, justamente por ser São José do Egito uma

¹²³ No questionário social solicitamos informações como esta, tendo em vista que o fato de o informante viajar e/ou ter contato com pessoas de outras cidades poderia influenciar no uso das variantes dos RADs, pois, conforme defendemos aqui, estes elementos parecem variar de uma comunidade para outra.

localidade que oferece determinada estrutura para a circunvizinhança. Entre as pessoas que circulam, com frequência em São José do Egito, encontramos muitas pessoas da cidade de Iguaracy, município que visitamos quando buscávamos escolher as cidades representantes da comunidade de fala.

O município de Iguaracy, com um pouco mais de 10 mil habitantes, segundo o IBGE, não foi selecionado para representar a comunidade de fala. No entanto, o contato com algumas pessoas da localidade nos fez notar o quanto a forma *visse?* é recorrente nesta localidade em que seus integrantes costumam se deslocar para São José do Egito ou para Recife quando necessitam de atendimentos médicos, educacionais, sobretudo em nível superior, e por questões de negócios profissionais, como é o caso dos comércios de maneira geral. Esta circulação das pessoas entre Recife, Iguaracy e São José do Egito pode disseminar o uso da variante *visse?* que, conforme apontaram Irineu e Costa (2015), é recorrente em Recife¹²⁴.

As ocorrências de *visse?* no Sertão do Pajeú podem, conforme colocamos acima, ser resultado deste contato linguístico muito frequente entre municípios do estado de Pernambuco. No entanto, por estarmos estudando a comunidade de fala Sertão do Pajeú, nos deteremos aqui, nesta seção, à análise de como este uso se dá dentro desta comunidade de fala.

Para procedermos à análise, selecionamos algumas das variáveis do presente estudo e, conseqüentemente, desprezamos outras destas variáveis, pois, por exemplo, para a variável extralinguística localidade, já identificamos que apenas em São José do Egito tivemos ocorrências. Da mesma forma, desprezamos a variável *massa fônica* e *classe de palavra originária da variante*, pois *visse?* é uma forma simples e nos parece claro que tem sua origem em verbos. Sendo assim, a análise da forma *visse?* levará em consideração, conforme apresentamos abaixo, as variáveis extralinguísticas *gênero*, *faixa-etária* e *escolaridade* e as variáveis linguísticas *classe de palavra que antecede o RAD* e a *entonação interrogativa*.

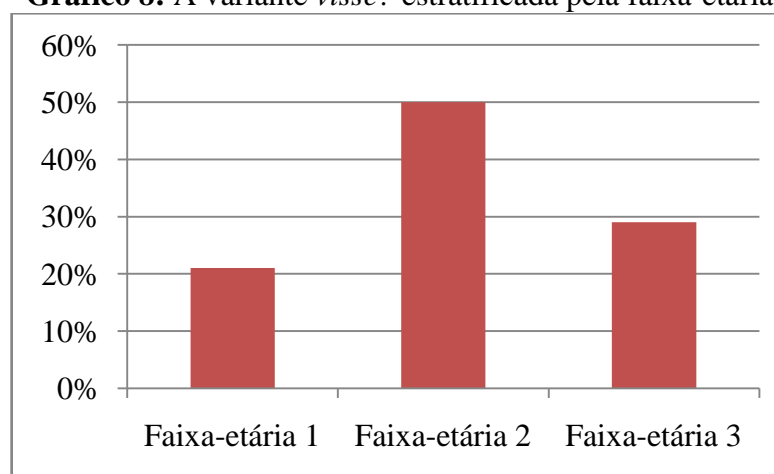
A variante *visse?* teve um total de quatorze ocorrências em São José do Egito e, semelhante à variante *hein?*, que mostrou ser específica em Serra Talhada, entre os municípios integrantes da comunidade de fala, e que só foi realizada por informantes do gênero feminino, as mulheres utilizaram mais esta variante, pois dos informantes do gênero masculino, apenas um homem utilizou-a. Direcionando-nos, assim, à concepção de que as

¹²⁴ É necessário destacarmos que o transitar das pessoas em diferentes cidades é comum aos moradores de Iguaracy, pois por ser um município pequeno, seus habitantes deslocam-se para outras localidades, como São José do Egito, em busca de atendimentos básicos da vida urbana, pois no que diz respeito aos participantes da pesquisa, eles afirmam não ter o costume de viajar para outras cidades, uma vez que no município eles têm o que lhes é necessário para a subsistência.

mulheres não apenas parecem utilizar mais os RADs, mas fazem uso de um maior índice das variantes que aqui consideramos típicas da comunidade de fala em estudo.

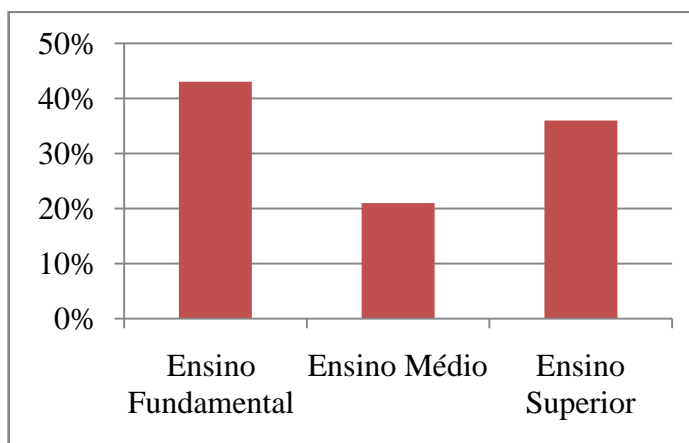
No que diz respeito à faixa-etária, o gráfico abaixo nos mostra um dado interessante. Analisemos.

Gráfico 8: A variante *visse?* estratificada pela faixa-etária



O percentual de uso da variante *visse?* apresenta-se maior na segunda faixa-etária, ou seja, com informantes entre 19 e 40 anos, algo semelhante ao que aconteceu com a variante *hein?*, inclusive, consideramos que a justificativa para este percentual de uso ser mais significativo na segunda faixa-etária pode ser a mesma que apresentamos para a variante *hein?*, a saber, a transição de uma faixa-etária para outra, ou seja, os jovens, por estarem aproximando-se da faixa-etária que esta variante mostra ser recorrente, e os mais velhos, da faixa-etária 3, a utilizam, possivelmente por manterem o uso da variante que se mostra comum na faixa-etária anterior.

As semelhanças entre o uso de *hein?* e *visse?* parecem não se manter no que diz respeito à escolaridade, pois a forma *visse?* apresenta um percentual de uso maior entre os informantes do Ensino Fundamental, 43%, e do Ensino Superior, 36%. Conforme observamos no gráfico abaixo.

Gráfico 9: A variante *visse?* estratificada pela escolaridade

O índice menor no Ensino Médio parece ser justificado quando retomamos a faixa-etária em que este RAD foi mais recorrente, a segunda faixa-etária, pois pessoas com idade entre 19 e 40 anos, que foram as que mais usaram o *visse?*, é mais comum que já tenham concluído o Ensino Médio, ou seja, estejam no nível superior, i. e., pessoas que, por questões que desconhecemos, abandonaram os estudos após o ensino fundamental, nível em que o uso desta forma foi mais recorrente.

Sendo assim, parece ser que as variáveis extralinguísticas condicionam o uso da variante *visse?*, conforme apresentamos, pois as mulheres a utilizaram mais, assim como os informantes da segunda faixa-etária e dos níveis de escolaridade fundamental e superior, mostrando, assim, que fatores sociais parecem interferir no uso desta variante que, conforme defendemos aqui, mostra-se como sendo peculiar neste município selecionado aqui como representante da comunidade de fala.

A partir destes dados, atentaremos para as variáveis linguísticas, no intuito de observar se estas condicionam o uso da variante *visse?* na fala dos egipsienses. Das variáveis linguísticas aqui selecionadas, a entonação interrogativa é uma que merece destaque no que diz respeito ao uso da variante *visse?*, pois, no *corpus*, tivemos um percentual de 100% de ocorrências para a entonação interrogativa apenas no marcador, ou seja, não foram produzidos em sentenças que já tenham uma entonação interrogativa, mas o RAD é o elemento que traz esta entonação, conforme podemos ver nas sentenças abaixo.

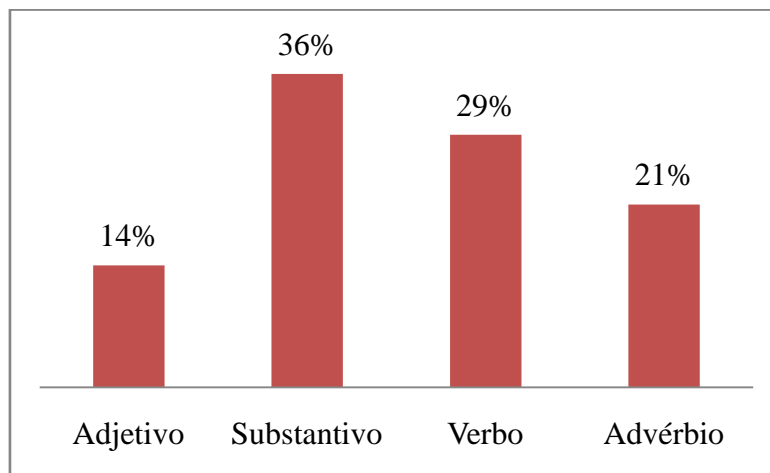
- (60) Homem, eu num lembro muito de meus professores não, *visse?* (Inf013 – SJE)
- (61) Já faz muito tempo, *visse?* (Inf008 – SJE)

O percentual de 100% de ocorrência para a variável *visse?* em sentenças que consideramos como a entonação interrogativa é advinda apenas do marcador é um dado relevante para a pesquisa, pois, estamos lidando aqui com uma variante que consideramos típica não apenas da comunidade de fala em análise, mas de um dos municípios que integram esta comunidade e essas características são exatamente as mesmas da variante *hein?* que apresentamos na seção anterior. No entanto, a variante *hein?* teve um percentual de 100% de ocorrências em sentenças interrogativas e com a forma *visse?*, os 100% de ocorrências foram exatamente na situação contrária, ou seja, a entonação interrogativa vem por meio do RAD, conforme mostramos nos exemplos (60) e (61).

Este dado nos faz ponderar acerca de uma possível característica do elemento em estudo, nesse caso, não do RAD, de maneira específica, mas de suas variantes, pois, conforme já discutimos, o número de variantes desta variável é bem significativo e parece oscilar de uma comunidade de fala para outra, justificando, assim, estudos voltados para a descrição e análises destas variantes. A característica que consideramos que as variantes dos RADs parecem possuir, de acordo com as análises até aqui realizadas, é que cada variante assenta em atender uma necessidade comunicativa diferente do falante, apesar de não excluir o seu papel de pedir aquiescência solicitada ao interlocutor, conforme já defendemos por meio de alguns exemplos, pois parece que há variantes como *hein?* que tem um uso voltado para pedir a atenção do ouvinte em sentenças totalmente interrogativas, já a variante *visse?* mostra-se como sendo recorrente para pedir a atenção do interlocutor em sentenças que não se caracterizam como interrogativas, mas o RAD traz essa entonação interrogativa.

Esta característica da variante é demonstrada também quando analisamos a variável independente linguística nomeada de classe de palavra que antecede o RAD, pois, conforme podemos ver no gráfico abaixo, as classes de palavras que antecedem esta variante são distintas.

Gráfico 10: A variante *visse?* estratificada pela classe de palavra que a antecede



Conforme observamos no gráfico acima, embora o substantivo seja a classe que mais antecede este RAD nas ocorrências deste *corpus*, é possível perceber que, juntamente com as demais classes de palavras, não há nenhuma que sobressaia de maneira muito acentuada, ou seja, há um percentual equilibrado que, embora algumas classes como os adjetivos antecedam menos estes RADs, não distancia de maneira tão significativa dos verbos e estes dos advérbios que, por sua vez, estão bem próximos também do percentual dos substantivos.

Apesar da diferença sinuosa, este dado aponta para uma possibilidade das inúmeras variantes atenderem à necessidade de finalizar as sentenças pedindo a anuência do interlocutor quando estas são finalizadas por diferentes classes de palavras, visto que as classes de palavras que antecedem o *visse?* oscilam entre quatro das cinco classes que selecionamos como possíveis, não tendo ocorrências de pronomes como antecedentes do *visse?*, dado que pode estar associado a ocorrência da variante em sentenças que não apresentam entonação interrogativa, pois com a variante *hein?* analisada em seção anterior, a classe pronome antecede esta forma e, nestes casos, o percentual de ocorrências do pronome antecedendo o RAD é 10%.

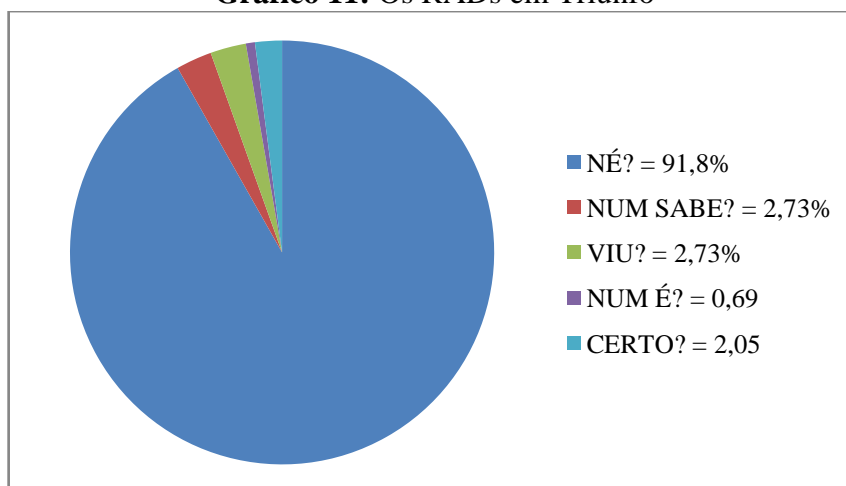
Com base nestas considerações, atentamos para os seguintes dados: a variante *visse?* possui ocorrências no Sertão do Pajeú e estas ocorrências de uso mostram-se como sendo condicionadas por fatores linguísticos e sociais, de acordo com o que apresentamos aqui.

Além de Serra Talhada e São José do Egito, selecionamos o município de Triunfo para representar a comunidade de fala em análise e, da mesma sorte, consideramos necessário apresentar e analisar o uso dos RADs nesta localidade, algo que nos propomos a fazer na seção seguinte.

5.2.3 Triunfo

O município de Triunfo já foi mencionado neste capítulo tendo em vista que o baixo número de realizações dos RADs chamou nossa atenção, sobretudo porque este dado pode estar relacionado ao aparente alto índice de realizações dos MCs iniciadores. No entanto, há, nas ocorrências das variantes, neste município, aspectos que consideramos importantes de serem destacados. É o caso, por exemplo, do número de variantes nesta cidade, cinco formas apenas atuam na fala dos triunfenses como sendo RAD, conforme podemos ver nos dados apresentados no gráfico abaixo.

Gráfico 11: Os RADs em Triunfo



A variante *né?*, conforme podemos ver no gráfico, teve um alto índice de realizações na cidade de Triunfo, 91,8% no total. Entretanto, é interessante destacarmos que, das três cidades, Triunfo foi a cidade que teve, além de um pequeno número de ocorrências dos RADs, um número menor de formas atuando como RADs, porém as variantes que são realizadas como RADs neste município, excetuando-se o *né?* e o *viu?*, são formas que aqui consideramos como formas peculiares da comunidade de fala.

Sendo assim, se nos demais municípios o número de formas já mencionadas como RADs foi maior, independente do número de ocorrências de cada variante, em Triunfo o número de variantes foi menor, mas o número de variantes consideradas como típicas da comunidade de fala foi maior, pois das onze variantes encontradas na comunidade de fala, cinco são consideradas formas que mostram ser específicas da comunidade e em Triunfo, três destas formas foram realizadas, a saber, *num sabe?*, *num é?* e *certo?*.

Além disso, consideramos pertinente destacar que a variante *num sabe?*, em termos percentuais, teve um número de realizações igual ao da forma *viu?*, ou seja, esta variante teve um percentual de uso igual ao da variante *num sabe?* que consideramos aqui como típica da comunidade de fala. A partir destas considerações, podemos mencionar não apenas o quanto o *né?* possui um alto índice de ocorrência na fala dos triunfenses, 91,8%, conforme apresentamos no parágrafo anterior, mas que as demais formas se alternam na fala dos triunfenses enquanto RADs, independentemente se são formas já descritas em estudos ou se formas que o presente estudo mostra que estão atuando como RADs nesta comunidade.

A variante *num é?* teve um percentual de uso de 0,69% em Triunfo, o que corresponde a uma única ocorrência na fala dos triunfenses. Pelo fato de nas demais cidades o percentual de uso desta variante não ter sido tão ínfimo assim, 1,5% em Serra Talhada e 4,1% em São José do Egito, decidimos observar de maneira mais atenta para esta realização. Para tanto, é necessário rever a sentença em que esta variante foi utilizada. Apresentamos o dado abaixo com o intuito de procedermos a análise.

(62) E: E a sua família mora toda aqui? todo mundo junto?

I: Vive assim fora uns dos outro, *num é?* (Inf 025 – TR)

Produzida por um homem com faixa etária acima de 40 anos, o dado mostra que a sentença não tem entonação interrogativa e que apenas o RAD dá esse caráter de pergunta, uma vez que a sentença traz uma declaração/resposta acerca do que o entrevistador perguntou ao informante. Este informante, com escolaridade de ensino fundamental, afirma, por meio do questionário social, que não costuma ter atividades de leitura, mas enquanto profissional que trabalha nas feiras populares da região, em cidades vizinhas, pode fazer uso deste RAD motivado por contatos linguísticos frequentes que mantém com pessoas de outras localidades, como Serra Talhada, por exemplo, em que a variante *num é?* teve um índice de ocorrência mais alto.

Com base nesta concepção, poderíamos ponderar que a variante *num é?* parece não ser específica de Triunfo, semelhante as variantes *sabe?*, *tá ligado?*, *entendeu?* *hein?*, *visse?* e *tá?* que não tiveram realizações neste município, mas uma variante que está presente na comunidade de fala, conforme o índice de realização nesse *corpus* mostra. Mesmo assim, não descartamos a consideração desta forma ser variante da variável em estudo no Sertão do Pajeú, porque em outras cidades desta comunidade seu uso se mostrou bastante significativo.

Consideramos, contudo, que o que pode estar ocorrendo é que esta forma tenha, em determinados municípios, um uso mais assíduo e, conforme destacamos na metodologia deste estudo, os municípios desta comunidade de fala, de tão próximos, acabam por ter um fluxo de pessoas que transitam de um para o outro a depender de suas necessidades educacionais, médicas, comerciais etc. E esse transitar, acreditamos, acaba por influenciar no uso e nas variantes dos RADs utilizadas na fala das pessoas.

Um caso diferente da variante *num é?* é o da forma *certo?*, pois, no caso desta, apenas em Triunfo, dos três municípios representantes da comunidade, obtivemos ocorrências, como se deu com a variante *hein?* e *visse?* em Serra Talhada e São José do Egito, respectivamente. Não desconsideramos a possibilidade de estar ocorrendo com estas formas o mesmo que defendemos que aparenta estar se dando com a variante *num é?* na fala do informante apresentada acima, ou seja, que ela esteja presente na fala das pessoas de outros municípios do Sertão do Pajeú, mas como apenas em Triunfo, das cidades em que os dados foram coletados, esta variante teve realizações, decidimos analisá-la separadamente, de acordo com o que expomos na seção abaixo.

5.2.3.1 Certo?

A variante *certo?*, neste *corpus* em análise, teve algumas peculiaridades, pois foi, por exemplo, em toda a comunidade de fala a variante com menor percentual de realizações, de acordo com o que apresentamos na segunda seção deste capítulo. Além disso, foi utilizada apenas por falantes do município de Triunfo o que justifica uma análise desta forma de RAD.

Para tanto, lançamos mão de algumas variáveis linguísticas e sociais para compreendermos melhor o uso de 2,05% deste RAD na cidade de Triunfo. Das variáveis sociais deste estudo, desprezamos a localidade, por questões já apresentadas aqui, ou seja, só em Triunfo tivemos o uso desta forma. Das variáveis linguísticas, desprezamos a classe de palavra originária do RAD, pois nos parece claro que trata-se de um adjetivo que, em contextos como estes, é utilizado para requisitar um apoio no discurso do falante. Além desta variável linguística, descartamos a massa fônica, pois é notório que a forma *certo?* é uma forma simples de RAD. Sendo assim, propomo-nos a observar, nos dados coletados, as variáveis *gênero*, *faixa-etária* e *escolaridade*, assim como, as variáveis linguísticas nomeadas de *classe de palavra que antecede o RAD* e *entonação interrogativa*.

Sendo assim, expomos as ocorrências de *certo?* na fala dos triunfenses.

(63) Deixando as famílias, *certo?* (Inf027 – TR)

(64) Miguel Falabela, *certo?* (Inf027 – TR)

(65) Sônia Braga, *certo?* (Inf027 – TR)

De acordo com as realizações expostas em (63), (64) e (65), notamos que todas as três sentenças que resultaram em um percentual de uso de 2,05% do total de ocorrências da variante *certo?* em Triunfo, são em sentenças cuja entonação interrogativa estava apenas no RAD, assim como, em todas as realizações, a variante *certo?* foi antecedida por substantivos, evidenciando assim a não variação no que diz respeito às variáveis linguísticas.

Após esta identificação, atentamos para as variáveis sociais e percebemos que as três ocorrências desta forma aconteceram na fala do mesmo informante, este, por sua vez, é um homem com idade acima de quarenta anos, com ensino superior e, por meio de seu questionário social não há informações que permitam justificar o uso desta variante, como um possível contato linguístico com informantes de outras localidades, por exemplo, pois o informante afirma que não costuma viajar para outros estados e que em sua rotina apenas transita em algumas cidades de Pernambuco¹²⁵.

É necessário destacar algumas considerações acerca deste RAD no que diz respeito às ocorrências no *corpus* em análise: i) consideramos como variante da comunidade de fala e específica do município de Triunfo, pois o informante que a utilizou possui todas as características que o inserem nestes grupos; ii) avaliamos a possibilidade de esta forma ser uma variante de outro município de Pernambuco, pois, de acordo com o uso de *hein?* em Serra Talhada, parece-nos possível que determinadas variantes sejam específicas não apenas da comunidade fala como um todo, como o Sertão do Pajeú, mas de uma comunidade de fala menor como falantes de um mesmo município, como é o caso de Serra Talhada. No entanto, só estudos que analisem estas comunidades podem debater com mais propriedade tal dado.

Por todas estas colocações, percebemos que o uso das diferentes variantes parece ser específico em cada cidade representativa da comunidade de fala em análise, possuindo variantes que só são realizadas no município, como é o caso das variantes *hein?*, *visse?* e *certo?*, que foram realizadas apenas em Serra Talhada, São José do Egito e Triunfo, respectivamente. Além disso, percebemos que há variantes que oscilam em termos de uso de uma cidade para outra, como é o caso da forma *viu?*, que em Serra Talhada possui um alto

¹²⁵ Apesar do pequeno índice de realização, acreditamos que estudos posteriores com outros municípios de Pernambuco podem objetivar identificar se esta forma é recorrente em outros municípios e/ou em determinada comunidade de prática presente no Sertão do Pajeú.

índice de ocorrência e em São José do Egito está entre as variantes com menor número de realizações.

Sendo assim, sugestionamos, após as análises realizadas, a possibilidade de os RADs terem, além de variantes específicas de determinada comunidade de fala, formas distintas que atuam no sistema linguístico a depender da sentença produzida, no que tange à estrutura da sentença, apontando para a interferência das variáveis linguísticas, bem como de fatores sociais do falante, ou seja, das variáveis sociais que parecem condicionar o uso de algumas das variantes dos RADs aqui estudados.

Na seção seguinte, os dados coletados serão analisados a partir das variáveis independentes linguísticas e sociais. No entanto, terão seus percentuais de uso e pesos relativos levantados a partir da variável dependente presença/ausência de marcador já identificado na língua portuguesa, ou seja, formas que mostram ser típicas da comunidade de fala em análise ou formas que estudos anteriores já apontaram como sendo RAD.

5.3 OS RESULTADOS DO VARBRUL

Os dados coletados são apresentados nesta seção a partir do tratamento estatístico dado por meio do Programa Computacional GoldVarbX, programa pertencente ao pacote VARBRUL. Para que o programa possa apresentar os percentuais de uso e peso relativo de cada variável independente linguística e social é necessário que a variável dependente seja estabelecida e, no presente estudo, esta foi considerada como sendo a presença/ausência de RADs – Requisitos de Apoio Discursivo – já descrito na Língua Portuguesa, uma vez que, a partir do objetivo principal deste estudo, temos como hipótese que podem existir formas que nesta comunidade de fala ainda não foram descritas em outras pesquisas.

Ao estabelecermos a variável dependente, é necessário também informar ao programa a partir de qual destas variáveis o valor de aplicação será dado, nesta pesquisa selecionamos a ausência de RAD já descrito na Língua Portuguesa, ou seja, os números apresentados pelo programa computacional são relacionados às formas que ainda não foram analisadas enquanto RADs, ficando subentendido para o pesquisador os números em percentual e peso relativo da presença de RADs já descritos na Língua Portuguesa¹²⁶.

¹²⁶ Afirmamos que os números da variável dependente não selecionada para dar o valor de aplicação fica subentendido ao pesquisador, porque o programa traz o peso relativo apenas da variável dependente selecionada

Antes de apresentarmos o peso relativo dado pelo programa para cada variável, julgamos pertinente apresentarmos os dados apontados pelo programa nas rodadas iniciais realizadas, trata-se do percentual de uso de cada variável dada a partir da variável dependente. No quadro abaixo, apresentamos em termos reais e percentuais o uso na comunidade fala dos RADS já descritos e dos não descritos em estudos com os marcadores.

Quadro 17: Percentual de uso das variantes dos RADS descritas e não descritas em Língua Portuguesa

Formas não descritas	Formas já descritas
72 10,7%	598 89,3%

A partir do quadro exposto acima, é possível termos um percentual de uso calculado a partir da distinção que realizamos, a saber, formas descritas e não descritas de RADS. Por meio dos dados expostos acima, podemos perceber que as variantes de RADS ainda não descritas na Língua Portuguesa tem um percentual de uso bem pequeno, 10,7%. No entanto, é válido destacar que estamos considerando estas formas como sendo específicas da comunidade de fala em análise, pois o fato delas não terem sido descritas aponta não apenas para a não descrição destas variantes, mas para o seguinte dado: os estudos já realizados com RADS em outras comunidades não apresentaram ou descreveram estas formas, o que nos leva a considerar a possibilidade destas formas não terem tido ocorrências de uso nas comunidades de fala analisadas nestes estudos.

Ainda que as formas não descritas tenham um percentual de uso pequeno, consideramos pertinente descrevê-las e analisá-las uma vez que este estudo pode ser um importante ponto de partida para estudos posteriores que busquem sequenciar a descrição de variantes de RADS em outras comunidades de fala e este percentual já mostra que o uso de outras formas de RADS é real na Língua Portuguesa.

Partindo ainda dos dados expostos no quadro acima, é necessário destacarmos uma informação importante, trata-se do número de ocorrência dos RADS, pois no início deste capítulo, apresentamos um total de 673 realizações no *corpus* em análise, no entanto, a soma do total de uso apresentada no quadro acima é de 670 ocorrências. Isso se dá pela seguinte razão, o programa apresentou, nas rodadas iniciais, um KnockOut, ou seja, umas das variantes realizadas não apresentou variação em nenhuma das variáveis selecionadas.

para dar o valor de aplicação, este valor vai de 0 a 1. Se para a variável selecionada o valor for de, por exemplo, 0.4, é evidente que o peso relativo da variável não selecionada é de 0.6.

Ao identificarmos o KnockOut apontado pelo GoldVarb X, constatamos que isso se deu por meio da variante *certo?*, ou seja, esta forma foi utilizada por um mesmo falante e em sentenças com estruturas semelhantes, não evidenciando, assim, variação e, conseqüentemente, condicionamento por meio de variáveis linguísticas e sociais¹²⁷.

Tendo em vista o KnockOut apresentado, foi necessário retirarmos as ocorrências desta variante dos dados informados ao programa computacional, uma vez que o programa necessita a retirada do KnockOut para a continuidade das rodadas a serem realizadas até a apresentação do peso relativo de cada variável. Sendo assim, foram excluídas as três ocorrências da variante *certo?* nesta etapa, resultando, assim, em 670 ocorrências de RADs, conforme podemos ver no quadro apresentado acima. É a partir deste número de ocorrências que os percentuais de uso que se seguem serão calculados, uma vez que a variante *certo?* foi desconsiderada pelo programa computacional.

Após a apresentação do percentual de uso das variantes dos RADs já catalogadas e das não catalogadas, o GoldVarb X nos apresentou o percentual de uso destas formas a partir das variáveis linguísticas e sociais selecionadas para a pesquisa. A apresentação e discussão destes percentuais nos são importantes, pois por meio deles já podemos perceber quais das hipóteses levantadas neste estudo apontam para uma confirmação ou refutação. Afirmamos que estes percentuais já podem apresentar estes indícios, porque levamos em consideração que os pesos relativos, que serão apresentados na seção seguinte deste capítulo, trazidos na última rodada do programa, é que serão, de fato, o material a nos dar condições de comprovarmos ou não as hipóteses deste estudo.

Após a apresentação dos termos percentuais das formas já catalogadas ou não, o programa nos forneceu o percentual de uso destas variantes estratificadas pelo gênero do informante. Os números são expostos no quadro abaixo.

Quadro 18: As formas descritas e não descritas de RADs estratificados por gênero

	Formas não descritas	Formas já descritas	Total
Homens	31 9,5%	296 90,5%	327 48,8%
Mulheres	41 12%	302 88%	343 51,2%

¹²⁷ Não aprofundaremos aqui a discussão em torno deste KnockOut tendo em vista que na seção anterior já apresentamos algumas considerações acerca desta variante e deste uso por parte de um único informante.

De acordo com os percentuais apresentados, percebemos que parece ser que esta variável não apresenta-se como sendo significativa para a variação no uso dos RADs encontradas no corpus, uma vez que, conforme podemos observar na coluna intitulada Total que traz o resultados expostos na linha horizontal do quadro, o uso dos RADs por homens e mulheres nesta comunidade de fala mostra-se bastante equiparados, 48,8% e 51,2%, respectivamente. O uso dos RADs sem uma significativa alteração a partir da variável social gênero tem sido apontado em pesquisas anteriores, como a de Valle (2001) e Freitag (2008), contrapondo às considerações acerca desta variável apresentadas por Macedo e Silva (1996), pois para estas autoras, e os resultados de suas pesquisas naquelas comunidades e naquele período de tempo, a variável gênero mostrava como sendo significativa para a variação, pois as mulheres, conforme já discutimos neste estudo, quando citamos a pesquisa das autoras, usaram mais os marcadores.

É válido destacar, no entanto, que as assertivas acerca da variável gênero apresentadas por Macedo e Silva (1996) dizem respeito ao uso dos MCs como um todo e os resultados de Valle (2001), Freitag (2008) e o que apresentamos aqui dizem respeito ao uso específico dos RADs, ou seja, a variável gênero pode condicionar o uso dos MCs, mas não parece ser tão significativa quando é analisada de maneira específica no uso dos RADs.

Como estamos lidando com formas descritas e não descritas dos RADs, optamos por observar o percentual de uso destas formas a partir do gênero e, de acordo com o que está exposto no quadro acima, apesar de o percentual de uso das formas não catalogadas ser 12% na fala das mulheres e 9,5% na fala dos homens, acreditamos que não são diferenças tão significativas, tendo em vista, inclusive, que o percentual de uso na comunidade de fala em análise das formas já descritas também se mostra equiparado, 90,5% na fala dos homens e 88% na fala das mulheres nos dados coletados.

Sendo assim, no *corpus* analisado a variável gênero parece não ser tão significativa para a alternância de uso das variantes de RADs encontradas na fala dos sertanejos, o mesmo parece não acontecer com a variável faixa-etária. Para decorrermos com as análises, expomos os números no quadro abaixo.

Quadro 19: Formas descritas e não descritas estratificadas pela faixa-etária

	Formas não descritas	Formas já descritas	Total
Faixa-etária 1 (12 a 18 anos)	8 4,4%	174 95,6%	182 27,2%
Faixa-etária 2 (19 a 40 anos)	24 10,1%	213 89,9%	237 35,4%
Faixa-etária 3 (acima de 40 anos)	40 15,9%	211 84,1%	251 37,4%

Ao analisarmos os números apresentados com base nesta variável social, percebemos que parece ser que quanto mais jovem os informantes menos eles usaram os RADs, pois para a faixa-etária 3, por exemplo, o percentual de uso foi de 37,4%, enquanto que na faixa-etária 1, este percentual foi 27,2%, ou seja, os jovens parecem que utilizam menos os RADs. Da mesma forma, a faixa-etária mostra-se como sendo um fator que pode condicionar o uso dos RADs, quando analisamos o percentual de uso das variantes já descritas e não descritas nos estudos com os RADs, sobretudo quando analisamos o percentual de realizações das formas não descritas, pois quanto menor a faixa-etária menor o número de ocorrências dos RADs, da faixa-etária 1 a faixa-etária 3, por exemplo, há um diferencial de 11,5% no percentual de uso.

No que tange as variantes que já foram descritas, algo contrário parece ocorrer, quanto menor a faixa-etária, maior o percentual de uso, pois na faixa-etária 1, por exemplo, obteve-se 95,6% de ocorrências e na faixa-etária 3 um número de 84,1%. Sendo assim, parece ser que as pessoas mais jovens apesar de utilizarem menos os RADs, parece ter este uso mais marcado pelas formas que já foram catalogadas em outras pesquisas, ou seja, formas já consideradas como RADs na Língua Portuguesa. As variantes, tratadas aqui como formas ainda não descritas, mostram-se mais recorrentes na fala das pessoas mais velhas. Tal dado, apresenta a possibilidade desta variável ser significativa para a variação linguística dos RADs, assim como, contrapõe nossa hipótese de que as pessoas mais jovens utilizam mais os RADs tendo em vista que participam de inúmeras atividades diárias de interação, pois os mais jovens parecem usar mais apenas os RADs já descritos, pois os ainda não catalogados tem um percentual de ocorrências bem menor da fala das pessoas mais jovens.

Além das variáveis gênero e faixa-etária, apresentamos no quadro abaixo as formas estudadas estratificadas a partir da localidade. Embora estejamos analisando uma única

comunidade de fala, ela é composta por diferentes municípios e julgamos que analisarmos um número de ocorrências desta variável dependente binária pode ser significativo.

Quadro 20: Formas descritas e não descritas estratificadas pela localidade

	Formas não descritas	Formas já descritas	Total
São José do Egito	42 15,8%	223 84,2%	265 39,6%
Triunfo	5 3,5%	138 96,5%	143 21,3%
Serra Talhada	25 9,5%	237 90,5%	262 39,1%

O percentual de uso dos RADs por localidade já mencionamos na seção anterior, por isso não iremos repeti-lo aqui, contudo, o percentual de uso das formas já catalogadas e não catalogadas nos chamam a atenção, pois nos três municípios o uso das variantes já analisadas é mais recorrente que o uso das variantes ainda não descritas. No entanto, em Triunfo as ocorrências das formas não catalogadas, 3,5%, é bem inferior ao dos outros municípios, pois em Serra Talhada este percentual é de 9,5% e em São José do Egito é de 15,8%. Porém, destacamos que este percentual inferior pode estar relacionado ao número de variantes, conforme apresentamos na seção anterior, este número foi bem pequeno no município de Triunfo, além disso, as ocorrências do RADs nesta cidade teve um percentual de realizações muito alto na variante *né?*, 91,8% o que pode justificar o baixo percentual de uso das formas não descritas.

Além disso, destacamos que a forma *certo?* foi retirada desta quantificação, interferindo no percentual de uso destas formas, tendo em vista que esta variante é aqui considerada como uma forma de RAD ainda não catalogada em Língua Portuguesa. Sendo assim, no *corpus* analisado há municípios em que os RADs foram menos recorrentes, como é o caso de Triunfo, e as formas já catalogadas são mais frequentes na fala dos sertanejos mesmo quando estratificadas pela localidade, mostrando-se, assim, como uma variável que parece condicionar o uso das variantes, sobretudo no que diz respeito às formas ainda não descritas em estudos anteriores.

Das variáveis sociais selecionadas, a escolaridade é uma das que menos mostrou apresentar interferência no uso das variantes dos RADs. Os números são expostos no quadro abaixo e a análises realizadas a partir deles são apresentadas posteriormente.

Quadro 21: Formas descritas e não descritas estratificadas pela escolaridade

	Formas não descritas	Formas já descritas	Total
Ensino Fundamental	34 13,7%	215 86,3%	249 37,2%
Ensino Médio	22 10,4%	190 89,6%	212 31,6%
Ensino Superior	16 7,7%	193 92,3%	209 31,2%

Apesar de observarmos que os informantes com Ensino Fundamental tenham utilizado mais os RADs, 37,2%, sendo o percentual de uso dos informantes com Ensino Médio e Superior semelhantes, 31,6% e 31,2%, respectivamente, não consideramos como uma variável significativa, pois além de, no total, os números de ocorrências serem bem próximos, notamos que algo semelhante acontece quando analisamos os números de ocorrências das variantes dos RADs descritas e não descritas na Língua Portuguesa, pois o percentual de realização é bem próximo entre os três níveis de escolaridade, ou seja, esta variável não parece condicionar o uso dos RADs na fala dos informantes desta pesquisa.

Com base nos números apresentados, percebemos que das variáveis independentes sociais selecionadas neste estudo, a saber, gênero, faixa-etária, localidade e escolaridade, apenas a faixa-etária apresenta-se como sendo significativa, pois os mais jovens parecem usar menos não apenas os RADs, mas as formas nomeadas aqui como não descritas na Língua Portuguesa.

A partir destas análises é conveniente apresentarmos o percentual de uso das variáveis independentes linguísticas selecionadas nesta pesquisa quantificadas pelo programa computacional a partir da variável dependente presença/ausência de formas já descritas como RADs em Língua Portuguesa. Começamos estas análises a partir da classe de palavra originária dos RADs.

No entanto, antes de apresentarmos o quadro com os números dados pelo GoldVarb X, é necessário destacarmos que selecionamos três classes de palavras como possíveis de dar origem aos RADs estudados, verbos, adjetivos e advérbios. Contudo, ao atribuir estas classes de palavras as formas, julgamos pertinente desconsiderar a classe dos advérbios, pois esta parece originar apenas uma única forma, o *hein?*, variante considerada como não descrita em

estudos dos marcadores, consideramos pertinente descartar tendo em vista que o programa calcula o percentual de uso a partir da variável binária presença/ausência de formas já descritas e, neste caso, a variante originada a partir dos advérbios é apenas a forma *hein?* o que não ocasiona a variação.

Além da classe dos advérbios como originária dos RADs, desconsideramos a de adjetivos, pois nos *corpus* analisado a única variante considerada como sendo originada por adjetivos foi a forma *certo?* e, como destacamos, as ocorrências desta variantes foram retiradas tendo em vista que o programa acusou um KnockOut.

Apesar da significativa redução na quantificação desta variável linguística, não consideramos a possibilidade de descartá-la tendo em vista que, conforme veremos nos quadro abaixo, podemos obter o percentual de ocorrências de formas já descritas ou não originárias de verbos.

Quadro 22: formas descritas e não descritas estratificados pela classe de palavra originária do RAD

	Formas não descritas	Formas já descritas	Total
Verbos	62 10,7%	597 90,6%	659 98,4%

Conforme destacam Marcuschi (1989) e Macedo e Silva (1996), os marcadores são elementos linguísticos originados, em sua maioria, por verbos. No *corpus* em análise, notamos que o número de ocorrências com RADs originários de verbos é significativo, 98,4%, sendo que nas formas ainda não descritas, o verbo origina 10,7%. Isso pode se dá tendo em vista o número menor de ocorrências destas formas no corpus em análise, além de que, as duas classes pertencentes a esta variável linguística desconsideradas nesta rodada foram com formas ainda não descritas, ou seja, interfere para que este número seja ainda menor.

Por termos desconsiderado dois fatores da variável e pelos resultados apresentados acima no que tange a classe dos verbos, consideramos que esta variável linguística não parece ser significativa para a variação dos RADs, uma vez que o próprio programa selecionou apenas uma das três possibilidades dadas para a variável linguística e nesta o percentual de uso foi muito acentuado nas formas já descritas, algo que estudos anteriores já apontavam: que os RADs parecem possuir essa característica de ser originários de verbos.

Diferentemente da variável citada acima, que não mostrou ter relevância para a variação, a variável independente linguística massa fônica apresentou um percentual de uso que merece destaque, sobretudo quando analisamos a partir da variável dependente presença/ausência, conforme podemos ver no quadro abaixo.

Quadro 23: Formas descritas e não descritas dos RADs estratificados pela massa fônica da variante

	Formas não descritas	Formas já descritas	Total
Forma simples	24 3,9%	595 96,1%	619 92,4%
Forma composta	48 94,1%	3 5,9%	51 7,6%

Os números apresentados no quadro mostram que, de maneira geral, os RADs são mais recorrentes em formas com massa fônica menor, ou seja, formas simples, pois o programa apresenta um percentual de uso de 92,4% para as formas simples e um percentual de apenas 7,6% para as formas compostas. Contudo, ao analisarmos esta variável a partir da variável dependente notamos que o contrário parece acontecer, pois as formas não descritas mostram-se como sendo formas com massa fônica maior, 94,1%, enquanto que as formas já descritas possuem um percentual de 96,1% em formas com massa fônica menor, ou seja, formas simples.

Embora tenhamos apresentado de maneira especulativa, na seção anterior, uma possível relação entre as formas descritas e não descritas dos RADs, consideramos que o dado apresentado acima corrobora com a hipótese citada na referida seção, pois as formas descritas, segundo os números de ocorrências apresentados, parecem ser condicionadas por menor massa fônica, ou seja, seriam as formas originais dos RADs, e as formas não descritas parecem ser formas mais compostas, ou seja, as formas originadas dos RADs já descritos. Contudo, tal consideração é uma hipótese levantada no presente estudo que só pode ser confirmada com estudos posteriores que, inclusive, façam uma abordagem diacrônica do uso dos RADs.

As discussões apresentadas no primeiro capítulo desta dissertação apontam uma característica marcante dos RADs, trata-se da posição do RAD na sentença, ou seja, eles ocupam a posição final da sentença, uma vez que tem o caráter de pedir uma resposta ao interlocutor. Sendo assim, temos como hipótese de que a classe de palavra que antecede o

RAD pode ser uma variável linguística relevante, tendo em vista que este aparece no fim da sentença.

Com base nisso, estratificamos, a partir da variável dependente, os dados coletados a partir da classe de palavra que antecede o RAD. Os números percentuais são expostos no quadro abaixo.

Quadro 24: Formas descritas e não descritas de RADs estratificadas pela classe de palavra que antecede o RAD

	Formas não descritas	Formas já descritas	Total
Substantivo	29 10,9%	236 89,1%	265 39,6%
Adjetivo	11 8,1%	125 91,9%	136 20,3%
Verbo	19 14%	117 86%	136 20,3%
Advérbio	9 8,3%	99 91,7	108 16,1%
Pronome	4 16%	21 84%	25 3,7%

Os números apresentados no quadro acima apontam para um dado interessante: os substantivos parecem ser a classe de palavra mais recorrente no que diz respeito a anteceder o RAD, 39,6% no total. Acreditamos que isso se dá em virtude da própria estrutura sintática do português, uma vez que pronomes, por exemplo, que teve um percentual de ocorrências de 3,7% parece ser menos comum na posição final da sentença. Os adjetivos e verbos tiveram um percentual equiparado no que diz respeito a anteceder o RAD, 20,6%, algo que, assim como os substantivos, pode se dá em virtude da estrutura sintática da Língua Portuguesa.

No que tange a variável dependente e a variável linguística classe de palavra que antecede o RAD, os resultados oscilam um pouco se comparados aos números totais apresentados no parágrafo anterior, pois as formas não descritas, por exemplo, parecem ser mais antecedida por pronomes e verbos, 16% e 14%, respectivamente. Enquanto que para as variantes já catalogadas, os percentuais apresentam-se bem equiparados entre as classes de palavras selecionadas neste estudo. Consideramos que isso pode se dá tendo em vista que estas formas já catalogadas possam estar presente no sistema linguístico do falante há uma tempo por se tratarem, possivelmente, de formas que são recorrentes na língua portuguesa e,

em sendo assim, o falante pode as utilizar em sentenças como diferentes estruturas, neste caso, em sentenças que são concluídas com classes de palavras distintas.

Convém apresentar ainda os números obtidos para a variável linguística entonação interrogativa visto que esta integra o grupo das quatro variáveis linguísticas selecionadas neste estudo. Sendo assim, o quadro abaixo apresenta os números percentuais desta variável obtidos a partir da variável dependente.

Quadro 25: Formas descritas e não descritas dos RADs estratificados pela entonação interrogativa

	Formas não descritas	Formas já descritas	Total
Entonação Interrogativa só no RAD	60 9,7%	558 90,3%	618 92,2%
Entonação Interrogativa em toda a sentença	12 23,1%	40 76,9%	52 7,8%

É pertinente destacar, antes mesmo das análises dos números apresentados no quadro acima, os fatores que integram esta variável, pois ao nomearmos entonação interrogativa só no RAD e entonação interrogativa em toda a sentença é necessário esclarecer como esta entonação interrogativa foi considerada aqui.

A escolha desta variável linguística foi motivada, assim como a classe de palavra originária dos RADs, por uma característica apontada por diferentes autores, já citados neste estudo, como sendo marcante dos RADs, trata-se do caráter interrogativo dos RADs. No entanto, ao selecionarmos os dados a serem analisados neste estudo, notamos que a entonação interrogativa parece se dá de diferentes formas, ou seja, o RAD parece aparecer tanto em sentenças que já possuem um caráter interrogativo, quanto em sentenças que não tem este caráter interrogativo, sendo o RAD o único elemento que marca esta interrogação.

Sendo assim, selecionamos a entonação interrogativa como podendo ser realizada só no RAD ou em toda a sentença em que o RAD foi realizado. Partindo desta concepção, o quadro acima mostra que, no total, os RADs parecem ser utilizados mais em sentenças em que a entonação interrogativa se dá apenas nele, 92,2% e não em toda a sentença, 7,8%. Tal percentual de ocorrência neste *corpus* analisado pode ser significativo para discussões posteriores acerca de concepções anteriores acerca dos RADs, pois estudos precedentes apontam que os RADs são recorrentes em contextos interrogativos e este dado pode provocar uma reflexão em torno desta percepção, pois pode ser que o RAD não seja recorrente em

contextos interrogativos, mas mesmo em sentenças não interrogativas, ele apresente esta marca de indagação.

Se, de maneira geral, os RADs tiveram um percentual de uso com entonação interrogativa só marcador, ao analisarmos este percentual a partir da variável dependente, notamos que as formas não descritas apresentam um percentual de uso maior em sentenças interrogativas, 23,1%, ou seja, o RAD é utilizado com mais frequência em sentenças que apresentam esta entonação interrogativa, diferentemente das formas já descritas, em que os RADs parecem ser utilizados com mais frequência em sentenças que a entonação interrogativa se dá apenas nele, 90,3%.

Tal dado nos parece importante tendo em vista que como o caráter interrogativo do RAD tem sido apresentado como uma característica destes elementos, algo que parece ser confirmado com o percentual de 90,3% no uso das formas já descritas, as formas não descritas parecem apontar para outras características dos RADs, ou seja, eles parecem ser realizados, de maneira bem recorrente, inclusive, em sentenças cuja entonação interrogativa se dá em toda a sentença e não apenas por intermédio do RAD.

Após todas estas análises, destacamos que os números percentuais apresentados confirmam ou refutam algumas das hipóteses levantadas na presente pesquisa. No entanto, antes de apresentarmos tais assertivas, o programa computacional aponta, junto com os pesos relativos, as rodadas consideradas mais e menos significativa para o estudo que está sendo realizado. Diante desta constatação, apresentamos na seção seguinte a *best stepping uprun* (melhor rodada).

5.3.1 Análises da rodada mais significativa¹²⁸

O programa selecionou a rodada de número 23 como sendo a mais significativa para o estudo. Nesta rodada, o programa traz o peso relativo das variáveis mais expressivas que fizeram com que esta rodada fosse, para o GoldVarb X, assim considerada.

¹²⁸ O programa apontou como *best stepping down run* (rodada menos significativa) a rodada de número 32. Nesta rodada, foram citados todos os grupos selecionados como variáveis independentes. No entanto, as análises apresentadas em seções anteriores contemplam cada uma destas variáveis. Sendo assim, optamos por não reapresentá-las, uma vez que fazer isso, seria apresentar o mesmo dado, agora em peso relativo. No que tange aos grupos citados *nabest stepping uprun* (melhor rodada), optamos por analisá-los individualmente tendo em vista que alguns pesos relativos diferem dos números em percentuais e, como a melhor rodada, apresenta o grupo de variáveis que são mais favoráveis para a variação, julgamos pertinente analisá-la nesta seção.

Esta rodada selecionou, a partir da variável dependente, os seguintes grupos: 1 (gênero), 3 (localidade), 5 (classe de palavra originária do RAD) e 6 (massa fônica) como sendo grupos em que os fatores são significativos para a variação. Iremos apresentar o peso relativo de cada grupo abaixo, elucidando, inclusive, diferenças entre o peso relativo dado por esta rodada do programa e o percentual de uso apontado pelo programa nas rodadas iniciais descritas e analisadas na seção anterior.

No grupo 1, gênero, o peso relativo dado ao valor de aplicação da variável dependente foi de .22 para o gênero masculino e de .76 para o gênero feminino, como estes pesos são calculados a partir de uma das variáveis dependentes, no nosso caso aqui a ausência de RAD já descrito na Língua Portuguesa, ou seja, o número restante para completar 1 (um) é referente a outra variável, a saber, presença de RAD já descrito na Língua Portuguesa. Para facilitar a compreensão, expomos estes dados no quadro abaixo para prosseguirmos com as análises.

Quadro 26: Peso relativo dado para o grupo 1 (Gênero)

Gênero ↓	Ausência de formas já descritas	Presença de formas não descritas
Masculino	.22	.78
Feminino	.76	.24

Os números apresentados no quadro apontam para o seguinte dado: as formas não descritas em estudos anteriores (ausentes) parecem ter um peso relativo de uso não muito significativo, tendo em vista que para as formas já descritas o peso foi de 0,78 para o gênero masculino. Já para o gênero feminino, o peso relativo dado pelo programa foi de 0,76 para as formas não descritas (ausentes), ou seja, as formas já descritas tiveram um peso relativo pequeno nas ocorrências, 0,24 evidenciando assim, que a variável social gênero apresenta, em peso relativo, uma diferença apontada pelo programa como significativa, algo que, por meio das porcentagens, embora tenhamos notado a sutil diferença nas ocorrências não tenhamos considerado como sendo significativa.

Tal dado, evidencia, de alguma forma, a importância de codificar, quantificar e rodar os dados coletados em programas computacionais como o GoldVarb X, pois o programa, ao atribuir os pesos relativos das variáveis identifica a rodada mais significativa e, como se deu aqui, pode apresentar variáveis que, em termos percentuais não tinha se apresentado como tão significativa assim.

O segundo grupo selecionado na *best stepping up run* foi o grupo 3 (localidade), faremos de maneira semelhante ao que fizemos para expor os resultados do grupo 1 porque entendemos que facilita a compreensão e exposição dos pesos relativos dados pelo programa.

Quadro 27: Peso relativo dado para o grupo 3 (Localidade)

Localidade ↓	Ausência de formas já descritas	Presença de formas não descritas
São José do Egito	.87	.13
Triunfo	.33	.67
Serra Talhada	.17	.83

Conforme podemos observar no quadro, as variantes não descritas (ausentes) tiveram um peso relativo maior de uso em São José do Egito, algo que nos mostra como não sendo recorrente nos outros dois municípios, pois o peso relativo maior em Triunfo e Serra Talhada foram dados para as formas já descritas (presentes). Tal dado nos apresenta como sendo significativo, pois embora tenhamos identificado no princípio destas análises que em Triunfo o número de ocorrências de RADs tenha sido menor se comparado aos demais municípios, não consideramos tal dado como sendo significativo, uma vez que estamos estudando os municípios como uma única comunidade de fala e o número de ocorrências em cada cidade identificado neste *corpus* não implica dizer que determinado município usa mais ou menos os RADs, apenas que durante a coleta de dados realizada para este estudo os RADs foram utilizados neste município com um determinado total de ocorrências.

Contudo, ao selecionar a localidade como uma variável social e o programa rodar estes dados a partir da variável dependente e ainda selecionar este grupo como pertencente a rodada mais significativa, notamos que o peso relativo de cada município muda quando analisado a partir da presença/ausência de formas já descritas em Língua Portuguesa.

Sendo assim, podemos perceber que em peso relativo, os municípios de Triunfo e Serra Talhada parecem usar mais as formas já descritas de RADs, enquanto que São José do Egito parece usar mais as variantes ainda não catalogadas dos RADs. A partir destas considerações, podemos perceber que podem haver municípios, neste caso, São José do Egito, em que o uso das formas ainda não descritas seja mais acentuado, essa percepção é necessária não apenas para a descrição das variantes que são utilizadas na comunidade de fala, mas também para servir como base para estudos posteriores que visem dar continuidade ao

processo de descrição das variantes dos RADs , uma vez que os dados destes municípios podem ser comparados com outros pertencentes a mesma ou a outra comunidade de fala.

O grupo 5 também foi apresentado na rodada mais significativa apontada pelo programa. Neste grupo temos um peso relativo de .46 para a variável classe de palavra que antecede o RAD em formas não descritas (ausentes) e .54 para esta mesma variável com formas já descritas (presentes). Apesar da indicação como relevante, entendemos que os verbos, fator selecionado nas rodadas, antecede as variantes já catalogadas e não catalogadas de maneira bem equiparada uma vez que o peso relativo foi de .46 e .54, respectivamente.

O último grupo citado na rodada mais relevante pelo programa foi o grupo 6, massa fônica. Neste caso, os pesos relativos apresentados convergem para o que já apresentamos na seção anterior, ou seja, as formas não descritas (ausentes) apresentaram um peso relativo de .32 para as formas simples, já as formas compostas tiveram um peso relativo de .68 para as variantes já descritas (presentes) e algo semelhante foi apontado quando apresentamos os números em termos percentuais, pois destacamos que as formas não descritas tiveram um percentual de uso maior com formas compostas.

Com base nos números e análises apresentadas, expomos as considerações finais do presente estudo, discorrendo acerca das hipóteses confirmadas e não confirmadas, além disso, destacamos passos futuros a serem dados em direção a descrição e análise linguística não apenas dos RADs, mas dos MCs da Língua Portuguesa.

6 CONSIDERAÇÕES FINAIS

O presente estudo buscou descrever e analisar os Requisitos de Apoio Discursivo – RADs – utilizados por pessoas do Sertão do Pajeú – Pernambuco. Para tanto, ancorados nos pressupostos teórico-metodológicos da Sociolinguística Variacionista, coletamos dados de fala de pessoas de três diferentes municípios desta região que, de acordo com o que consideramos, podem representar a comunidade de fala analisada. A partir dos dados obtidos, a pesquisa aponta alguns resultados acerca do uso dos RADs, assim como, aponta para estudos posteriores a serem realizados com estes elementos linguísticos da Língua Portuguesa.

A realização deste estudo foi motivada, a partir do estudo inicial de Alcântara (2015), pela seguinte hipótese: é possível que existam formas que atuam como RADs na fala dos sertanejos que diferem das formas de RADs já descritas em outras pesquisas realizadas em outras comunidades de fala. Sendo assim, iniciamos o estudo procurando observar o uso dos RADs e quais as variantes desta variável linguística na fala dos informantes.

Após a análise dos dados coletados, identificamos onze formas no *corpus*, a saber: *né?*, *sabe?*, *viu?*, *tá ligado?*, *entendeu?*, *num sabe?*, *num é?*, *hein?*, *tá?*, *visse?* e *certo?*. Dentre estas variantes, identificamos, com base em estudos anteriores, que algumas destas formas parecem ser recorrentes em outras comunidades de fala, pois algumas pesquisas apontam estas variantes nas comunidades de fala analisada, são elas: *né?*, *tá? sabe?*, *viu?* e *entendeu?*. Dentre estas formas, destacamos o alto índice de ocorrência da variante *né?*, 62%. Tal dado parece confirmar que esta é a variante de RAD mais utilizada na Língua Portuguesa, pois estudos anteriores, como Marcuschi (1989), Macedo e Silva (1996) e Freitag (2008), destacam o alto índice de realizações no *né?* nos *corpus* estudados.

Além disso, notamos que algumas formas como *ok?* e *tá certo?*, apresentadas por Alcântara (2015) como presente em uma comunidade de prática do Sertão do Pajeú, não tiveram ocorrências no comunidade fala do Sertão do Pajeú apontando, assim, para uma possibilidade dos RADs serem elementos que variam não apenas de uma comunidade de fala para outra, mas nas comunidades de prática, pois podem existir formas que sejam específicas de seus falantes, algo que só estudos futuros com outras comunidades de prática poderão afirmar.

No que tange as variantes *visse?*, *tá ligado?*, *num é?*, *num sabe?*, *certo?* e *hein?*, consideradas aqui como formas ainda não descritas dos RADs, os resultados alcançados

confirmam nossa hipótese inicial, ou seja, há formas que estão atuando como RADs no Sertão do Pajeú que são distintas de outras formas já descritas em outras comunidades de fala e descrevê-las e analisá-las é, certamente uma importante contribuição que os estudos linguísticos pode oferecer, pois são elementos linguísticos presentes e recorrentes na língua e, por considerarmos a língua de um povo como integrante de sua identidade, estudar e reconhecer os itens que integram esta língua é dar ênfase a algo que compõe a identidade de um povo, além de ser um registro de uma língua falada pela comunidade de fala.

A despeito destas considerações, as análises realizadas mostram algo que não tivemos como hipótese inicial, mas reconhecemos que pode ser um importante dado para estudos posteriores, sobretudo, pesquisas dentro da perspectiva diacrônica de estudo da língua, trata-se de algumas formas terem o que chamamos aqui de relação com algumas formas já descritas em outros estudos. Esta relação pode ser entendida como um resquício de uma variante para outra, pois temos formas como *viu?*, muito recorrente na Língua Portuguesa, e formas como *visse?*, que, conforme Irineu e Costa (2015) são recorrentes na fala do recifense. Contudo, apresentamos este dado apenas como uma especulação levantada a partir dos dados deste estudo e entendemos que só estudos diacrônicos podem confirmar ou refutar tal consideração.

As ocorrências das variantes de RADs não descritas anteriormente tiveram ainda outra característica que entendemos como sendo significativa, conforme destacamos no capítulo de análise dos dados, trata-se de algumas variantes terem realizações apenas em alguns dos municípios integrantes da comunidade de fala analisada, é caso de *hein?* que teve ocorrências apenas em Serra Talhada, de *visse?* com ocorrências apenas em São José do Egito e *certo?* em Triunfo. Parece que estas formas não se apresentam como estranhas, por exemplo, aos indivíduos que não pertencem ao referido município, mas ao termos realizações que se deram em apenas um dos municípios, entendemos que o informante ao selecionar esta e não outra forma de RADs pode fazer isto motivado por alguns fatores, como semânticos, por exemplo. Contudo, é necessário que pesquisas posteriores, que abordem esta perspectiva semântica da língua busquem analisar esta possibilidade que se apresenta, em nossa opinião, como algo importante para os estudos dos RADs.

Com base nestas assertivas, notamos que descrever as variantes dos RADs presentes em uma determinada comunidade de fala, além de ser uma contribuição para estudos linguísticos, é um processo de descobertas nunca antes hipotetizadas, pois, conforme discorremos, a hipótese que tínhamos inicialmente era de que poderiam existir formas distintas na fala dos sertanejos e, após a análises realizadas, não apenas confirmamos nossa

hipótese, mas fomos levados pelos dados alcançados a novas hipóteses acerca do uso dos RADs, como, por exemplo, a possibilidade de uma relação entre uma forma e outra, estas, por sua vez, só poderão ser confirmadas em pesquisas futuras.

Após a descrição das variantes encontradas, as análises seguiram com o intuito de confirmarmos ou não as demais hipóteses levantadas neste estudo. Para tanto, por se tratar de hipóteses relacionadas a variáveis linguísticas e sociais, recorreremos aos números, em termos percentuais, das ocorrências de cada variante estratificadas por cada uma das variáveis selecionadas.

As variáveis sociais selecionadas foram: localidade, gênero, faixa-etária e escolaridade. Para a variável localidade, tínhamos como hipótese que poderiam existir formas distintas que atuam como RAD nos diferentes municípios que integram a comunidade de fala em análise. Esta hipótese foi confirmada, pois as formas *hein?*, *visse?* e *certo?* forma realizadas apenas nos municípios de Serra Talhada, São José do Egito e Triunfo, respectivamente. No entanto, o percentual de 2,05% de uso da variante *certo?* em Triunfo, por exemplo, nos chamou atenção, pois se compararmos com formas já descritas, como é caso de *viu?*, com 2,73, são bem próximos, tal dado nos faz acreditar que parece ser que não apenas há formas ainda não descritas em cada localidade, como também, estas formas podem ser tão presentes quanto as formas já descritas e analisadas em outras comunidades.

No que diz respeito à variável gênero, nossa hipótese era de que as mulheres usariam mais os RADs. Esta hipótese foi levantada com base em estudos citados nesta dissertação que afirmam, com base em algumas considerações, que as mulheres usam mais os RADs. No entanto, a proximidade no percentual de uso, na comunidade de fala, estratificadas por município e a partir das variantes descritas e não descritas na língua portuguesa, entendemos que não mostrou como sendo significativa, conforme discutimos no capítulo de análise. No entanto, entendemos que isso pode se dá em virtude da seguinte consideração: os estudos que apontam esta variável, gênero, como significativa para a variação dos RADs, são estudos iniciais e que foram realizados tempos atrás. Na atualidade isto pode ter sofrido alterações, ou seja, o gênero não influenciam no uso dos RADs, contudo, só uma sequência de estudos com os RADs que analisem esta variável poderá concluir se o gênero, nos corpus atuais, não condiciona o uso dos RADs.

A hipótese levantada acerca da faixa-etária, ou seja, os mais jovens usariam mais os RADs, foi refutada, pois, após as análises dos dados coletados, observamos que, na comunidade de fala em estudo, os informantes mais jovens não utilizam mais os RADs, pelo

contrário, conforme apresentamos, os mais jovens parecem usar menos os RADs. No entanto, isso pode se dá em virtude da própria coleta de dados, pois para o jovem interagir com o pesquisador pode ser que não seja tão interativo para ele quanto interagir com um amigo ou alguém que já conheça, uma vez que o uso dos RADs está relacionado a interação entre as pessoas.

Com relação a escolaridade, tínhamos como hipótese que as pessoas com maior nível de escolaridade usariam menos os RADs, tendo em vista que eles são elementos estigmatizados pela Gramática Tradicional. No entanto, esta variável não mostrou condicionar o uso dos RADs, pois os resultados em termos percentuais foram bastante equiparados. O uso dos RADs por pessoas de diferentes níveis de escolaridade pode apontar para algo que Freitag (2007) considera: os marcadores não constituem-se vícios de linguagem, mas, acreditamos nós, em uma necessidade em momentos de interação que perpassa a fala de pessoas com diferentes níveis de escolaridade.

Além das variáveis sociais, levantamos hipóteses para as variáveis linguísticas selecionadas: massa fônica, classe de palavra que origina e antecede o RAD e entonação interrogativa. Após a realização do estudos, consideramos que estas variáveis linguísticas podem ser significativas no que diz respeito ao uso dos RADs.

Para a variável massa fônica tínhamos como hipótese que as formas não descritas poderiam ser mais recorrentes com formas compostas e esta hipótese foi confirmada, pois as formas já descritas tiveram um percentual de 96,1% com formas simples Já as formas compostas, foram mais recorrentes nas variantes não descritas, 94,1%. Tal dado nos levou a considerar a possibilidade de relação que já mencionamos anteriormente, que, segundo discutimos, só estudos diacrônicos podem confirmar.

Para a classe de palavra que origina o RAD, por estarmos lidando com formas descritas e não descritas em estudos com RADs, tínhamos como hipótese que as formas não descritas poderiam ser originadas de classes de palavras distintas das dos RADs já descritos. Esta hipótese não foi confirmada, pois as formas novas parecem emergir das mesmas classe de palavras dos RADs já catalogados, a saber, verbos, adjetivos e advérbios.

Com relação à variável linguística classe de palavras que antecede o RAD, em virtude da posição sintática do RAD, hipotetizamos que algumas classes de palavras podem favorecer o uso dos RADs. Após a análises, observamos que das classes de palavras selecionadas, substantivos, adjetivos, verbos e pronomes, o substantivo parece ser a classe que mais antecede o RAD, ou seja, a hipótese confirmada, consideramos que isso acontece, como dito,

pela estrutura sintática da língua. Contudo, os dados analisados mostram que as variantes não descritas parecem ser mais antecidas por pronomes e verbos, diferenciando-se, assim, das variantes já catalogadas, pois estas as classes de palavras que antecidem o RAD são bem equiparadas em termos percentuais. Consideramos que isso pode se dá em virtude do uso recorrente dos RADs já descritos, pois este uso frequente pode fazer com que o informante consiga usá-los em diferentes tipos de sentenças e, conseqüentemente, sentenças que sejam concluídas com palavras de diferentes classes.

A entonação interrogativa teve como hipótese que o RAD pode ser utilizado em sentenças que não tenham a entonação interrogativa, mas ele é o elemento linguístico que traz esta entonação interrogativa. Após as análises realizadas, confirmamos esta hipótese, pois 92,2% do uso dos RADs neste *corpus* foi em sentenças que não possuem uma entonação interrogativa, mas o RAD é quem traz esta entonação após uma sentença afirmativa, por exemplo. O RAD teve um percentual de uso de 7,8% em sentenças completamente interrogativas e este dado, assim como, o que mencionamos anteriormente acerca do uso em sentenças sem entonação interrogativa, confirma a hipótese do presente estudo para esta variável. Este dado pode ser significativo para a definição dos RADs em língua Portuguesa, pois, como vimos, são elementos que podem trazer uma entonação interrogativa em sentenças que não tem esta característica de perguntar, mas declarar algo, apenas. Sendo que o RAD traz esta marca interrogativa pedindo o assentimento do interlocutor acerca do que foi afirmado.

Por fim, as rodas no GoldVarb X, selecionaram, de todas as variáveis aqui mencionadas, a rodada em que as variáveis gênero, localidade, classe de palavra originária dos RADs e massa fônica foram analisadas. É importante destacar que o programa selecionou estas variáveis a partir da variável dependente presença/ausência dos RADs já descritos em Língua Portuguesa. Tais variáveis já foram mencionadas e apresentadas, nesta seção, a partir das hipótese que foram ou não confirmadas, mas a indicação do programa computacional como sendo a mais significativa nos leva a considerar que, em termos de variação no uso do RADs, este estudo sociolinguístico considera que o gênero, a localidade, a classe de palavras originária do RAD e a massa fônica podem ser significativas no estudos das variantes dos RADs, que sejam elas formas já descritas ou não na Língua Portuguesa.

Com base nas considerações apresentadas aqui e tomando como base a metáfora da floresta virgem proposta por Macedo e Silva (1996), tomamos a decisão de concluir este estudo de maneira diferente ao que Alcântara (2015) fez ao afirmar que o referido estudo

poderia ser considerado um passo a frente dentro da floresta dos marcadores. Em nosso caso aqui, concluímos o estudo com o sentimento e a necessidade de fazer o que as autoras também sugerem na metáfora: “voltar sobre os passos dados” (MACEDO E SILVA, 1996, p.16), pois no que observamos aqui, há muito ainda a ser descoberto nesta e em outras comunidades de fala no que diz respeito ao uso dos Requisitos de Apoio Discursivo da Língua Portuguesa, inclusive, ‘voltar sobre os passos’ utilizando outros recursos, pois nos instiga a possibilidade de analisar estes e/ ou outras RADs em diferentes perspectivas linguísticas, como sugerimos em vários trechos desta dissertação.

REFERÊNCIAS

- ALCÂNTARA, M. C. N. S. *O uso dos Requisitos de Apoio Discursivo por radialistas do Sertão do Pajeú em dois níveis de formalidades diferentes: um estudo sociolinguístico*. Monografia (Licenciatura Plena em Letras) – Curso de Letras, Universidade Federal Rural de Pernambuco – Unidade Acadêmica de Serra Talhada, Serra Talhada – Pernambuco, 2015.
- BEINHAUER, W. *El español coloquial*. Madrid: Gredos, 1964.
- BELINE, R. A variação linguística. In: FIORIN, J. L. (org). *Introdução à Linguística*. 4. Ed. São Paulo: Contexto, 2006.
- BRUSTOLIN, A. K. B. da S. *et al* O preenchimento do sujeito pronominal em textos escritos de alunos adolescentes de Florianópolis. *Working papers in linguistics*, n.esp. Florianópolis, 2009. p. 94-107.
- CARDOSO, S. A. M. da S. et al. *Atlas Linguístico do Brasil*. Londrina: Eduel, v. 1. p. 212, 2014.
- CARDOSO, S. A. M. da S. et al. *Atlas Linguístico do Brasil*. Londrina: Eduel, v. 2. p. 368, 2014.
- CASTILHO, A. T. de. *A língua falada no ensino de português*. 6. Ed. – São Paulo: Contexto, 2004.
- CASTILHO, A. T. de. *Nova gramática do português brasileiro*. 1. ed., 1. reimpressão. – São Paulo: Contexto, 2010.
- CASTILHO, A. T. de. *Nova gramática do português brasileiro*. 1. ed., 4ª. reimpressão. – São Paulo: Contexto, 2016.
- COELHO, I. L. *et al Sociolinguística*. Florianópolis: LLV/CCE/UFSC, 2010.
- DE PAULA, A. S. O trabalho de campo sociolinguístico. In: COSTA, J. F. da.; SANTOS, R. L. de A.; VITÓRIO, E. G. de S. L. A. *Variação e mudança linguística no estado de alagoas*. Maceió: EDUFAL, 2011.
- FREITAG, Raquel MeisterKo. Marcadores discursivos interacionais: análise contrastiva entre duas variedades do português falado no Brasil. In: *Anais do VIII Encontro do CELSUL*. Pelotas: Educat, 2008.
- FREITAG, R. M. K.; *Marcadores Discursivos não são vícios de linguagem*. Interdisciplinar: revista de estudos em Língua e Literatura. São Cristóvão – Sergipe, v. 4. n. 4. p. 22-43 2007.
- GÜLICH, E. *Makrosyntax der Gliederungssignale im gesprochenen Französisch*. München: W. Fink, 1970.

GIVÓN, T. *English Grammar: a function-based introduction*. Amsterdam/Philadelphia: J. Benjamins, 1993.

GIVÓN, T. *Functionalism and Grammar*. Amsterdam/Philadelphia: J Benjamins, 1995.

GUY, G. R. *A identidade linguística da Comunidade de fala: paralelismo Interdialetal nos padrões de Variação linguística*. Organon, Revista do Instituto de Letras da UFRGS, Porto Alegre, v. 28 e 29. p. 17-32, 2001.

GUY, G. R.; ZILLES, A. *Sociolinguística quantitativa: instrumental de análise*. São Paulo: Parábola Editorial, 2007.

HOLTUS, G.; RADTKE, E. *Gesprochenes Italienisch in Geschichte und Gegenwart*. Tübingen: Günter Narr, 1985.

IRINEU, L. M.; COSTA, W. P. A. Marcadores Discursivos no gênero comentário de blog futebolístico: constatações sobre o falar pernambucano. In. SEDRINS, A. P.; SÁ, E. J. de. (orgs.). *Aspectos descritivos e sócio-histórico da língua falada em Pernambuco*. Recife: EDUFRPE/FACEPE, 2015.

KELLER, E. Gambits: conversational strategy signals. *Journal of pragmatics* 5, 1979: 219-238.

LABOV, W. Where does the linguistic variable stop? A response to Beatriz Lavandera. In: *Sociolinguistics Working Paper*. Texas, n. 44, 1978. p. 1-13.

LABOV, W. *Padrões sociolinguísticos*. São Paulo: Parábola Editorial, 2008 [1972].

LABOV, W. *Principles of linguistic Change: Internal Factors*. Vol. 1. Blackwell publishing, 1994.

LABOV, W. *Principles of linguistic Change: Social Factors*. Vol. 2. Blackwell publishing, 2001.

MACEDO, A. T.; CARMO, C. S. Marcadores nas respostas. Anais do III Congresso da Assel-RIO. Niterói, Instituto de Letras da UFF, 1994.

MACEDO, A. T. Aquisição de marcadores em primeira e em segunda língua. In: RONCARATI, C.; MOLLICA, M. C. *Variação e aquisição*. Rio de Janeiro: Tempo Brasileiro, 1997.

MACEDO, A. T.; SILVA, G. M. de O. e. Análise sociolinguística de alguns marcadores conversacionais. In: MACEDO, A. T.; RONCARATI, C.; MOLLICA, M. C. *Variação e Discurso*. Rio de Janeiro: Tempo Brasileiro, 1996.

MARCUSCHI, L. A. Marcadores conversacionais do português brasileiro: formas, posições e funções. In: CASTILHO, A. T. de. *Português culto falado no Brasil*. Campinas: Editora da UNICAMP, 1989.

MARCUSCHI, L. A. *Análise da Conversação*. São Paulo: Ática: 1986.

MOLLICA, M. C. Fundamentação teórica: conceituação e delimitação. In: MOLLICA, M. C.; BRAGA, M. L. *Introdução à sociolinguística: o tratamento da variação*. São Paulo: Contexto, 2003. p. 9-14.

MOURA NEVES, M. H. de. *Gramática funcional*. São Paulo: Martins Fontes, 1997.

MEYER-HERMANN, R. *Studienzur Funktion von Metakommunikation: am beispielgesprochenerportugiesischer und französischer Sprache*. Habilschrift: M. S. Bielefeld, 3° vol. 1979.

MEYER-HERMANN, Formen und Funktionen von “Abschwächung” imgesprochenenPortugiesische. In: SCHMIDT-RADFELDT, J. *Portugiesische Sprachwissenschaft*. Tübingen: Günter Narr, 1983, 21-55.

MURPHY, R. *English Grammar in Use*. São Paulo: Cambridge University Press, 2004.

NARO, A. J. Modelos quantitativos e tratamento estatístico. In: MOLLICA, M. C.; BRAGA, M. L. *Introdução à sociolinguística: o tratamento da variação*. – São Paulo: Contexto, 2003. p. 15-25.

PAIVA, M. da C. de. Transcrição dos dados linguísticos. In: MOLLICA, M. C.; BRAGA, M. L. *Introdução à sociolinguística: o tratamento da variação*. – São Paulo: Contexto, 2003. p. 135-146.

PEIXOTO, S. P. L. A hipo e hipersegmentação reveladas nos textos de aquisição da escrita: um estudo sociolinguístico. In: COSTA, J. F. da.; SANTOS, R. L. de A.; VITÓRIO, E. G. de S. L. A. *Variação e mudança linguística no estado de alagoas*. – Maceió: EDUFAL, 2011.

PEREIRA, D. K. F. *A realização do artigo definido no português falado na região do Sertão do Pajeú – Pernambuco*. Dissertação (Mestrado em Linguística) – Programa de Pós-graduação em Letras, Universidade Federal de Pernambuco. Recife – Pernambuco, 2017.

REHBEIN, J. Sprechhandlungsaugmente: zurOrganisations der Hörersteuerung.In: WEYDT, H., Hsg. *Die Partikeln der desutschen Sprache*. Berlin: de Gruyter, 58-74, 1979.

RISSO, S. M.; SILVA, G. M. O.; URBANO, H. Traços definidores dos Marcadores Discursivos. In. JUBRAN, C. S. (org.) *A Construção do texto falado*. São Paulo: Contexto, 2015.

RISSO, S. M.; SILVA, G. M. O.; URBANO, H. Marcadores Discursivos: Traços definidores. In. KOCH, I. V. (org.) *Gramática do Português Falado*. Vol. VI. Campinas – São Paulo: UNICAMP/FAPESP, 1996.

RISSO, S. M.; SILVA, G. M. O.; URBANO, H. *Marcadores Discursivos: Traços definidores*. Projeto Gramática do Português Falado. Campos do Jordão: Mimeo, 1994.

SANTOS, R. L. A. A metodologia da pesquisa em sociolinguística variacionista. *Revista Espaço Acadêmico*, nº 97, 2009. p. 68-71.

SANTOS, R. L. A. *A concordância verbal na fala de menores carentes que vivem em entidades filantrópicas de Maceió*. Dissertação (Mestrado em Linguística) – Faculdade de Letras, Universidade Federal de Alagoas, Maceió, 2010.

SANTOS, R. L. de. A.; VITÓRIO, E. G. de. S. L. A. Teoria da variação e Mudança Linguística. In: COSTA, J. F. da.; SANTOS, R. L. de A.; VITÓRIO, E. G. de S. L. A. *Variação e mudança linguística no estado de alagoas*. – Maceió: EDUFAL, 2011.

SAUSSURE, F. de. *Curso de Linguística Geral*, 27. ed., São Paulo: Cultrix, 2006.

SCHERRE, M. M. P.; NARO, A. J. Análise quantitativa e tópicos de interpretação do Varbrul. In: MOLLICA, M. C.; BRAGA, M. L. *Introdução à sociolinguística: o tratamento da variação*. – São Paulo: Contexto, 2003. p. 15-25.

SCHIFFRIN, D. *Discourse markers*. Cambridge: Cambridge University Press, 1987.

SCHIFFRIN, D. Discovering the context of an utterance. *Linguistics*, 25, 11-32, 1987.

SEDRINS, A. P.; SIQUEIRA, A. L. S.; ARAÚJO, R. L. de. A variação na concordância nominal de número na língua falada no Sertão Pernambucano. In: SEDRINS, A. P.; SÁ, E. J. de. (orgs.). *Aspectos descritivos e sócio-histórico da língua falada em Pernambuco*. Recife: EDUFRPE/FACEPE, 2015.

SEDRINS, A. P.; PEREIRA, D. K. F.; SIQUEIRA, A. L. S.; Variação na realização do artigo definido diante de antropônimos em dados de fala e escrita no sertão de Pernambuco. In: SEDRINS, A. P.; SÁ, E. J. de. (orgs.). *Aspectos descritivos e sócio-histórico da língua falada em Pernambuco*. Recife: EDUFRPE/FACEPE, 2015.

SEVERO, C. G. O lugar do indivíduo na teoria laboviana. *Revista língua(gem)*. Macapá, v. 1. n. 2. 2004.

SILVA, G. M. O.; ASSAFIN, S. A incidência de vocábulos pouco frequentes após hesitação e partícula ‘assim’. *Anais do IV Congresso da ASSEL-Rio*. Niterói, Instituto de Letras da UFF, 1994.

SILVA, V. L. P. da. Relevância das variáveis linguísticas. In: MOLLICA, M. C.; BRAGA, M. L. *Introdução à sociolinguística: o tratamento da variação*. – São Paulo: Contexto, 2003. p. 67-71.

TARALLO, F. *A pesquisa sociolinguística*. São Paulo: Ática, 2007.

VALLE, C. R. M. *Sabe? Não tem? Entende?*: itens de origem verbal em variação como requisitos de apoio discursivo. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Linguística, Universidade Federal de Santa Catarina, Florianópolis, 2001.

VANIN, A. A. Considerações relevantes sobre definições de ‘comunidade de fala’, 2009. Disponível em <http://eduem.uem.br/ojs/index.php/ActaSciLangCult/article/viewFile/6367/6367>. Acesso em 15 de dezembro de 2017.

VERSYCLE, N. I. et al. Microrregião Pajeú: economia, clima e desenvolvimento da agricultura através de modelo digital do terreno. Revista GEAMA, Recife, v.1 n. 1. 2015.

VITÓRIO, E. *Ter/haver existenciais na escrita de alunos dos ensinos fundamental e médio da cidade de Maceió/AL*. Dissertação (Mestrado em Linguística) – Programa de Pós-Graduação em Letras e Linguística, Universidade Federal de Alagoas, Maceió, 2008.

WEIREICH, U.; LABOV. W. &HERZOG, M. I. *Fundamentos empíricos para uma teoria da mudança linguística*. São Paulo: Parábola. 2006 [1968].

APÊNDICE A - Roteiro de entrevista



Universidade Federal de Pernambuco (UFPE)
Centro de Artes e Comunicação (CAC)
Programa de Pós-Graduação em Letras (PPGL)

Orientador: Prof. Dr Marcelo Amorim Sibaldo
Orientando: Meiriany Cristinaide Nascimento Souza Alcântara

ROTEIRO DE ENTREVISTA

- 1) Você é desta cidade? Já morou fora daqui?
- 2) Como é viver aqui? Você gosta de morar aqui?
- 3) Fale sobre a sua cidade. Os problemas, as qualidades.
- 4) O que poderia existir aqui para ser um lugar melhor?
- 5) Você estudou aqui? Em qual escola?
- 6) Quais suas matérias preferidas?
- 7) Fale sobre os seus professores
- 8) Quem você destacaria como o melhor professor, ou seja, aquele que poderia ser um exemplo para você?
- 9) Você tem amigos? Fale o que acha sobre amizade.
- 10) Como é sua relação com sua família?
- 11) Qual a importância que sua família tem pra você?
- 12) Você já teve algum problema familiar? Fale sobre ele.
- 13) Qual foi o momento de sua vida em que sua família foi importante?
- 14) Alguém de sua família mora em outro estado? Em qual estado?

(Caso a resposta da questão 14 tenha sido positiva, faça as perguntas de número 15, 16 e 17).

- 15) Você tem contato com ela?
- 16) Costuma viajar para visitar esse familiar?
- 17) Quanto tempo demora sua visita?
- 18) Você tem algum ídolo? Qual?
- 19) Você costuma ouvir música?
- 20) Quais seus cantores e cantoras preferidos?
- 21) Costuma assistir filmes ou novelas?
- 22) Quais atores/atrizes você destacaria como sendo o melhor, em sua opinião?
- 23) Qual seu filme preferido? Lembra-se do nome dos personagens?
- 24) Qual sua novela preferida? Lembra-se do nome dos personagens?
- 25) Gosta de algum programa de TV? Qual? Quem é o apresentador (a)?
- 26) Gosta de alguma série de TV? Qual? Lembra-se do nome dos personagens? Conte um pouco sobre os personagens. Identifica-se com algum? Qual?
- 27) Qual a pessoa mais importante na sua vida e por quê?
- 28) Se você pudesse mudar de vida com alguém, quem seria essa pessoa e por quê?
- 29) Como seria o dia perfeito em sua vida?
- 30) Quem você citaria como uma pessoa exemplo para sua vida?
- 31) Qual sua opinião acerca do país? O que realmente precisa ser feito para ser um país melhor?

APÊNDICE B - Ficha social do informante



Universidade Federal de Pernambuco (UFPE)
Centro de Artes e Comunicação (CAC)
Programa de Pós-Graduação em Letras (PPGL)
Orientador: Prof. Dr Marcelo Amorim Sibaldo
Orientando: Meiriany Cristinaide Nascimento Souza Alcântara

Ficha Social do informante		
DADOS PESSOAIS DO INFORMANTE		
1. NOME:		2. APELIDO:
3. DATA DE NASCIMENTO:	4. IDADE:	5. SEXO/GÊNERO:
6. ENDEREÇO:		
7. ESTADO CIVIL:		8. NÚMERO DE FILHOS: HOMENS: MULHERES:
9. NATURALIDADE:	10. COM QUE IDADE CHEGOU A ESTA CIDADE? (CASO NÃO SEJA NATURAL DA LOCALIDADE)	
11. DOMICÍLIOS E TEMPO DE PERMANÊNCIA FORA DA LOCALIDADE:		
12. ESCOLARIDADE:		13. OUTROS CURSOS:
14. COSTUMA VIAJAR PARA OUTROS MUNICÍPIOS:		15. COSTUMA VIAJAR PARA OUTROS ESTADOS:
DADOS PROFISSIONAIS DO INFORMANTE		
16. QUAL A PROFISSÃO QUE EXERCE?		
17. QUAIS PROFISSÕES JÁ EXERCEU?		
18. POR QUANTO TEMPO EXERCE A ATUAL PROFISSÃO?		
19. FEZ CURSOS ESPECÍFICOS PARA A SUA PROFISSÃO?		

RENDA		
20. FONTES DE RENDA	21. VALOR DA RENDA () Entre 1 e 3 salários mínimos () Entre 4 e 6 salários mínimos () Mais de 7 salários mínimos	
CONTATO COM MEIOS DE COMUNICAÇÃO		
22. ASSISTE TV?	23. PROGRAMAS FAVORITOS () Novelas () Prog. Religioso () Esportes () Filmes () Noticiário () Prog. Auditório () Outros	
24. OUVI RÁDIO? (FORA DO EXPEDIENTE DE TRABALHO)	25. PROGRAMAS PREFERIDOS: () Noticiário geral () Esportes () Prog. Religioso () Música () Noticiário policial () Prog. c/ partic. do ouvinte () Outros	
26. LÊ JORNAL?	27. NOME DO JORNAL:	28. SEÇÕES QUE GOSTA DE LER NO JORNAL:
29. LÊ REVISTA:	30. NOME DA REVISTA (TIPO):	
31. FAZ OUTROS TIPOS DE LEITURA?	32. QUAIS E POR QUANTO TEMPO (POR SEMANA)	
33. QUE RELIGIÃO PRÁTICA?		
34. PARTICIPAÇÃO EM DIVERSÕES () CINEMA () TEATRO () SHOWS () MAN. FOLCLÓRICAS () FUTEBOL () OUTROS ESPORTES () OUTROS _____		
PARA PREENCHIMENTO APÓS A ENTREVISTA		
35. ESPONTANEIDADE DA ELOCUÇÃO:		
36. GRAU DE CONHECIMENTO ENTRE INFORMANTE E INQUIRIDOR:		
37. INTERFERÊNCIA OCASIONAL DE CIRCUNSTANTES:		
38. CARACTERIZAÇÃO SUMÁRIA DO(S) CIRCUNSTANTE(S):		
39. AMBIENTE DO INQUÉRITO:		
40. OBSERVAÇÕES:		
41. INQUIRIDOR:	42. CIDADE:	43. DATA:
		44. DURAÇÃO: